A Comparative Study of Statistical-based Change Detection Methods for Multidimensional and Multitemporal SAR Images

Ammar Mian [1,2]

[1]SONDRA, CentraleSupélec, 3 rue Joliot Curie, 91192 Gif-sur-Yvette, France
[2]LISTIC, Université Savoie Mont-Blanc, F-74944, Annecy le Vieux, France

E-mail: ammar.mian@centralesupelec.fr

Frédéric Pascal

Laboratoire des Signaux et Systèmes, CentraleSupélec, 3 rue Joliot Curie, 91192 Gif-sur-Yvette, France

E-mail: frederic.pascal@centralesupelec.fr

July 2019

Abstract. This paper addresses the problem of activity monitoring through change detection in a time series of multidimensional Synthetic Aperture Radar (SAR) images. Thanks to SAR sensors' all-weather and all-illumination acquisitions capabilities, this technology has become widely popular in recent decades when it concerns the monitoring of large areas. As a consequence, a plethora of methodologies to process the increasing amount of data has emerged.

In order to present a clear picture of available techniques from a practical standpoint, the current paper aims at presenting an overview of statistical-based methodologies which are adapted to the processing of noisy and multidimensional data obtained from the latest generation of sensors. To tackle the various big data challenges, namely the problems of missing data, outliers/corrupted data, hetergenenous data, robust alternatives are studied in the statistics and signal processing community. In peculiar, we investigate the use of advanced robust approaches considering non-Gaussian modeling which appear to be better suited to handle high-resolution heterogeneous images.

To illustrate the attractiveness of the different methodologies presented, a comparative study on real high-resolution data has been realized. From this study, it appears that robust methodologies enjoy better detection performance through a complexity trade-off with regards to other non-robust alternatives.

Keywords: Change detection, Synthetic Aperture Radar, Time Series, Activity monitoring

1. Introduction

1.1. Motivations

Change detection (CD) for remotely sensed images of the Earth has been a popular subject of study for recent decades. It has indeed attracted a plethora of scholars due to the various applications to military (activity monitoring) or civil (geophysics, disaster assessment, etc) applications. With the increase in the number of spatial missions with Synthethic Aperture Radar (SAR) sensors, the amount of readily available data has led to a big data era. In order to process this quantity, automatic algorithms to help analyzing time series of images have to be developed.

Thus, CD algorithms have been thoroughly investigated in the recent literature [1]. In that context, statistical based methodologies have been successful to process the images that are subjected to the well-known speckle noise. Notably, when it concerns the processing of multidimensional (such as Polarimetric SAR images) and multitemporal images, various methodologies have been proposed in the literature. The present paper aims at giving a short overview of parametric methodologies based on popular distribution models of the SAR data.

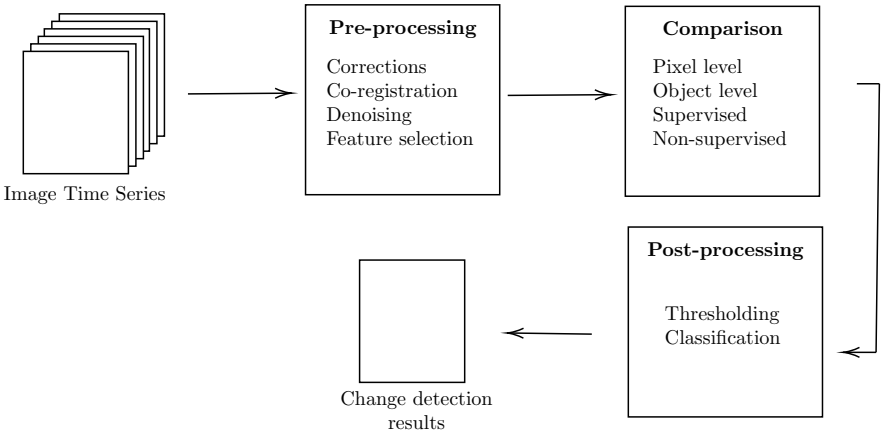

Figure 1. General procedure for a change detection methodology.

1.2. Change detection overview

The literature on the subject is dense, and a variety of methodologies are available‡. Broadly speaking, as illustrated in Figure 1, a change detection algorithm relies on three separate elements:

- A pre-processing phase in which the time series of images have to be co-registered, which means that by applying geometric transforms, each pixel of every image corresponds to the same physical localization. Various methodologies also consider a denoising step in which the speckle noise is reduced thanks to filtering techniques [3, 4].
  Lastly, features can be selected from the images for more concise or descriptive representation for CD purposes. Some possibilities include wavelet decomposition [5], Markov fields [6] or Principal Component Analysis [7] among others.
- A comparison step in which the features at each date are compared among themselves. This step consists in finding a relevant distance to measure dissimilarities of features between the different images of the series. Many techniques exist based on various principles.
- A post-processing step which varies depending on the methodology used for the comparison phase. It can either correspond to a thresholding [8, 9] or involves machine learning classification algorithms [10].

We will deal presently with comparison step methodologies. More specifically, we will place ourself in the context of pixel-level methodologies as opposed to object-level methodologies as defined in [1]. In this paradigm, the comparison is done on a local spatial neighborhood through the principle of the sliding windows: for each pixel a spatial neighborhood is defined and a distance function is computed to compare the similarities of this neighboorhood between the images at different time.

Statistical-based methodologies rely on a probability model in order to infer about changes. Those methods can rely on a model with associated parameters in which case they are referred to as parametric methodologies. Nonparametric methods consider statistical tools without considering parameters [11] or Bayesian models [12] with a prior on the distribution of those parameters. We will consider presently the parametric approach which has been the most popular subject of study in the recent decade thanks to the seminal work by Conradsen et al. [13] which showed the attractiveness of applying statistical results on the equality of covariance matrix for CD.

We will first introduce probability models used to model SAR images with associated parameters in section 2. Then in section 3, we will detail several dissimilarity functions referenced in the literature for SAR change detection. Finally, we will compare the performance of these functions on two distinct datasets in section 5 and conclude in section 6.

2. Statistical modeling of SAR images

2.1. The data

Denote by $\mathbb{W} = \{\mathbf{X}_1, \ldots, \mathbf{X}_T\}$ a collection of T mutually independent groups of p-dimensional i.i.d complex vectors: $\mathbf{X}_t = [\mathbf{x}_1^t, \ldots, \mathbf{x}_N^t] \in \mathbb{C}^{p\times N}$. With regards to the Single Look Complex (SLC) SAR

‡ Notably, see [1] or [2] for an overview.

images these sets correspond to the local observations on a spatially sliding windows as illustrated in Figure 2. The subscript k correspond to a spatial index while the superscript t corresponds to a time index.

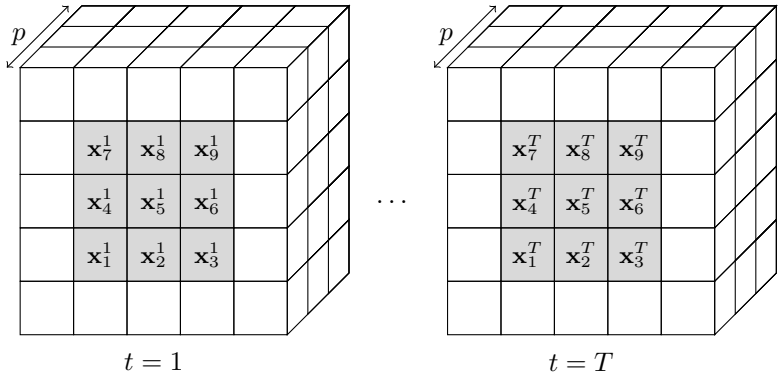

Figure 2. Illustration of the sliding windows (in grey) approach ($p = 3$, $N = 9$). The central pixel ($\mathbf{x}_5^t$) corresponds to the test pixel.

When it concerns Multilook Complex Images (MLC), the raw data $\mathbf{x}_k^t$ is not made available. The Sample Covariance Matrix (SCM) is then released to filter the speckle noise and compress the data. The number of pixels L used for the averaging is called the number of looks.

2.2. Gaussian model

The centered complex Gaussian distribution is the most encountered one in the SAR literature, notably for modelling Polarimetric images. Indeed, since each pixel consists of the coherent sum of the contribution of many scatterers, it is expected that, thanks to the central limit theorem, the Gaussian model is not too far from the actual empirical distribution. Given a random vector $\mathbf{x}$, it is said that $\mathbf{x}$ is distributed along the complex Gaussian distribution (which we denote $x \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, \mathbf{\Sigma})$) if its probability density function (p.d.f) is:

$$p_{\mathbf{x}}^{\mathbb{C}\mathcal{N}}(\mathbf{x}; \mathbf{\Sigma}) = \frac{1}{\pi^p \ |\mathbf{\Sigma}|} \exp\left(-\mathbf{x}^{\mathrm{H}} \mathbf{\Sigma}^{-1} \mathbf{x}\right). \tag{1}$$

For multilook images, the matrix obtained through the averaging data is assumed to be Wishart distributed as a consequence.

2.3. Non-Gaussian modeling

While the Gaussian distribution is popular, it fails to accurately describe the heterogeneity observed in very high-resolution images as described in [14], [15] or in [16]. Indeed, in those images, the amount of scatterers in each pixel has been greatly reduced with regards to low-resolution images.

To better describe the observed distribution of data, other models have been considered. For example, the K-distribution has been considered in [17, 18], the Weibull distribution in [19] or inverse Generalized Gaussian distribution in [20]. These various models belong to the family of complex elliptical distributions which generalizes them as discussed in [21] which is a model depending on a density generator function $g : \mathbb{R}^+ \to \mathbb{R}^+$ that satisfies the condition $\mathfrak{m}_{p,g} = \int_{\mathbb{R}^+} t^{p-1} g(t) \mathrm{d}t < \infty$ and a positive definite matrix $\mathbf{\Xi} \in \mathbb{S}_{\mathbb{H}}^p$: a random vector $\mathbf{x}$ follows a complex elliptical distribution (denoted $\mathbf{x} \sim \mathbb{C}\mathcal{E}(\mathbf{0}_p, g, \mathbf{\Xi})$) if its p.d.f is the following:

$$p_{\mathbf{x}}^{\mathbb{C}\mathcal{E}}(\mathbf{x}; \mathbf{\Xi}, g) = \mathfrak{C}_{p,g} |\mathbf{\Xi}|^{-1} g\left(\mathbf{x}^{\mathrm{H}} \mathbf{\Xi}^{-1} \mathbf{x}\right), \tag{2}$$

where $\mathfrak{C}_{p,g}$ is a normalization constraint ensuring that $\int_{\mathbb{C}^p} p_{\mathbf{x}}^{\mathbb{C}\mathcal{E}}(\mathbf{x}; \mathbf{\Xi}, g) \mathrm{d}\mathbf{x} = 1$.

Another representation of those distributions can be found through the compound-Gaussian model, sometimes referred as product model. In this case, a random vector $\mathbf{x}$ follows a complex compound-Gaussian (CCG) distribution if:

$$\mathbf{x} = \sqrt{\tau} \mathbf{z}, \tag{3}$$

where $\mathbf{z} \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, \boldsymbol{\xi})$, called the speckle and τ, called the texture, follows a distribution probability on $\mathbb{R}^+$. This quantity is often assumed to be deterministic in order to generalize the Gaussian distribution

without having to consider a model on the texture. In this case, the model on the pixels of the sliding windows is written:

$$\mathbf{x}_k^t \sim \sqrt{\tau_k^t}\mathbf{z}_k^t, \tag{4}$$

where $\tau_k^t \in \mathbb{R}^+$ is deterministic and $\mathbf{z}_k^t \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, \boldsymbol{\xi}_t)$.

The elliptical and CCG distributions are probability models which are in the scope of the robust statistical literature initiated by works such as [22, 23, 24]. More details can be found in the books [25] and [26]. Following their definition, a method is said to be robust, in the context of this paper, when its statistical properties are independent of the density generator function in the elliptical case or independent of the set of texture parameters for the deterministic CCG case.

Using these various models, we will describe, hereafter, various dissimilarity measures for those models which can be used for change detection.

3. Dissimilarity measurement: an overview

3.1. Posing the problem

Under a parametric approach, CD can be achieved by deciding between the two following alternative hypotheses:

$$\begin{cases} \mathrm{H}_0 : & \boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_T = \boldsymbol{\theta}_0 \quad (no\, change), \\ \mathrm{H}_1 : & \exists (t,t'),\, \boldsymbol{\theta}_t \neq \boldsymbol{\theta}_{t'} \quad (change) \end{cases}. \tag{5}$$

where $\boldsymbol{\theta}$ corresponds to the parameters of the distribution used as a model.

In order to decide, a dissimilarity measure between the data over time is needed. It can be seen as a function, also called statistic:

$$\hat{\Lambda} : \begin{array}{lll} \mathbb{C}^{p\times N\times T} & \rightarrow & \mathbb{R} \\ \mathbb{W} & \rightarrow & \hat{\Lambda}(\mathbb{W}), \end{array}$$

such that $\hat{\Lambda}(\mathbb{W})$ is high when H_1 is true and low otherwhise.

An interesting property of such function consists in the Constant False Alarm Rate (CFAR) property. This property is valid when the distribution of the statistic under H_0 hypothesis is not a function of the parameters of the problem. This allows selecting a threshold value for the detection which is directly linked to the probability of false alarms. The threshold is often obtained by deriving the distribution of the statistic or can be obtained by Monte-Carlo simulations when it is not available.

3.2. Hypothesis testing statistics

The first kind of statistics comes from the statistical literature on hypothesis testing. Several techniques have been developed to obtain a statistic of decision.

A comparative study of this framework can be found in [27]. The analysis presented allowed to show that many approaches led to equivalent results. Thus, three statistically independent statistics remain:

- the Generalized Likelihood Ratio Test (GLRT) statistic:

$$\hat{\Lambda}_{\mathrm{G}} = \frac{\left|\hat{\boldsymbol{\Sigma}}_0^{\mathrm{SCM}}\right|^{TN}}{\prod_{t=1}^{T}\left|\hat{\boldsymbol{\Sigma}}_t^{\mathrm{SCM}}\right|^{N}}, \tag{6}$$

  where:

$$\forall t, \hat{\boldsymbol{\Sigma}}_t^{\mathrm{SCM}} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_k^t{\mathbf{x}_k^t}^{\mathrm{H}} \text{ and } \hat{\boldsymbol{\Sigma}}_0^{\mathrm{SCM}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\Sigma}}_t^{\mathrm{SCM}}. \tag{7}$$

  This statistic is well known in the literature and its properties have been well studied. [28] has derived the distribution under null hypothesis and proposed an approximation as a function of χ^2 distributions.

- the t_1 statistic:

$$\hat{\Lambda}_{\mathrm{t}_1} = \frac{1}{T}\sum_{t=1}^{T}\mathrm{Tr}\left[\left(\left(\hat{\boldsymbol{\Sigma}}_0^{\mathrm{SCM}}\right)^{-1}\hat{\boldsymbol{\Sigma}}_t^{\mathrm{SCM}}\right)^2\right]. \tag{8}$$

  This statistic has the CFAR property and has a χ^2 asymptotic distribution (when $N \rightarrow \infty$).

- the Wald statistic:

$$\hat{\Lambda}_{\text{Wald}} = N \sum_{t=2}^{T} \text{Tr}\left[\left(\mathbf{I}_p - \hat{\mathbf{\Sigma}}_1^{\text{SCM}}(\hat{\mathbf{\Sigma}}_t^{\text{SCM}})^{-1}\right)^2\right] \tag{9}$$

$$-q\left(N\sum_{t=1}^{T}(\hat{\mathbf{\Sigma}}_t^{\text{SCM}})^{-\dagger} \otimes (\hat{\mathbf{\Sigma}}_t^{\text{SCM}})^{-1}, \text{vec}\left(\sum_{t=2}^{T}\mathbf{\Upsilon}_t\right)\right), \tag{10}$$

where † is the transpose operator and:

$$\mathbf{\Upsilon}_t = N\left((\hat{\mathbf{\Sigma}}_t^{\text{SCM}})^{-1} - (\hat{\mathbf{\Sigma}}_t^{\text{SCM}})^{-1}\hat{\mathbf{\Sigma}}_1^{\text{SCM}}(\hat{\mathbf{\Sigma}}_t^{\text{SCM}})^{-1}\right), \tag{11}$$

$$q(\mathbf{x}, \mathbf{\Sigma}) = \mathbf{x}^{\text{H}}\mathbf{\Sigma}^{-1}\mathbf{x}. \tag{12}$$

This statistic has the CFAR property and has a χ^2 asymptotic distribution (when $N \to \infty$). It is, however, computationally more complex than the previous ones due to the Kroenecker product.

In order to deal with multilook images and for a pair of two images ($T = 2$), [29] proposed to use the Complex Hotelling-Lawley statistic given by:

$$\hat{\Lambda}_{\text{HTL}} = \text{Tr}((\hat{\mathbf{\Sigma}}_0^{\text{SCM}})^{-1}\hat{\mathbf{\Sigma}}_1^{\text{SCM}}), \tag{13}$$

where the estimates correspond to the multilook covariance matrices at date 0 and 1. The authors also derived the distribution under null hypothesis and showed that it can be approximated by a Fisher-Snecdor distribution. The study showed a potential improvement of detection rate compared to the GLRT statistic.

Concerning non-Gaussian models, [30] has considered the derivation of the GLRT using the compound-Gaussian model with deterministic texture parameters:

$$\hat{\Lambda}_{\text{MT}} = \frac{\left|\hat{\boldsymbol{\xi}}_0^{\text{MT}}\right|^{TN}}{\prod_{t=1}^{T}\left|\hat{\boldsymbol{\xi}}_{\mathbb{W}_t}^{\text{TE}}\right|^{N}} \prod_{k=1}^{N} \frac{\left(\sum_{t=1}^{T} q\left(\hat{\boldsymbol{\xi}}_0^{\text{MT}}, \mathbf{x}_k^t\right)\right)^{Tp}}{T^{Tp}\prod_{t=1}^{T}\left(q\left(\hat{\boldsymbol{\xi}}_{\mathbb{W}_t}^{\text{TE}}, \mathbf{x}_k^t\right)\right)^{p}}, \tag{14}$$

where $\mathbb{W}_t = \{\mathbf{x}_k^t : 1 \leq k \leq N\}$,

$$\hat{\boldsymbol{\xi}}_{\mathbb{W}_t}^{\text{TE}} = \frac{p}{N}\sum_{\mathbf{x}_k^t \in \mathbb{W}_t} \frac{\mathbf{x}_k^t \mathbf{x}_k^{t\,\text{H}}}{q(\hat{\boldsymbol{\xi}}_{\mathbb{W}_t}^{\text{TE}}, \mathbf{x}_k^t)}, \text{ and } \hat{\boldsymbol{\xi}}_0^{\text{MT}} = \frac{p}{N}\sum_{k=1}^{N} \frac{\sum_{t=1}^{T}\mathbf{x}_k^t \mathbf{x}_k^{t\,\text{H}}}{\sum_{t=1}^{T} q\left(\hat{\boldsymbol{\xi}}_0^{\text{MT}}, \mathbf{x}_k^t\right)}, \tag{15}$$

and $q(\boldsymbol{\xi}, \mathbf{x})$ is defined at eq. (12).

This statistic has been shown to have better robustness to very heterogeneous data for which the Gaussian model is very far. This comes however, at a cost of complexity. Indeed, there is a need to compute the solution of two fixed point equations which is computationally expensive. The statistic has the CFAR property but the distribution under null hypothesis has not yet been derived.

3.3. Information-theoretic measures

Information theory tools are based on another approach which consists in measuring quantities related to the information at disposition on the data. In order to compare two distributions, distances can be computed. When a parametric model is used, the distance is often dependent on the parameters of the two distributions compared. Since in many applications, the parameters are not known, estimates are plugged into the distance function.

Kullback-Leibler divergence is a popular measure encountered, notably in SAR change detection problems [31]. For the Gaussian case, a distance can be obtained through the symmetrised version of the divergence:

$$\hat{\Lambda}_{\text{KL}} = \frac{1}{2}\left(d(\hat{\mathbf{\Sigma}}_0^{\text{SCM}}, \hat{\mathbf{\Sigma}}_1^{\text{SCM}}) + d(\hat{\mathbf{\Sigma}}_1^{\text{SCM}}, \hat{\mathbf{\Sigma}}_0^{\text{SCM}})\right), \tag{16}$$

where $d(\mathbf{A},\mathbf{B}) = \mathrm{Tr}(\mathbf{B}^{-1}\mathbf{A}) + \log\frac{|\mathbf{A}|}{|\mathbf{B}|}$. [32] has shown that this distance is asymptotically distributed as a χ^2 distribution.

Other measures consist in Renyi entropy or Bhattacharya distance which have been shown to behave similarly in [32]. They will thus be omitted in the present paper.

For non-Gaussian model, [33] proposed a measure in the multilook case and for $T = 2$, based on the principle of mutual information. This distance, based on Gamma model on the texture parameters, is a generalization of the one proposed in [34]:

$$\hat{\Lambda}_{\mathrm{MLL}} = \mathrm{MLL}(\mathbb{W}_1) + \mathrm{MLL}(\mathbb{W}_2) - \mathrm{MLL}(\mathbb{W}_{12}), \tag{17}$$

where $\mathbb{W}_{12} = \mathbb{W}_1 \cup \mathbb{W}_2$,

$$\begin{aligned}\mathrm{MLL}(\mathbb{W}) = N\frac{\nu+pL}{2}(\log(L\nu) - \log(\mu)) - NL\log|\boldsymbol{\xi}_{\mathbb{W}}^{\mathrm{TE}}| \\ + \frac{\nu-pL}{2}\sum_{\mathbf{x}_k^t\in\mathbb{W}}\log\left[\mathrm{Tr}\left((\boldsymbol{\xi}_{\mathbb{W}}^{\mathrm{TE}})^{-1}\mathbf{x}_k^t{\mathbf{x}_k^t}^{\mathrm{H}}\right)\right] - N\log|\Gamma(\nu)| \\ + \sum_{\mathbf{x}_k^t\in\mathbb{W}}\log\left[K_{\nu-pL}\left(2\sqrt{\mathrm{Tr}\left((\boldsymbol{\xi}_{\mathbb{W}}^{\mathrm{TE}})^{-1}\mathbf{x}_k^t{\mathbf{x}_k^t}^{\mathrm{H}}\right)L\nu/\mu}\right)\right],\end{aligned} \tag{18}$$

Γ is the Gamma function, L is the number of equivalent looks, $\boldsymbol{\xi}_{\mathbb{W}}^{\mathrm{TE}}$ is defined at eq. (15), $K_\nu(z)$ is the modified Bessel function of the second kind and (μ,ν) are the parameters of a Gamma distribution found by fitting the distribution of the texture parameters set

$$\mathcal{T} = \{\hat{\tau}_k^t = q(\boldsymbol{\xi}_{\mathbb{W}}^{\mathrm{TE}}, \mathbf{x}_k^t)/p : \mathbf{x}_k^t \in \mathbb{W}\}. \tag{19}$$

The method used yields a high computational cost, since the test statistic relies on a Bessel function as well as a fitting of a Gamma distribution.

3.4. Riemannian geometry distances

Riemannian geometry is an alterative concept which allows to compare distributions. Indeed, when the parameters of a distribution lie in a Riemannian manifold (for example, covariance matrices lie in Riemannian manifold of positive definite matrices $\mathbb{S}_{\mathbb{H}}^p$ [35]), it is possible to define a metric which can be related to the Kullback-Leibler divergence. Consequently, distances between two probability distributions which take into account the geometry properties of the parameter space, have been considered in the litterature. Notably, for a Gaussian model, we have the following distance [36]:

$$\hat{\Lambda}_{\mathcal{RE}} = \|\log\left((\hat{\boldsymbol{\Sigma}}_1^{\mathrm{SCM}})^{-\frac{1}{2}}\hat{\boldsymbol{\Sigma}}_2^{\mathrm{SCM}}(\hat{\boldsymbol{\Sigma}}_1^{\mathrm{SCM}})^{-\frac{1}{2}}\right)\|_F^2, \tag{20}$$

where log is the logarithm of matrices.

For the elliptical case, a Riemannian CES distance can be obtained as well [37]:

$$\hat{\Lambda}_{\mathcal{RE}} = \alpha\sum_{i=1}^{p}\log^2\lambda_i + \beta\left(\sum_{i=1}^{p}\log\lambda_i\right)^2 \tag{21}$$

where λ_i is the i-th eigenvalue of $(\hat{\boldsymbol{\Sigma}}_1^{\mathrm{MLE}})^{-1}\hat{\boldsymbol{\Sigma}}_2^{\mathrm{MLE}}$, $\hat{\boldsymbol{\Sigma}}_{\epsilon\in\{0,1\}}^{\mathrm{MLE}}$ is the MLE of the scatter matrix in CES case and α,β depends on the density generator function. For a Student-t distribution, we have:

$$\alpha = \frac{d+p}{d+p+1}, \beta = \alpha - 1, \hat{\boldsymbol{\Sigma}}_\epsilon^{\mathrm{MLE}} = \frac{d+p}{N}\sum_{k=1}^{N}\frac{\mathbf{x}_k^\epsilon{\mathbf{x}_k^\epsilon}^{\mathrm{H}}}{d+{\mathbf{x}_k^\epsilon}^{\mathrm{H}}(\hat{\boldsymbol{\Sigma}}_\epsilon^{\mathrm{MLE}})^{-1}\mathbf{x}_k^\epsilon}. \tag{22}$$

where d is the number of degrees of freedom of the Student-t distribution.

The concept of Riemannian geometry has also been considered in [38], where a methodology has been proposed for polarimetric SAR images based on Kennaugh matrix decomposition. This distance, however, is not based on a probability model and will be thus omitted in the present paper.

Table 1. Summary of statistics with their respective properties.

Statistic	Reference	Model	Eq.	$T > 2$	SLC/MLC	CFAR	Asymptotic null distribution
$\hat{\Lambda}_{\mathrm{G}}$	[13]	Gaussian	(6)	✓	Both	✓	χ^2 (asymptotic)
$\hat{\Lambda}_{t_1}$	[27]	Gaussian	(8)	✓	Both	✓	χ^2 (asymptotic)
$\hat{\Lambda}_{\mathrm{Wald}}$	[27]	Gaussian	(10)	✓	Both	✓	χ^2 (asymptotic)
$\hat{\Lambda}_{\mathrm{HTL}}$	[29]	Gaussian	(13)	×	Both	✓	Fisher-Snecdor (approximation)
$\hat{\Lambda}_{\mathrm{MT}}$	[30]	Deterministic CCG	(14)	✓	SLC	✓	×
$\hat{\Lambda}_{\mathrm{KL}}$	[32]	Gaussian	(16)	×	Both	✓	χ^2 (asymptotic)
$\hat{\Lambda}_{\mathrm{MLL}}$	[33]	Compound-Gaussian	(17)	×	MLC	×	×
$\hat{\Lambda}_{\mathcal{RG}}$	[36]	Gaussian	(20)	×	Both	×	×
$\hat{\Lambda}_{\mathcal{RE}}$	[37]	Student-t	(21)	×	SLC	×	×
$\hat{\Lambda}_{\mathcal{WG}}$	[40]	Gaussian	(24)	×	Both	×	×
$\hat{\Lambda}_{\mathcal{WE}}$	[40]	Elliptical	(25)	×	SLC	×	×

3.5. Optimal transport

Finally a distance can be obtained through the concept of optimal transport which defines a distance depending on a cost to *displace* parts of a reference distribution into a target distribution. It has been shown in [39], that this distance has a closed-form for elliptical distributions:

$$d(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \mathrm{Tr}(\boldsymbol{\Sigma}_1) + \mathrm{Tr}(\boldsymbol{\Sigma}_2) - 2\mathrm{Tr}\left[\left(\boldsymbol{\Sigma}_1^{\frac{1}{2}}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{\frac{1}{2}}\right)^{\frac{1}{2}}\right]. \tag{23}$$

Thanks to this result, we can define two statistics for Gaussian and Elliptical case by plugging estimates of the covariance matrices. A Wasserstein statistic for Gaussian can be computed by taking:

$$\hat{\Lambda}_{\mathcal{WG}} = d(\hat{\boldsymbol{\Sigma}}_1^{\mathrm{SCM}}, \hat{\boldsymbol{\Sigma}}_2^{\mathrm{SCM}}). \tag{24}$$

A Wasserstein distance in the elliptical can be obtained by taking:

$$\hat{\Lambda}_{\mathcal{WE}} = d(\hat{\tau}_1\hat{\boldsymbol{\xi}}_1^{\mathrm{TE}}, \hat{\tau}_2\hat{\boldsymbol{\xi}}_2^{\mathrm{TE}}), \tag{25}$$

where $\hat{\boldsymbol{\xi}}_{\epsilon\in\{0,1\}}^{TE}$ is defined at eq. (15) and $\hat{\tau}_{\epsilon\in\{0,1\}} = (p/N)\sum_{k=1}^{N} q(\hat{\boldsymbol{\xi}}_{\epsilon}^{\mathrm{TE}}, \mathbf{x}_k^{\epsilon})$.

3.6. Summary

A summary of the different statistics with their respective properties can be found in Table 1.

4. Analysis of complexity

To compare the attractiveness of the different statistics described in section 3 with regards to time consumption, we consider an experimental Monte-Carlo analysis on synthetic Gaussian data.

A set of data $\mathbb{W}$ is generated for parameters values $p = 10$, $N = 25$ and $T = 2$. The Gaussian data is generated with a covariance matrix $(\boldsymbol{\Sigma})_{ij} = 0.5^{|i-j|}$ and 4000 Monte-Carlo trials have been done. For statistics involving fixed-point equation, we fixed the number of iterations to 15 and we choose $L = 1$ (Single Look data). Finally, the simulation has been done on a 2.50 GHz processor in Python 3.7. The mean time consumption obtained is presented in Table 2.

Table 2. Time consumption in seconds.

$\hat{\Lambda}_{\mathrm{G}}$	$\hat{\Lambda}_{t_1}$	$\hat{\Lambda}_{\mathrm{Wald}}$	$\hat{\Lambda}_{\mathrm{HTL}}$	$\hat{\Lambda}_{\mathrm{MT}}$	$\hat{\Lambda}_{\mathrm{KL}}$	$\hat{\Lambda}_{\mathrm{MLL}}$	$\hat{\Lambda}_{\mathcal{RG}}$	$\hat{\Lambda}_{\mathcal{RE}}$	$\hat{\Lambda}_{\mathcal{WG}}$	$\hat{\Lambda}_{\mathcal{WE}}$
0.001	0.001	0.003	0.001	0.004	0.001	0.006	0.011	0.007	0.002	0.007

Several observations can be made:

- Most of the Gaussian derived statistics have similar time consumption since the statistic are based on the computation of the SCM which is less expensive than a fixed-point estimation. Wald statistic is more expensive than the Gaussian GLRT and $t1$ statistics due to the inverse of a Kroenecker products which is more expensive than a simple inverse. $\hat{\Lambda}_{\mathcal{RG}}$ has the highest time consumption, even compared to non-Gaussian methodologies, due to the fact that it requires to compute the square root of matrices which can take a lot of time depending on the implementation (the one used correspond to the Scipy implementation which can be improved with regards to complexity).

- Concerning non-Gaussian statistics, the time consumption is generally higher than Gaussian methodologies due to the fixed-point estimation. The best methodology with regards to time consumption is $\hat{\Lambda}_{\mathrm{MT}}$ since it requires fewer operations than the others.

5. Experiments on real data

5.1. Methodology and aim of the study

We consider comparing the performance of detection for all the statistics presented in section 3 on real SLC polarimetric SAR datasets where a ground truth is available.

In order to compute the statistics, we use a sliding windows of size 5×5 ($N = 25$). The fixed-point estimates are computed with 5 iterations to save time. For the statistic $\hat{\Lambda}_{\mathrm{MLL}}$, since it consider a multilook image, we choose $L = 1$. For the statistic $\hat{\Lambda}_{\mathcal{RE}}$ based on the Student-t distribution, we choose $d = 3$ in order to have a distribution far from the Gaussian one. When the data correspond to a series of length $T > 2$, only the adapted statistics are computed.

The performance of detection is measured through the Receiver Operator Curve (ROC), which quantify the amount of good detection as a function of false alarms.

5.2. Datasets used

We consider three scenes of UAVSAR polarimetric data obtained from https://uavsar.jpl.nasa.gov (Courtesy NASA/JPL-Caltech) presented in Figures 3 and 4 with their associated ground truth. The first two scenes are referenced under label SanAnd_26524_03 Segment 4 and correspond to a pair of images at dates April 23, 2009, and May 11, 2015 ($T = 2$). The groud truth for these two scenes has been fetched from [41]. The third scene is referenced under label Snjoaq_14511 and correspond to a series of $T = 17$ images and the ground truth have been realized by the present authors.

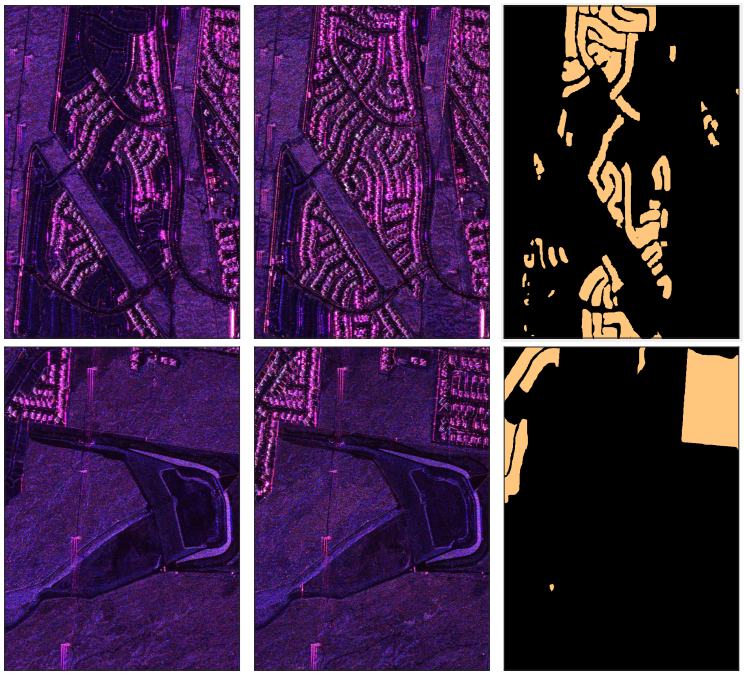

Figure 3. UAVSAR Scenes 1 (top) and 2 (bottom) in Pauli representation. Left: Image at first date. Middle: Image at second date. Right: Ground Truth.

5.3. Reproducibility

The codes (in Python 3.7) are available at *https://github.com/AmmarMian/WCCM-2019*. The datasets can be obtained from UAVSAR website (https://uavsar.jpl.nasa.gov) using the labels given and the ground truth may be obtained from the authors.

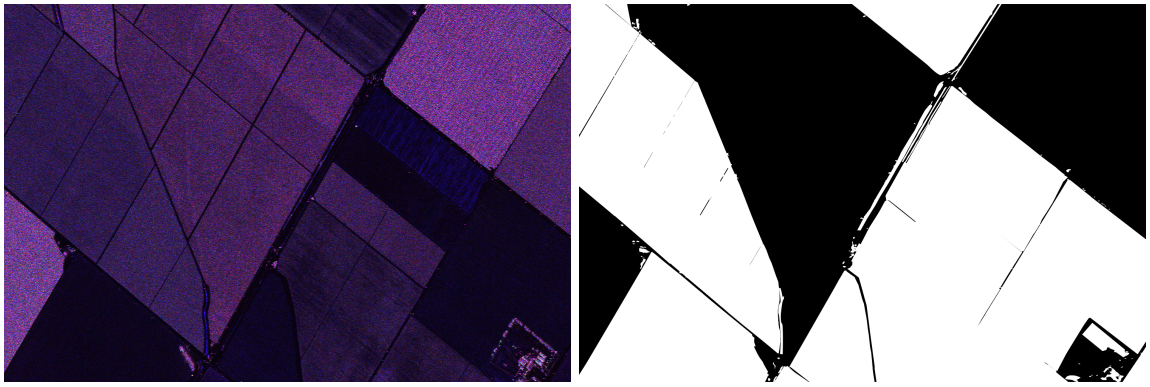

Figure 4. UAVSAR Scene 3 in Pauli representation. Left: Image at $t = 1$. Right: Ground Truth on the whole series.

5.4. Results and discussion

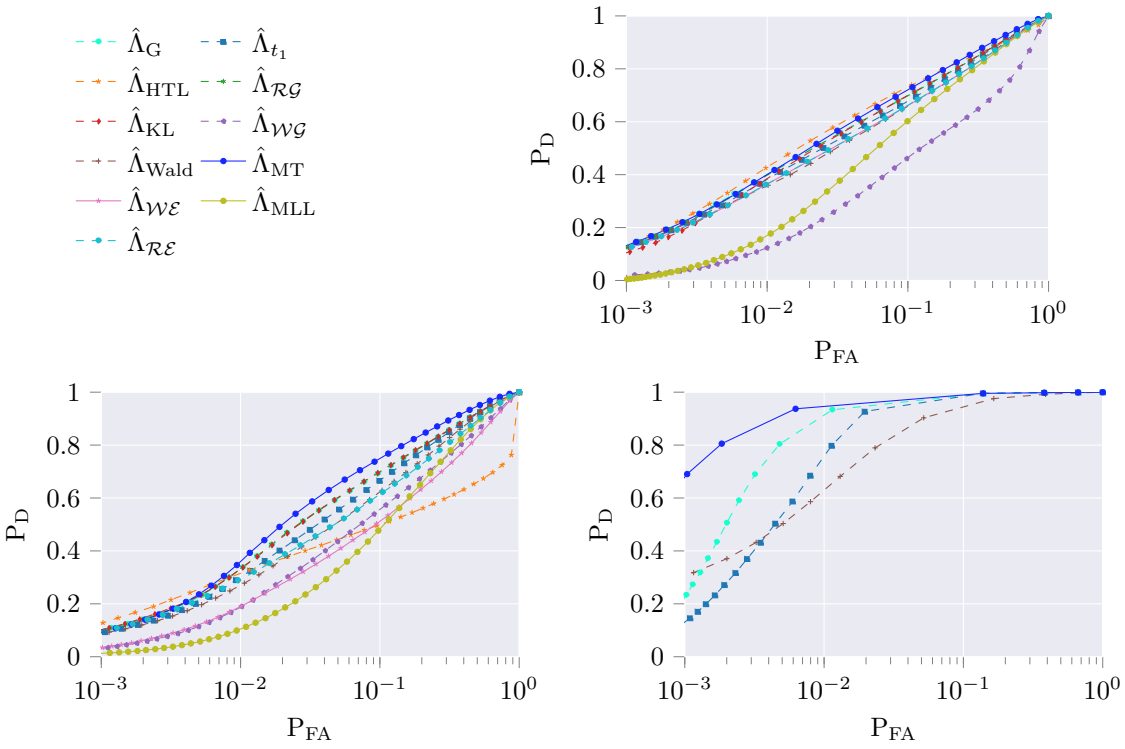

Figure 5. ROC plots for a window size of 5×5.

The ROC for all statistics and three scenes are presented in Figure 5. A transcription of the detection performance at $\mathrm{P_{FA}} = 0.01$ is also given in Table 3.

Table 3. Probability of detection at a false alarm rate of 1%.

Statistic	$\hat{\Lambda}_{\mathrm{G}}$	$\hat{\Lambda}_{t_1}$	$\hat{\Lambda}_{\mathrm{Wald}}$	$\hat{\Lambda}_{\mathrm{HTL}}$	$\hat{\Lambda}_{\mathrm{MT}}$	$\hat{\Lambda}_{\mathrm{KL}}$	$\hat{\Lambda}_{\mathrm{MLL}}$	$\hat{\Lambda}_{\mathcal{RG}}$	$\hat{\Lambda}_{\mathcal{RE}}$	$\hat{\Lambda}_{\mathcal{WG}}$	$\hat{\Lambda}_{\mathcal{WE}}$
Scene 1	0.41	0.41	0.35	**0.43**	0.41	0.40	0.18	0.41	0.39	0.12	0.36
Scene 2	0.33	0.30	0.27	0.32	**0.39**	0.32	0.11	0.33	0.29	0.19	0.19
Scene 3	0.90	0.75	0.62		**0.95**						

The first observation that can be made is that $\hat{\Lambda}_{\mathrm{MLL}}$ generally have poorer performance of detection than the other statistics. This is mostly explained by the fact that it has been developed for a multilook scenario which was no the case here. Thus, even by taking $L = 1$, the statistic yield poorer detection performance than most of the other statistics.

The statistics obtained through the Wasserstein distance, $\hat{\Lambda}_{\mathcal{WG}}$ and $\hat{\Lambda}_{\mathcal{WE}}$ appear to not work well in this problem of change detection since the performance of detection compared to others distance is lower, especially for Scene 2.

For the case $T = 2$, considering the Gaussian-derived statistics $\hat{\Lambda}_{\mathrm{G}}$, $\hat{\Lambda}_{t_1}$, $\hat{\Lambda}_{\mathrm{KL}}$, $\hat{\Lambda}_{\mathrm{HTL}}$ and $\hat{\Lambda}_{\mathcal{RG}}$, it is difficult to discern a best statistic since depending on the scene, their relative performance vary. When the number of images is high, $\hat{\Lambda}_{\mathrm{Wald}}$ appears to be lower in all cases but the performance of detection is not far. For $T > 2$, $\hat{\Lambda}_{\mathrm{G}}$ seems to have better performance than its counterpart but no conclusion can be made since only one set of data with $T > 2$ was available.

Finally, concerning robust statistics $\hat{\Lambda}_{\mathrm{MT}}$ has the overall best performance for all scenes. In scene 1, it has slightly lower performance than $\hat{\Lambda}_{\mathrm{HTL}}$ but on the other scenes (especially Scene 3), the performance of detection is greatly improved. The gain of this statistic compared to Gaussian one is explained by the heterogeneity of the images at the transition between objects. As an example, the output of both $\hat{\Lambda}_{\mathrm{G}}$ and $\hat{\Lambda}_{\mathrm{MT}}$ are presented in Figure 6. For this data, the Gaussian derived statistic, which does not take into account the heterogeneity, yields a high value of the statistic at the transition between the fields while the CCG statistic does not, which constitutes a big improvement concerning false alarms. The natural Student-t statistic $\hat{\Lambda}_{\mathcal{WE}}$ does not have good performance of detection which is explained by a mismatch in the degrees of freedom chosen to model the distribution of the data. Moreover, the Student-t model can be inaccurate to model the actual data. In those regards, $\hat{\Lambda}_{\mathrm{MT}}$ is able to obtain better performance without relying on a specific elliptical model and has thus a better robustness to mismatch scenarios.

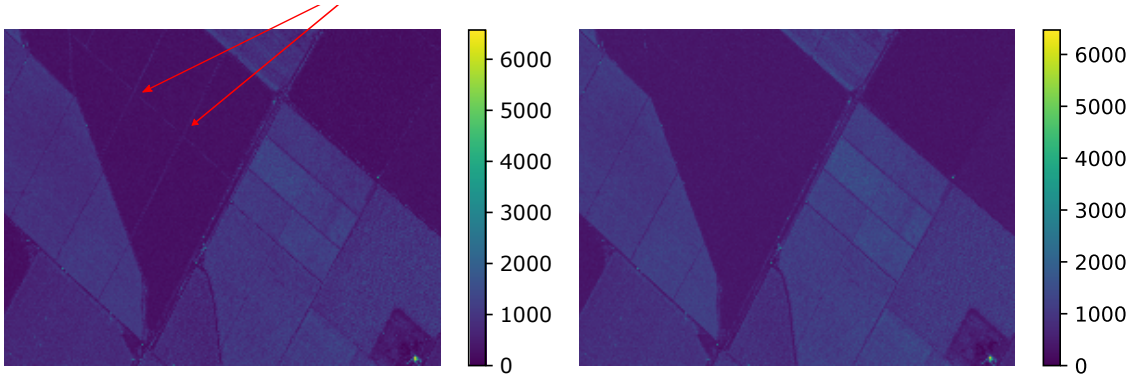

Figure 6. Output of $\hat{\Lambda}_{\mathrm{G}}$ (left) and $\hat{\Lambda}_{\mathrm{MT}}$ (right) on scene 3. The major difference between the two statistics is higlihted by red arrows.

6. Conclusions

We have described several parametric approaches for CD in multidimensional and multitemporal SAR images obtained from various principles and detailed their properties. The main result obtained through the comparative study on real data undertaken in section 5, was to show that the statistic $\hat{\Lambda}_{\mathrm{MT}}$, based on a deterministic CCG modelling, has allowed to obtain the best performance of detection thanks to its inherent robustness properties. Indeed, by taking into account the heterogeneity of the backscattering power on the local window size, the statistic has allowed to reduce the number of false alarms and kept the amount of true detection coherent with Gaussian-derived detectors. This improvement comes with a higher computational cost as explored in section 4 which has been shown to be reasonable with regards to its detection performance.

References

[1] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," {*ISPRS*} *Journal of Photogrammetry and Remote Sensing*, vol. 80, pp. 91 – 106, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271613000804

[2] A. Hecheltjen, F. Thonfeld, and G. Menz, *Recent Advances in Remote Sensing Change Detection – A Review*. Dordrecht: Springer Netherlands, 2014, pp. 145–178.

[3] A. Achim, P. Tsakalides, and A. Bezerianos, "Sar image denoising via bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 8, pp. 1773–1784, Aug 2003.

[4] S. Foucher and C. Lopez-Martinez, "Analysis, evaluation, and comparison of polarimetric SAR speckle filtering techniques," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1751–1764, April 2014.

[5] A. Mian, J.-P. Ovarlez, G. Ginolhac, and A. M. Atto, "Multivariate change detection on high resolution monovariate SAR image using linear Time-Frequency analysis," in *25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece, Aug. 2017.

[6] F. Wang, Y. Wu, Q. Zhang, P. Zhang, M. Li, and Y. Lu, "Unsupervised change detection on SAR images using triplet markov field model," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 697–701, July 2013.

[7] O. Yousif and Y. Ban, "Improving urban change detection from multitemporal SAR images using pca-nlm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2032–2041, April 2013.

[8] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[9] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2866–2878, Oct 2006.

[10] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, "Change detection in synthetic aperture radar images based on deep neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 1, pp. 125–138, Jan 2016.

[11] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and C. Zoppetti, "Nonparametric change detection in multitemporal SAR images based on mean-shift clustering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2022–2031, April 2013.

[12] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J. Tourneret, "Change detection for optical and radar images using a bayesian nonparametric model coupled with a markov random field," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 1513–1517.

[13] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "Change detection in polarimetric SAR data and the complex Wishart distribution," in *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, vol. 6, 2001, pp. 2628–2630 vol.6.

[14] M. S. Greco and F. Gini, "Statistical analysis of high-resolution SAR ground clutter data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 566–575, March 2007.

[15] G. Gao, "Statistical modeling of SAR images: A survey," *Sensors*, vol. 10, no. 1, pp. 775–795, 2010.

[16] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Compound-Gaussian clutter modeling with an inverse Gaussian texture distribution," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 876–879, Dec 2012.

[17] S. Yueh, J. Kong, J. Jao, R. Shin, and L. Novak, "K-distribution and polarimetric terrain radar clutter," *Journal of Electromagnetic Waves and Applications*, vol. 3, no. 8, pp. 747–768, 1989. [Online]. Available: https://doi.org/10.1163/156939389X00412

[18] H. . Muller, "K statistics of terrain clutter in high resolution SAR images," in *Proceedings of IGARSS '94 - 1994 IEEE International Geoscience and Remote Sensing Symposium*, vol. 4, Aug 1994, pp. 2146–2148 vol.4.

[19] T. Bucciarelli, P. Lombardo, C. J. Oliver, and M. Perrotta, "A compound weibull model for SAR texture analysis," in *1995 International Geoscience and Remote Sensing Symposium, IGARSS '95. Quantitative Remote Sensing for Science and Applications*, vol. 1, July 1995, pp. 181–183 vol.1.

[20] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric g distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/env.658

[21] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, Nov 2012.

[22] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, January 1976.

[23] V. J. Yohai, "Robust estimation in the linear model," *Ann. Statist.*, vol. 2, no. 3, pp. 562–567, 05 1974. [Online]. Available: https://doi.org/10.1214/aos/1176342717

[24] B. Martin and D. Pierre, "Robust estimation of the sur model," *Canadian Journal of Statistics*, vol. 28, no. 2, pp. 277–288. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.2307/3315978

[25] R. A. Maronna, D. R. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*, 1st ed., ser. Wiley Series in Probability and Statistics. Wiley, 2006.

[26] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge University Press, 2018.

[27] D. Ciuonzo, V. Carotenuto, and A. D. Maio, "On multiple covariance equality testing with application to SAR change detection," *IEEE Transactions on Signal Processing*, vol. 65, no. 19, pp. 5078–5091, Oct 2017.

[28] T. Anderson, *An Introduction to Multivariate Statistical Analysis*, ser. Wiley Series in Probability and Statistics. Wiley, 2003. [Online]. Available: https://books.google.es/books?id=Cmm9QgAACAAJ

[29] 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 Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 3953–3966, 2016.

[30] A. Mian, G. Ginolhac, J.-P. Ovarlez, and A. M. Atto, "New robust statistics for change detection in time series of multivariate SAR images," *IEEE Transactions on Signal Processing*, vol. 67, no. 2, pp. 520–534, Jan 2019.

[31] 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 Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1432–1445, May 2007.

[32] A. C. Frery, A. D. C. Nascimento, and R. J. Cintra, "Analytic Expressions for Stochastic Distances Between Relaxed Complex Wishart Distributions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, pp. 1213–1226, Feb. 2014.

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

[34] J. . Beaulieu and R. Touzi, "Segmentation of textured polarimetric sar scenes by likelihood approximation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 10, pp. 2063–2072, Oct 2004.

[35] L. T. Skovgaard, "A riemannian geometry of the multivariate normal model," *Scandinavian Journal of Statistics*, pp. 211–223, 1984.

[36] S. T. Smith, "Covariance, subspace, and intrinsic crame/spl acute/r-rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, May 2005.

[37] A. Breloy, G. Ginolhac, A. Renaux, and F. Bouchard, "Intrinsic cramér–rao bounds for scatter and shape matrices estimation in ces distributions," *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 262–266, Feb 2019.

[38] D. Ratha, S. De, T. Celik, and A. Bhattacharya, "Change Detection in Polarimetric SAR Images Using a Geodesic Distance Between Scattering Mechanisms," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 1066–1070, 2017.

[39] M. Gelbrich, "On a formula for the l2 wasserstein metric between measures on euclidean and hilbert spaces," *Mathematische Nachrichten*, vol. 147, no. 1, pp. 185–203, 1990.

[40] N. Ghaffari and S. Walker, "On multivariate optimal transportation," *arXiv preprint arXiv:1801.03516*, 2018.

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