3

An Overview of Covariance-based Change Detection Methodologies in Multivariate SAR Image Time-Series

AQ1

Ammar MIAN[1, 2], Guillaume GINOLHAC[2], Jean Philippe OVARLEZ[3], Arnaud BRELOY[4] and Frédéric PASCAL[1]

[1] *CentraleSupélec, Paris-Saclay University, Gif-sur-Yvette, France*
[2] *University Savoie Mont Blanc, Annecy, France*
[3] *ONERA, Palaiseau, France*
[4] *Paris Nanterre University, France*

3.1. Introduction

Change detection (CD) for remotely sensed images of the Earth has been a popular subject of study in the past decades. It has indeed attracted a plethora of scholars due to the various applications, in both military (activity monitoring) and civil (geophysics, disaster assessment, etc.) contexts. With the increase in the number of spatial missions with embedded synthetic aperture radar (SAR) sensors, the amount of readily available observations has now reached the "big data" era. To efficiently process and analyze this data, automatic algorithms have therefore to be developed. Notably, CD algorithms have been thoroughly investigated: the literature on the subject is dense, and a variety of methodologies can be envisioned[1].

1. See Hussain *et al*. (2013) or Hecheltjen *et al*. (2014) for overviews.

Change Detection and Image Time-Series Analysis 1,
coordinated by Abdourrahmane M. ATTO, Francesca BOVOLO and Lorenzo BRUZZONE.
© ISTE Ltd 2021.

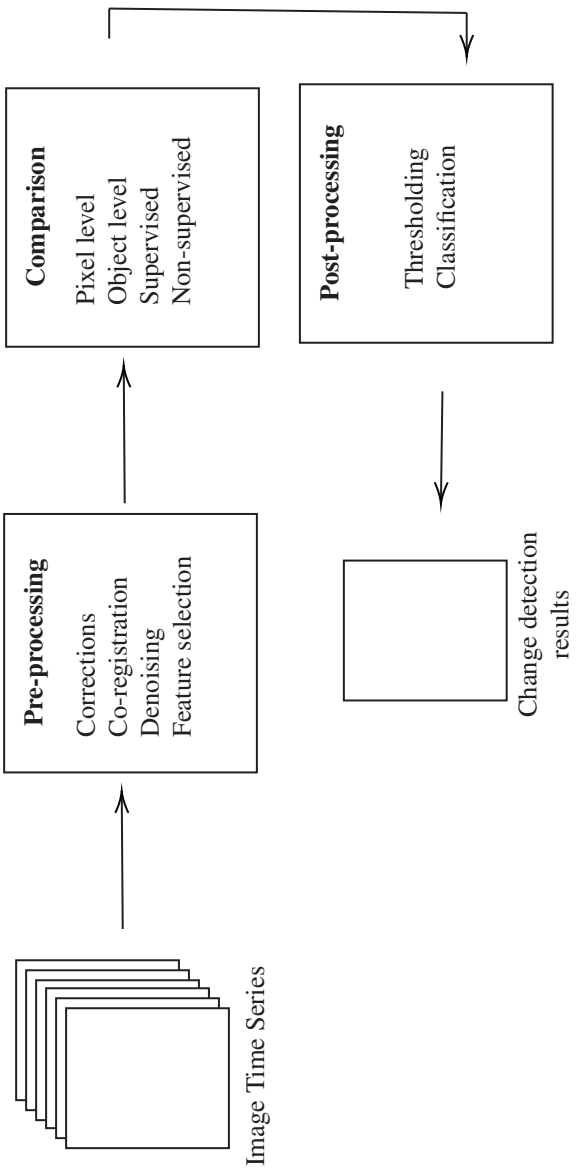


Figure 3.1. *General procedure for a change detection methodology*

Broadly speaking, a change detection algorithm can be synthesized as in Figure 3.1, and it relies on three main separate elements:

– a pre-processing phase, in which the time series of images have to be co-registered, meaning that a geometric transformation is applied so that each pixel of every image corresponds to the same physical localization. Various methodologies also considered a denoising step in which the speckle noise is reduced thanks to filtering techniques (Achim *et al*. 2003; Foucher and Lopez-Martinez 2014). Finally, local features can be selected to obtain a more concise (or descriptive) representation of the data. Among others, some possibilities include wavelet decompositions (Mian *et al*. 2017, 2019b), Markov fields (Wang *et al*. 2013) or principal component analysis (Yousif and Ban 2013);

– a comparison step, in which the selected features from each date are compared among themselves. This step consists of building a relevant distance to measure dissimilarities of the features. Depending on the feature space, there generally exist many principles (geometry, statistics, etc.) to choose/build this distance;

– a post-processing step, which varies depending on the methodology used for the comparison step. It can either correspond to a thresholding (Bruzzone and Prieto 2000; Kervrann and Boulanger 2006) or involve machine learning classification algorithms (Gong *et al*. 2016).

In this large context, this chapter proposes an overview of the comparison step methodologies based on the local covariance matrices of the multivariate pixels. The scope of this overview can thus be specified by the following remarks:

– the focus is put on unsupervised CD methodologies. The CD problem is indeed challenging due to the lack of available ground truths, which does not allow applying supervised methods from the image processing literature. Moreover, it is well known that SAR images are subjected to speckle noise, which makes traditional optical approaches prone to high false alarm rates. In this case, unsupervised methodologies, often based on statistical tools, have yielded interesting approaches;

– we consider multivariate pixels, where the channels can correspond to a polarimetric diversity, or another kind of diversity depending on the pre-processing. The main example will be the spectro-angular diversity, obtained through wavelet transforms (Mian *et al*. 2019b);

– the presented methodologies are *pixel-level*, as opposed to *object-level* ones (Hussain *et al*. 2013). In this paradigm, the comparison is done on a local spatial neighborhood (patches) 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 neighborhood between the images at different time;

– we consider statistical parametric methodologies, as opposed to non-parametric ones (Aiazzi *et al*. 2013; Prendes *et al*. 2015). Hence, the CD process will rely on a probability model and its associated parameters. Within this methodology, the

covariance matrix of the local patches will be considered as a main feature to infer about changes in the time series.

The covariance-based approach has been a popular subject of study thanks to the seminal work (Conradsen *et al*. 2001), which demonstrated its use for CD in multivariate SAR image time series (ITS). One of its main interests is notably the possibility to leverage well-established results from the statistical literature, which is why numerous works consequently followed such an approach. Our aim is therefore to present a general overview of this topic, as well as recent developments considering the choice of statistical models, the dissimilarity measure and the use of structured covariance matrices.

The rest of this chapter is organized as follows. Section 3.2 describes the dataset and the different scenes used to illustrate the performance of the change detectors[2]. Section 3.3 introduces several families of elliptical distributions (and associated parameters) that can be used to model multivariate SAR images. Section 3.4 details several dissimilarity functions based on covariance matrices, which are then compared for CD on the considered datasets. Section 3.5 presents an extension of a statistical detection methodology that allows us to account for low-rank structures in the covariance matrix, whose interest is also illustrated on the real dataset. Finally, section 3.6 draws the conclusions and perspectives of this study.

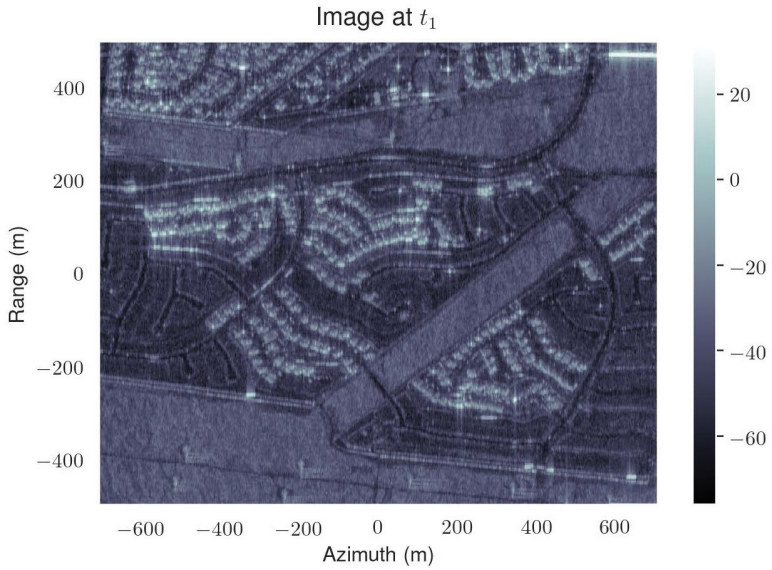
3.2. Dataset description

We consider three SAR ITS datasets from UAVSAR (courtesy of NASA/JPL-Caltech). The two first scenes are displayed, respectively, in Figures 3.2 (four images) and 3.3 (two images). The CD ground truths for these scenes are collected from Ratha *et al*. (2017) and Nascimento *et al*. (2019) and are shown in Figures 3.5. The third scene, displayed in Figure 3.4, is referenced under label Snjoaq_14511 and corresponds to a series of $T = 17$ images. The ground truth has been realized by the present authors. The SAR images correspond to full-polarization data with a resolution of 1.67 m in range and 0.6 m in azimuth. Since the scatterers present in the scenes exhibit an interesting spectro-angular behavior, each polarization of these images can be subject to the wavelet transform from Mian *et al*. (2019b), yielding pixels of dimension $p = 12$ (only used in the last section). Table 3.1 gives an overall perspective of the scenes used in the study.

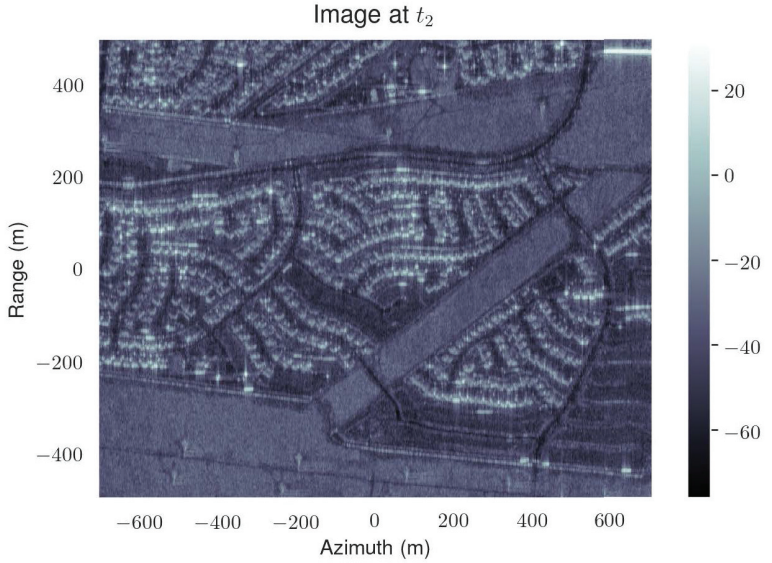
2. The codes (in Python 3.7) for all the change detectors are available at *https://github.com/AmmarMian/WCCM-2019* for section 3.4 and at *https://github.com/AmmarMian/Robust-Low-Rank-CD* for section 3.4. The datasets can be obtained from UAVSAR website (https://uavsar.jpl.nasa.gov) using the labels given and the ground truth may be obtained from the author's website: *https://ammarmian.github.io/*.

Dataset	Url	Resolution	Scene	p	T	Size	Coordinates (top-left px)
UAVSAR SanAnd_26524_03 Segment 4 April 23, 2009 - May 15, 2011	https://uavsar.jpl.nasa.gov	Rg: 1.67m Az: 0.6m	Scene 1	3 or 12	2 or 4	2360×600 px	[Rg, Az] = [2891, 28891]
			Scene 2	3 or 12	2	2300×600 px	[Rg, Az] = [3236,25601]
Snjoaq_14511			Scene 3	3	17	2300×600 px	[Rg, Az] = [3236,25601]

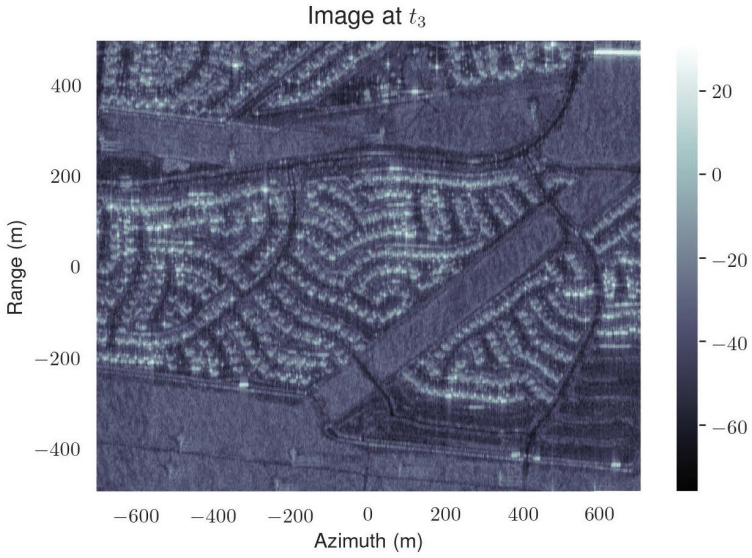
Table 3.1. *Description of SAR data*



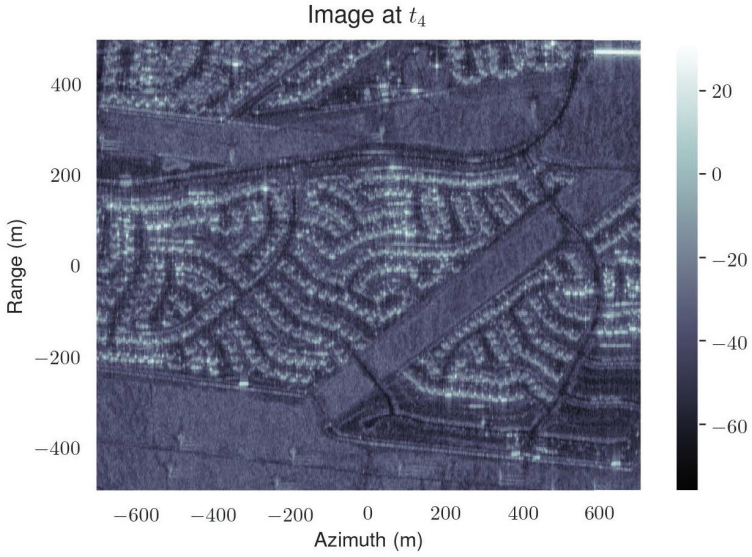
(a) Scene at t_1



(b) Scene at t_2

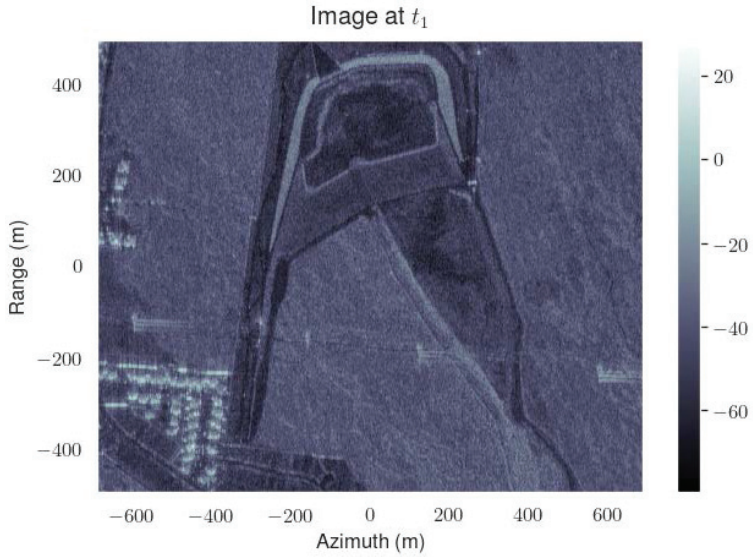


(c) Scene at t_3

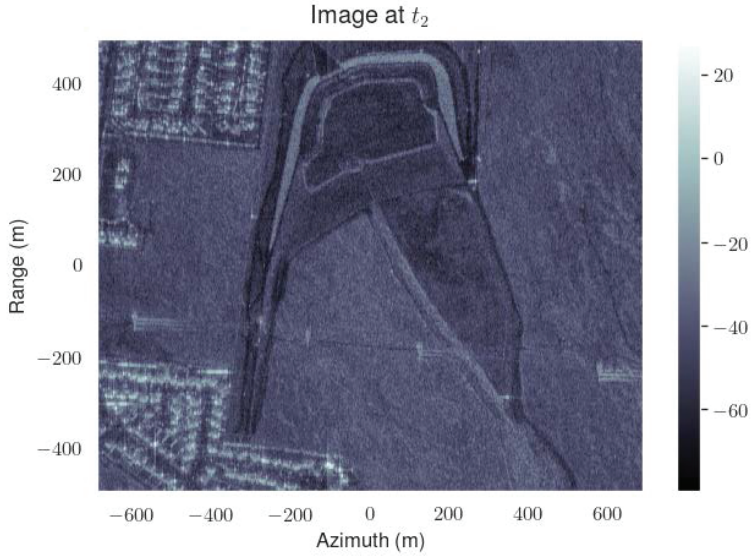


(d) Scene at t_4

Figure 3.2. *UAVSAR dataset used in this study for Scene 1. Four dates are available between April 23, 2009, and May 15, 2011. For a color version of this figure, see www.iste.co.uk/atto/change1.zip*



(a) Scene at t_1



(b) Scene at t_2

Figure 3.3. *UAVSAR dataset used in this study for Scene 2. Two dates are available between April 23, 2009, and May 15, 2011. For a color version of this figure, see www.iste.co.uk/atto/change1.zip*

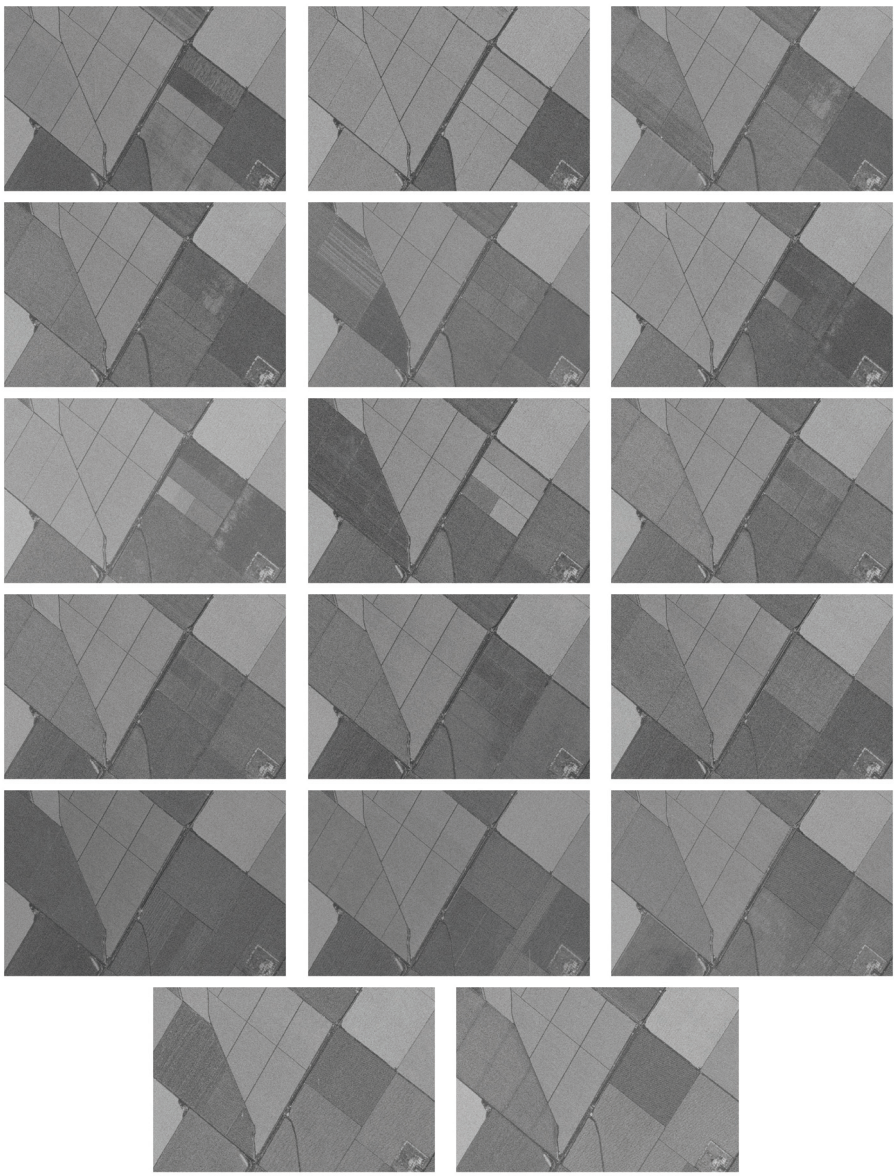
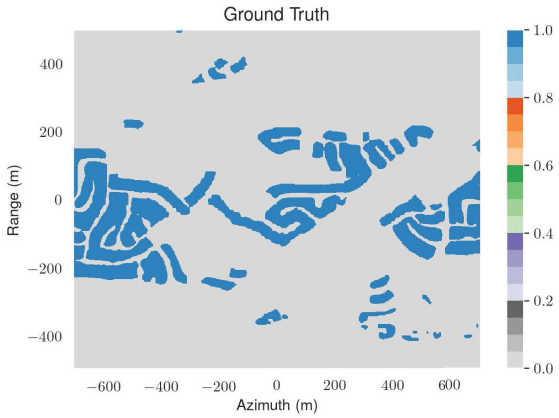
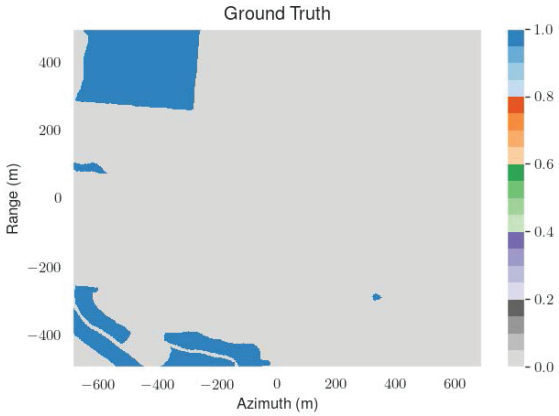


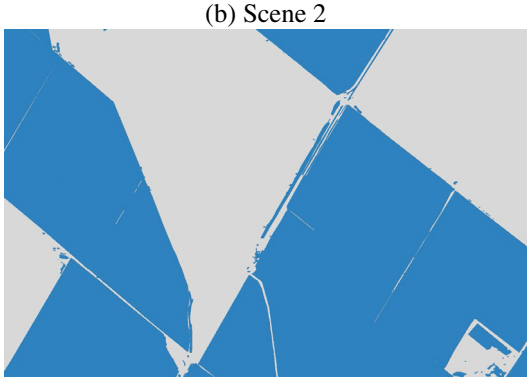
Figure 3.4. *UAVSAR dataset used in this study for Scene 3*



(a) Scene 1



(b) Scene 2



(c) Scene 3

Figure 3.5. *Ground truth for Scenes 1, 2 and 3. For a color version of this figure, see www.iste.co.uk/atto/change1.zip*

3.3. Statistical modeling of SAR images

3.3.1. *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 regard to the single look complex (SLC) SAR images, these sets correspond to the local observations in a spatially sliding windows as illustrated in Figure 3.6. The subscript k corresponds to a spatial index, while the superscript t corresponds to a time index.

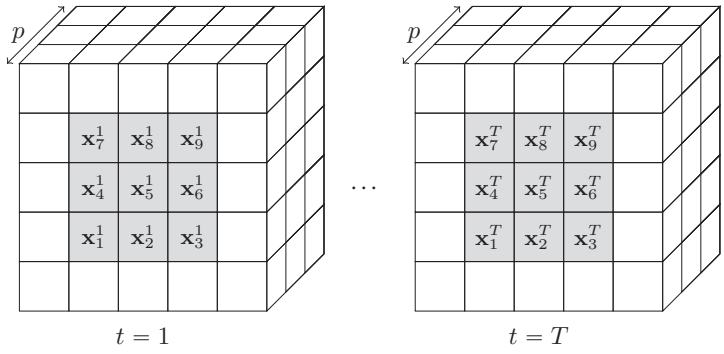


Figure 3.6. *Illustration of the sliding windows (in gray) approach ($p = 3$, $N = 9$). The central pixel ($\mathbf{x}_5^t$) corresponds to the test pixel*

3.3.2. *Gaussian model*

The centered complex Gaussian distribution is the most encountered one in the SAR literature, notably for modeling 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. A random vector $\mathbf{x} \in \mathbb{C}^p$ is said to be distributed along the complex Gaussian distribution (which we denote $x \sim \mathbb{C}\mathcal{N}$) if its probability density function (p.d.f) is:

$$p_{\mathbf{x}}^{\mathbb{C}\mathcal{N}}(\mathbf{x}; \boldsymbol{\Sigma}) = \frac{1}{\pi^p \left|\boldsymbol{\Sigma}\right|} e^{(} - \mathbf{x}^H \boldsymbol{\Sigma}^{-1} \mathbf{x}) \qquad [3.1]$$

When we have a dataset $\{\mathbf{x}_k\}_{k \in [1,N]}$ of i.i.d. data, the maximum likelihood estimator (MLE) of the covariance matrix is the well-known sample covariance matrix estimator (SCM):

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{x}_k \mathbf{x}_k^H \qquad [3.2]$$

This estimate matrix obtained through data averaging is assumed to be Wishart distributed as a consequence. In our time-series configuration, the statistical model for the pixels on a sliding window is:

$$\mathbf{x}_k^t \sim \mathcal{CN}(\mathbf{0}_p, \mathbf{\Sigma}_t) \quad [3.3]$$

3.3.3. *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 Greco and Gini (2007), Gao (2010) and Ollila *et al*. (2012b). Indeed, in those images, the amount of scatterers in each pixel has been greatly reduced with regard 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 Yueh *et al*. (1989) and Muller (1994), the Weibull distribution in Bucciarelli *et al*. (1995) or inverse generalized Gaussian distribution in Freitas *et al*. (2005). These various models belong to the family of complex elliptical symmetric (CES) distributions, which generalize them as discussed in Ollila *et al*. (2012a), which is a model depending on a density generator function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ that satisfies the condition $\mathfrak{m}_{p,g} = \int_{\mathbb{R}^+} t^{p-1} g(t)\, dt < \infty$ and a positive definite matrix $\mathbf{\Xi} \in \mathbb{S}_{\mathbb{H}}^p$: a random vector $\mathbf{x} \in \mathbb{C}^p$ follows a complex elliptical distribution (denoted $\mathbf{x} \sim \mathbb{CE}[\mathbf{0}_p][g][\mathbf{\Xi}]$) if its p.d.f is the following:

$$p_{\mathbf{x}}^{\mathbb{CE}}(\mathbf{x}; \mathbf{\Xi}, g) = \mathfrak{C}_{p,g}\, |\mathbf{\Xi}|^{-1}\, g\left(\mathbf{x}^H\, \mathbf{\Xi}^{-1}\, \mathbf{x}\right), \quad [3.4]$$

where $\mathfrak{C}_{p,g}$ is a normalization constraint ensuring that $\int_{\mathbb{C}^p} p_{\mathbf{x}}^{\mathbb{CE}}(\mathbf{x}; \mathbf{\Xi}, g) d\mathbf{x} = 1$. The MLE depends on the function g. For example, the t-distribution is also a particular case of the CES family, and in this case, g is equal to:

$$g(t) = \left(1 + \frac{2\,t}{d}\right)^{-(2p+d)/2}, \quad [3.5]$$

where d is the degree of freedom of the distribution. For this distribution and when we have a dataset $\{\mathbf{x}_k\}_{k\in[1,N]}$ of i.i.d. data, the MLE of the covariance matrix is the result of the following fixed-point equation:

$$\hat{\mathbf{\Sigma}}_{student} = \frac{d+p}{N} \sum_{k=1}^{N} \frac{\mathbf{x}_k\, \mathbf{x}_k^H}{d + \mathbf{x}_k^H\, \hat{\mathbf{\Sigma}}_{student}^{-1}\, \mathbf{x}_k} \quad [3.6]$$

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:

$$\mathbf{x} = \sqrt{\tau}\, \mathbf{z}, \quad [3.7]$$

where $\mathbf{z} \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, \mathbf{\Sigma})$, called the speckle, and τ, called the texture, follow 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. For this distribution and when we have a dataset $\{\mathbf{x}_k\}_{k\in[1,N]}$ of i.i.d. data, the MLE of the covariance matrix is the result of the following fixed-point equation:

$$\hat{\mathbf{\Sigma}}_{tyl} = \frac{p}{N}\sum_{k=1}^{N}\frac{\mathbf{x}_k\,\mathbf{x}_k^H}{\mathbf{x}_k^H\,\hat{\mathbf{\Sigma}}_{tyl}^{-1}\,\mathbf{x}_k} \qquad [3.8]$$

In our time-series configuration, the statistical model for the pixels in a sliding window is written as:

$$\mathbf{x}_k^t \sim \sqrt{\tau_k^t}\,\mathbf{z}_k^t, \qquad [3.9]$$

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

The elliptical and CCG distributions are probability models in the scope of the robust statistical literature initiated by works such as Yohai (1974), Maronna (1976) and Martin and Pierre (2000). More details can be found in the books Maronna *et al.* (2006) and Zoubir *et al.* (2018). Following their definition, a method is said to be robust, in this chapter's context, 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.

3.4. Dissimilarity measures

3.4.1. *Problem formulation*

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

$$\begin{cases} \mathrm{H}_0 : \boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_T = \boldsymbol{\theta}_0 & (\text{no change}), \\ \mathrm{H}_1 : \exists(t,t'),\ \boldsymbol{\theta}_t \neq \boldsymbol{\theta}_{t'} & (\text{change}) \end{cases}, \qquad [3.10]$$

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:

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

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

An interesting property of such a function is 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 alarm. 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.

3.4.2. *Hypothesis testing statistics*

The first kind of dissimilarity measures comes from the statistical literature on hypothesis testing. These functions, also known as test statistics, are obtained by adapting semi-closed formulas to the problem of detection undertaken. Many of such formulas have been designed and studied in statistical literature giving an insight into the expected distribution of the resulting statistic under null hypothesis. In some cases, the resulting statistic from different techniques can yield statistically equivalent tests as shown in Ciuonzo *et al*. (2017).

Among those possibles techniques, we recall here, under the problem formulation at [3.10], the most well-known ones:

– **Generalized likelihood ratio test (GLRT):**

$$\hat{\Lambda} = \frac{\max\limits_{[\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_T]} p_{\mathbb{W}_{1,T}}\left(\mathbb{W}_{1,T} \,/\, \mathrm{H}_1; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T\right)}{\max\limits_{\boldsymbol{\theta}_0} p_{\mathbb{W}_{1,T}}\left(\mathbb{W}_{1,T} \,/\, \mathrm{H}_0; \boldsymbol{\theta}_0\right)} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda. \qquad [3.11]$$

– **Terell gradient statistic:**

$$\lambda_{\text{grad}} = \left.\frac{\partial \log p_{\mathbb{W}_{1,T}}\left(\mathbb{W}_{1,T}/\mathrm{H}_1; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T\right)}{\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} (\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_0) \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda. \qquad [3.12]$$

– **Wald statistic:**

$$\lambda_{\text{Wald}} = (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)^T \left(\left[\mathbf{I}^{-1}(\hat{\boldsymbol{\theta}}_1)\right]_{\boldsymbol{\theta}}\right)^{-1} (\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda, \qquad [3.13]$$

where $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix of the problem of estimation under the H_1 hypothesis.

3.4.2.1. *Gaussian assumption*

Under Gaussian assumption, the derivation of the GLRT has the following expression:

$$\hat{\Lambda}_{\mathrm{G}} = \frac{\left|\hat{\boldsymbol{\Sigma}}_0^{\mathrm{SCM}}\right|^{TN}}{\prod\limits_{t=1}^{T} \left|\hat{\boldsymbol{\Sigma}}_t^{\mathrm{SCM}}\right|^{N}}, \qquad [3.14]$$

where:

$$\forall t, \hat{\mathbf{\Sigma}}_t^{\mathrm{SCM}} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{x}_k^t \mathbf{x}_k^{t\,\mathrm{H}} \text{ and } \hat{\mathbf{\Sigma}}_0^{\mathrm{SCM}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\mathbf{\Sigma}}_t^{\mathrm{SCM}}. \quad [3.15]$$

This statistic is well known in the literature, and its properties have been well studied. Anderson (2003) has derived the distribution under the null hypothesis and proposed an approximation as a function of χ^2 distributions only depending on the parameters p, T and N. Therefore, this statistic has the CFAR property.

Concerning the Terell gradient statistic, it has been shown in Ciuonzo *et al*. (2017) that its form is statistically equivalent to the following reduced expression:

$$\hat{\Lambda}_{\mathrm{t_1}} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{Tr}\left[\left(\left(\hat{\mathbf{\Sigma}}_0^{\mathrm{SCM}}\right)^{-1} \hat{\mathbf{\Sigma}}_t^{\mathrm{SCM}}\right)^2\right]. \quad [3.16]$$

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

Finally, the Wald statistic yields the following result:

$$\begin{aligned}\hat{\Lambda}_{\mathrm{Wald}} =& N\sum_{t=2}^{T} \mathrm{Tr}\left[\left(\mathbf{I}_p - \hat{\mathbf{\Sigma}}_1^{\mathrm{SCM}}\left(\hat{\mathbf{\Sigma}}_t^{\mathrm{SCM}}\right)^{-1}\right)^2\right] \\ & - q\left(N\sum_{t=1}^{T}\left(\hat{\mathbf{\Sigma}}_t^{\mathrm{SCM}}\right)^{-\dagger} \otimes (\hat{\mathbf{\Sigma}}_t^{\mathrm{SCM}})^{-1}, \mathrm{vec}\left(\sum_{t=2}^{T} \mathbf{\Upsilon}_t\right)\right),\end{aligned} \quad [3.17]$$

where $\dagger$ is the transpose operator,

$$q(\mathbf{x}, \mathbf{\Sigma}) = \mathbf{x}^{\mathrm{H}}\, \mathbf{\Sigma}^{-1}\, \mathbf{x}, \quad [3.18]$$

and where

$$\mathbf{\Upsilon}_t = N\left((\hat{\mathbf{\Sigma}}_t^{\mathrm{SCM}})^{-1} - (\hat{\mathbf{\Sigma}}_t^{\mathrm{SCM}})^{-1}\hat{\mathbf{\Sigma}}_1^{\mathrm{SCM}}(\hat{\mathbf{\Sigma}}_t^{\mathrm{SCM}})^{-1}\right). \quad [3.19]$$

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.

3.4.2.2. *Non-Gaussian assumption*

Concerning non-Gaussian models, Mian *et al*. (2019a) have considered the derivation of the GLRT using the compound Gaussian model with deterministic

texture parameters:

$$\hat{\Lambda}_{\mathrm{MT}} = \frac{\left|\hat{\boldsymbol{\Sigma}}_0^{\mathrm{MT}}\right|^{TN}}{\prod_{t=1}^{T}\left|\hat{\boldsymbol{\Sigma}}_{\mathbb{W}_t}^{\mathrm{TE}}\right|^{N}} \prod_{k=1}^{N} \frac{\left(\sum_{t=1}^{T} q\left(\hat{\boldsymbol{\Sigma}}_0^{\mathrm{MT}}, \mathbf{x}_k^t\right)\right)^{Tp}}{T^{Tp}\prod_{t=1}^{T}\left(q\left(\hat{\boldsymbol{\Sigma}}_{\mathbb{W}_t}^{\mathrm{TE}}, \mathbf{x}_k^t\right)\right)^{p}}, \qquad [3.20]$$

where $\mathbb{W}_t = \{\mathbf{x}_k^t : 1 \leqslant k \leqslant N\}$ and $\hat{\boldsymbol{\Sigma}}_{\mathbb{W}_t}^{\mathrm{TE}}$ is defined in [3.8] and

$$\hat{\boldsymbol{\Sigma}}_0^{\mathrm{MT}} = \frac{p}{N}\sum_{k=1}^{N} \frac{\sum_{t=1}^{T} \mathbf{x}_k^t \left(\mathbf{x}_k^t\right)^H}{\sum_{t=1}^{T} q\left(\hat{\boldsymbol{\Sigma}}_0^{\mathrm{MT}}, \mathbf{x}_k^t\right)}, \qquad [3.21]$$

where $q(\boldsymbol{\Sigma}, \mathbf{x})$ is defined in [3.18].

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.

Concerning the derivation of Terell gradient and Wald statistics under this non-Gaussian model has not been considered yet and remains an open problem.

3.4.3. *Information-theoretic measures*

Information theory tools are based on another approach, which consists of 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.

The Kullback–Leibler (KL) divergence is a popular measure between probability density functions, encountered notably in SAR change detection problems (Inglada and Mercier 2007). It is defined as:

$$d\left(p_{\mathbf{X}}, p_{\mathbf{Y}}\right) = E\left[\log\left\{\frac{p_{\mathbf{X}}(\mathbf{x})}{p_{\mathbf{Y}}(\mathbf{x})}\right\}\right] = \int_{\mathbb{C}^p} p_{\mathbf{X}}(\mathbf{x}) \log\left\{\frac{p_{\mathbf{X}}(\mathbf{x})}{p_{\mathbf{Y}}(\mathbf{x})}\right\} d\mathbf{x} \qquad [3.22]$$

For two zero-mean multivariate Gaussian distributions, it leads to:

$$d_{\mathrm{KL}}\left(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\right) = d\left(p_{\mathbf{X}}, p_{\mathbf{Y}}\right) = \frac{1}{2}\left(\mathrm{Tr}\left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2\right) + \log\left(\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|}\right)\right) \qquad [3.23]$$

For the Gaussian case, a statistical test can be defined through the symmetrized version of the divergence (J-Divergence) between the two covariance matrix estimates:

$$\hat{\Lambda}_{\mathrm{KL}} = \frac{1}{2}\left(d_{\mathrm{KL}}\left(\hat{\boldsymbol{\Sigma}}_0^{\mathrm{SCM}}, \hat{\boldsymbol{\Sigma}}_1^{\mathrm{SCM}}\right) + d_{\mathrm{KL}}\left(\hat{\boldsymbol{\Sigma}}_1^{\mathrm{SCM}}, \hat{\boldsymbol{\Sigma}}_0^{\mathrm{SCM}}\right)\right), \qquad [3.24]$$

Frery *et al*. (2014) have shown that this statistics is asymptotically distributed as a χ^2 distribution.

Other measures include Renyi entropy or Bhattacharya distance, which have been shown to behave similarly in Frery *et al*. (2014). They will thus be omitted in this chapter. An alternative complex Hotelling–Lawley statistic has been proposed by Akbari *et al*. (2016) for the case $T = 2$. The distance used is given by:

$$\hat{\Lambda}_{\mathrm{HTL}} = \mathrm{Tr}\left(\left(\hat{\boldsymbol{\Sigma}}_0^{\mathrm{SCM}}\right)^{-1}\hat{\boldsymbol{\Sigma}}_1^{\mathrm{SCM}}\right). \qquad [3.25]$$

The authors also derived the distribution under null hypothesis and showed that it can be approximated by a Fisher–Snedecor distribution. The study showed a potential improvement of detection rate compared to the GLRT statistic.

For non-Gaussian model, Liu *et al*. (2014) proposed a measure for $T = 2$, based on the principle of mutual information. The proposed approach assumes a scenario where the pixels have been averaged on a small window of size L to reduce the speckle noise before performing the change detection (multi-look scenario). The distance, based on a gamma model on the texture parameters, is a generalization of the one proposed in Beaulieu and Touzi (2004):

$$\hat{\Lambda}_{\mathrm{MLL}} = \mathrm{MLL}(\mathbb{W}_1) + \mathrm{MLL}(\mathbb{W}_2) - \mathrm{MLL}(\mathbb{W}_{12}), \qquad [3.26]$$

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

$$\begin{aligned}\mathrm{MLL}(\mathbb{W}) &= N\,\frac{\nu + p\,L}{2}\,\left(\log(L\nu) - \log(\mu)\right) - N\,L\log\left|\boldsymbol{\Sigma}_{\mathbb{W}}^{\mathrm{TE}}\right| \\ &+ \frac{\nu - p\,L}{2}\sum_{\mathbf{x}_k^t \in \mathbb{W}}\log\left[\mathrm{Tr}\left(\left(\boldsymbol{\Sigma}_{\mathbb{W}}^{\mathrm{TE}}\right)^{-1}\mathbf{x}_k^t\,\mathbf{x}_k^{t\,\mathrm{H}}\right)\right] - N\,\log\left|\Gamma(\nu)\right| \\ &+ \sum_{\mathbf{x}_k^t \in \mathbb{W}}\log\left[K_{\nu - p\,L}\left(2\sqrt{\mathrm{Tr}\left(\left(\boldsymbol{\Sigma}_{\mathbb{W}}^{\mathrm{TE}}\right)^{-1}\mathbf{x}_k^t\,\mathbf{x}_k^{t\,\mathrm{H}}\right)\,L\,\nu/\mu}\right)\right],\end{aligned} \qquad [3.27]$$

where $\Gamma(.)$ is the Gamma function, $\mathbf{\Sigma}_{\mathbb{W}}^{\mathrm{TE}}$ is defined in equation [3.8], $K_\nu(z)$ 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

$$\mathcal{T} = \left\{ \hat{\tau}_k^t = q\left(\mathbf{\Sigma}_{\mathbb{W}}^{\mathrm{TE}}, \mathbf{x}_k^t\right)/p : \mathbf{x}_k^t \in \mathbb{W} \right\}. \quad [3.28]$$

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

3.4.4. *Riemannian geometry distances*

Riemannian geometry is an alternative concept that allows us to compare distributions. Indeed, when the parameters of a distribution lie in a Riemannian manifold (e.g. covariance matrices lie in Riemannian manifold of positive definite matrices $\mathbb{S}_{\mathbb{H}}^p$ (Skovgaard 1984)), 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 literature. Notably, for a Gaussian model, we have the following distance (Smith 2005):

$$\hat{\Lambda}_{\mathcal{RE}} = \left\| \log\left(\left(\hat{\mathbf{\Sigma}}_1^{\mathrm{SCM}}\right)^{-\frac{1}{2}} \hat{\mathbf{\Sigma}}_2^{\mathrm{SCM}} \left(\hat{\mathbf{\Sigma}}_1^{\mathrm{SCM}}\right)^{-\frac{1}{2}} \right) \right\|_F^2, \quad [3.29]$$

where $\log$ is the logarithm of matrices, which for $\mathbf{\Sigma} \in \mathbb{S}_{\mathbb{H}}^p$, having the eigenvalue decomposition $\mathbf{\Sigma} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^{\mathrm{H}}$, is defined as follows:

$$\log(\mathbf{\Sigma}) = \mathbf{U}\,\log_{\odot}(\mathbf{\Lambda})\,\mathbf{U}^{\mathrm{H}}, \quad [3.30]$$

where $\log_{\odot}$ is the pointwise logarithm.

For the elliptical case, a Riemannian CES distance can be obtained as well (Breloy *et al*. 2019):

$$\hat{\Lambda}_{\mathcal{RE}} = \alpha \sum_{i=1}^{p} \log^2 \lambda_i + \beta \left(\sum_{i=1}^{p} \log \lambda_i \right)^2, \quad [3.31]$$

where λ_i is the i-th eigenvalue of $\left(\hat{\mathbf{\Sigma}}_1^{\mathrm{MLE}}\right)^{-1} \hat{\mathbf{\Sigma}}_2^{\mathrm{MLE}}$, $\hat{\mathbf{\Sigma}}_{\epsilon\in\{0,1\}}^{\mathrm{MLE}}$ is the MLE of the scatter matrix in CES case and α, β depend on the density generator function. For a Student's t distribution, we have:

$$\alpha = \frac{d+p}{d+p+1}, \beta = \alpha - 1, \quad [3.32]$$

and $\hat{\mathbf{\Sigma}}_{\epsilon}^{\mathrm{MLE}}$ is defined in equation [3.6].

The concept of Riemannian geometry has also been considered in Ratha *et al.* (2017), 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 this chapter.

3.4.5. *Optimal transport*

Finally, a distance can be obtained through the concept of optimal transport, which defines a distance $d_W(.,.)$ depending on a cost to *displace* parts of a reference distribution into a target distribution. Let $p_{\mathbf{X}}(.)$ and $p_{\mathbf{Y}}(.)$ be two multivatiate distributions. Assuming a quadratic cost, the optimal transport problem can be stated as finding the joint distributions $p_{\mathbf{X},\mathbf{Y}}(.,.)$ to minimize:

$$d\left(p_{\mathbf{X}}, p_{\mathbf{Y}}\right) := \inf_{p_{\mathbf{X},\mathbf{Y}}(.,.)\geqslant 0} \left\{ E\left(\|\mathbf{X}-\mathbf{Y}\|^2\right) \mid \int_{\mathbf{x}} p_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})\, d\mathbf{x} = p_{\mathbf{Y}}(\mathbf{y}), \right.$$
$$\left. \int_{\mathbf{y}} p_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})\, d\mathbf{y} = p_{\mathbf{X}}(\mathbf{x}) \right\}. \quad [3.33]$$

It has been shown in Gelbrich (1990) that this distance has a closed form for Gaussian distributions, $\mathbf{X} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_1)$, $\mathbf{Y} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_2)$ and $E\left[\mathbf{X}\,\mathbf{Y}^H\right] = \mathbf{C}$. In that case, $E\left(\|\mathbf{X}-\mathbf{Y}\|^2\right) = \mathrm{Tr}\left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 - 2\,\mathbf{C}\right)$ and the problem remains to solve:

$$d\left(p_{\mathbf{X}}, p_{\mathbf{Y}}\right) = \min_{\mathbf{C}} \left\{ \mathrm{Tr}\left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 - 2\,\mathbf{C}\right) \mid \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{C} \\ \mathbf{C}^H & \boldsymbol{\Sigma}_2 \end{bmatrix} \succcurlyeq \mathbf{0} \right\}, \quad [3.34]$$

which leads to $\mathbf{C} = \boldsymbol{\Sigma}_2^{-1/2} \left(\boldsymbol{\Sigma}_2^{1/2}\, \boldsymbol{\Sigma}_1\, \boldsymbol{\Sigma}_2^{1/2}\right)^{1/2} \boldsymbol{\Sigma}_2^{1/2}$. The final result leads to:

$$d_{\mathrm{W}}\left(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2\right) = d\left(p_{\mathbf{X}}, p_{\mathbf{Y}}\right),$$
$$= \mathrm{Tr}\left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 - 2\left(\boldsymbol{\Sigma}_2^{1/2}\, \boldsymbol{\Sigma}_1\, \boldsymbol{\Sigma}_2^{1/2}\right)^{1/2}\right). \quad [3.35]$$

The corresponding optimal transport problem for two Gaussian distributions can be derived through a distance between two covariance matrices. Plugging estimates of the two covariance matrices leads to define the associated Wasserstein statistic for Gaussian hypothesis:

$$\hat{\Lambda}_{\mathcal{WG}} = d_{\mathrm{W}}\left(\hat{\boldsymbol{\Sigma}}_1^{\mathrm{SCM}}, \hat{\boldsymbol{\Sigma}}_2^{\mathrm{SCM}}\right). \quad [3.36]$$

Statistic	Reference	Model	Equation	$T > 2$	CFAR	Asymptotic null distribution
$\hat{\Lambda}_{\mathrm{G}}$	(Conradsen *et al*. 2001)	Gaussian	[3.14]	✓	✓	χ^2 (asymptotic)
$\hat{\Lambda}_{t_1}$	(Ciuonzo *et al*. 2017)	Gaussian	[3.16]	✓	✓	χ^2 (asymptotic)
$\hat{\Lambda}_{\mathrm{Wald}}$	(Ciuonzo *et al*. 2017)	Gaussian	[3.17]	✓	✓	χ^2 (asymptotic)
$\hat{\Lambda}_{\mathrm{HTL}}$	(Akbari *et al*. 2016)	Gaussian	[3.25]	×	✓	Fisher–Snedecor (approximation)
$\hat{\Lambda}_{\mathrm{MT}}$	(Mian *et al*. 2019a)	Deterministic CCG	[3.20]	✓	✓	×
$\hat{\Lambda}_{\mathrm{KL}}$	(Frery *et al*. 2014)	Gaussian	[3.24]	×	✓	χ^2 (asymptotic)
$\hat{\Lambda}_{\mathrm{MLL}}$	(Liu *et al*. 2014)	Compound Gaussian	[3.26]	×	×	×
$\hat{\Lambda}_{\mathcal{RG}}$	(Smith 2005)	Gaussian	[3.29]	×	×	×
$\hat{\Lambda}_{\mathcal{RE}}$	(Breloy *et al*. 2019)	Student's t	[3.31]	×	×	×
$\hat{\Lambda}_{\mathcal{WG}}$	(Ghaffari and Walker 2018)	Gaussian	[3.36]	×	×	×
$\hat{\Lambda}_{\mathcal{WE}}$	(Ghaffari and Walker 2018)	Elliptical	[3.37]	×	×	×

Table 3.2. *Summary of statistics with their respective properties*

Thanks to this result, we can define a statistic elliptical case by plugging estimates of the covariance matrices. A Wasserstein statistic in the elliptical can be obtained by taking:

$$\hat{\Lambda}_{\mathcal{WE}} = d_{\mathrm{W}}\left(\hat{\tau}_1\,\hat{\mathbf{\Sigma}}_1^{\mathrm{TE}}, \hat{\tau}_2\,\hat{\mathbf{\Sigma}}_2^{\mathrm{TE}}\right), \qquad [3.37]$$

where $\hat{\mathbf{\Sigma}}^{TE}_{\epsilon\in\{0,1\}}$ is defined in equation [3.21] and $\hat{\tau}_{\epsilon\in\{0,1\}} = \frac{p}{N}\sum_{k=1}^{N} q\left(\hat{\mathbf{\Sigma}}_{\epsilon}^{\mathrm{TE}}, \mathbf{x}_k^{\epsilon}\right)$.

3.4.6. *Summary*

A summary of the different statistics with their respective properties can be found in Table 3.2.

3.4.7. *Results of change detectors on the UAVSAR dataset*

We recall that the dataset and the used scenes are described in section 3.2. For Scene 1, we only use two dates (the first and fourth images) like for Scene 2. All the dates of Scene 3 are considered in the tests.

The ROC (Receiver Operating Characteristic) for all statistics and the three scenes are presented in Figure 3.7. A transcription of the detection performance at $\mathrm{P_{FA}} = 0.01$ is also given in Table 3.3.

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 scenario where the data is denoised through an averaging before change detection, which was not the case here. Thus, the statistic yields 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 for this CD problem since the performance of detection compared to others distance is lower, especially for Scene 2.

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**						

Table 3.3. *Probability of detection at a false alarm rate of 1%*

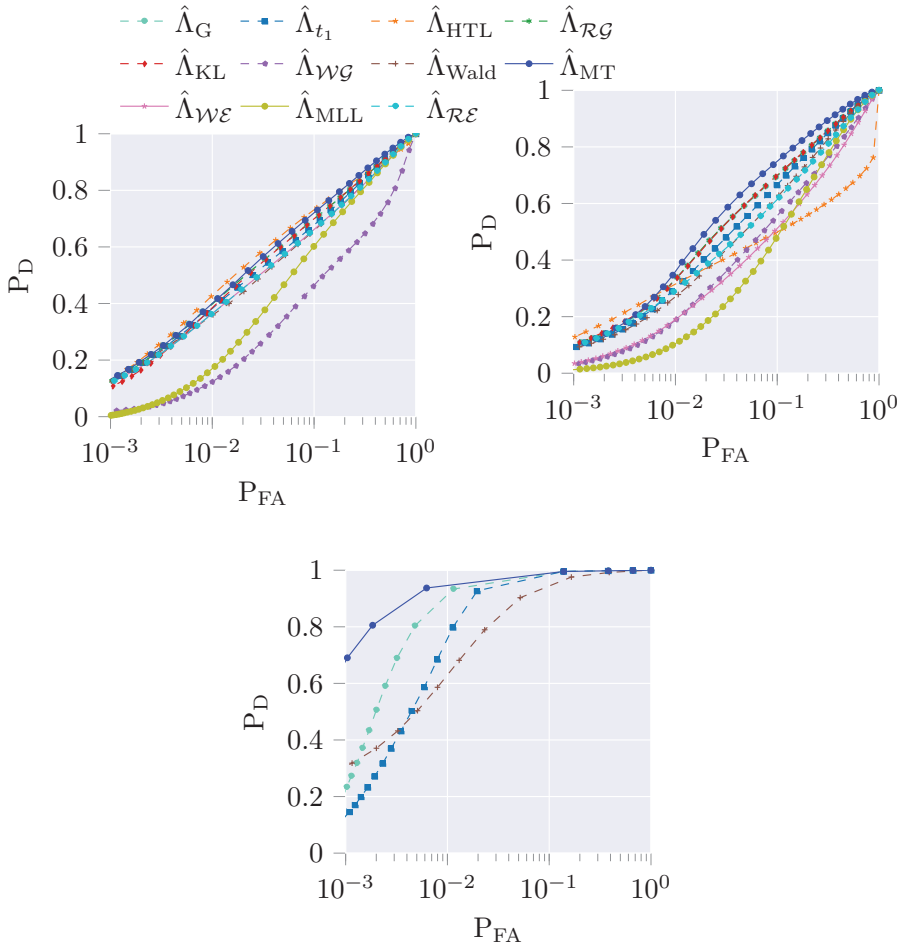


Figure 3.7. *ROC plots for a window size of* 5×5 *for the three scenes. Top: Scene 1; middle: Scene 2; bottom: Scene 3. For a color version of this figure, see www.iste.co.uk/atto/change1.zip*

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.

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

Table 3.4. *Time consumption in seconds*

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 the Gaussian one is explained by the heterogeneity of the images at the transition between objects. As an example, the outputs of both $\hat{\Lambda}_{\mathrm{G}}$ and $\hat{\Lambda}_{\mathrm{MT}}$ are presented in Figure 3.8. 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's 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, Student's t model can be inaccurate to model the actual data. In those regards, $\hat{\Lambda}_{\mathrm{MT}}$ is able to achieve better performance without relying on a specific elliptical model and has thus a better robustness to mismatch scenarios.

To compare the attractiveness of the different statistics described in this section with regard to computational complexity, we consider an experimental Monte-Carlo analysis on synthetic Gaussian data.

A set of data $\mathbb{W}$ is generated with parameter values $p = 10$, $N = 25$ and $T = 2$. The Gaussian data is generated with a covariance matrix $(\mathbf{\Sigma})_{ij} = 0.5^{|i-j|}$ and $4,000$ Monte-Carlo trials have been considered. 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 execution times are provided in Table 3.4.

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 Kronecker product, 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 corresponds to the Scipy implementation, which can be improved with regard 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 regard to time consumption is $\hat{\Lambda}_{\mathrm{MT}}$ since it requires fewer operations than the others. Moreover, some solutions based on stochastic Riemannian gradient (Zhou and Said 2019) should allow us to reduce the computational cost of the covariance estimator used in $\hat{\Lambda}_{\mathrm{MT}}$.

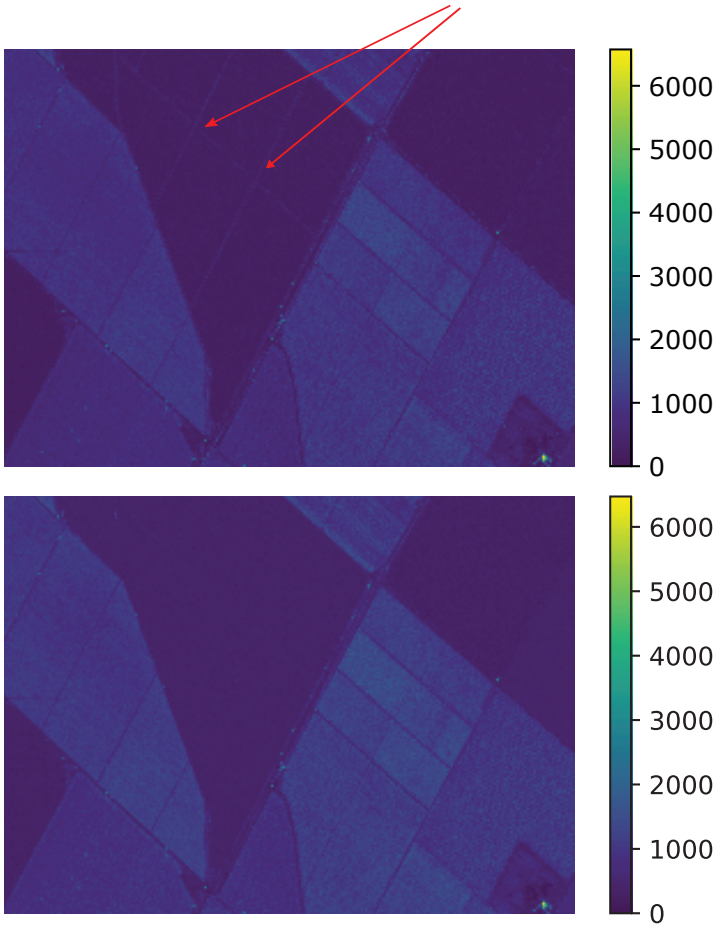


Figure 3.8. *Output of $\hat{\Lambda}_{\mathrm{G}}$ (top) and $\hat{\Lambda}_{\mathrm{MT}}$ (bottom) on Scene 3. The major difference between the two statistics is highlighted by red arrows. For a color version of this figure, see www.iste.co.uk/atto/change1.zip*

3.5. Change detection based on structured covariances

In Mian *et al*. (2017), we have shown that the performance of the detectors (Gaussian and non-Gaussian) improves with the increase of the diversity (here the spectro-angular diversity obtained with the method proposed in Mian *et al*. (2019b)) and so on p. When p is increased, the number of data to estimate the covariance matrix has to be increased too. The relation $N = 2p$ (obtained in array processing (Reed *et al*. 1974)) is often used to choose this number. But in our configuration, it can be difficult to handle with large N. First, the hypothesis of i.i.d. data to estimate the covariance is not fulfilled if N is too large, and second, we loose resolution when N increases (because of the method used in Mian *et al*. (2019b) to highlight the spectro-angular diversity). Therefore, if we want to choose a large p to obtain good performance, we have to find a method to keep a value of N reasonable.

It exists two kinds of approaches to reduce the number of N needed to the estimation of the covariance matrix: the first one is to shrink the covariance towards to the identity matrix (Ledoit and Wolf 2004; Ollila and Tyler 2014; Pascal *et al*. 2014), and the second one is to resort on the inherent structure of data. A common structure for RADAR data is that they lie in a low-dimensional subspace. In this case, the covariance matrix has a low-rank (LR) structure, i.e. it can be written as the sum of an LR matrix and a diagonal matrix (where we will assume that all eigenvalues are equal).

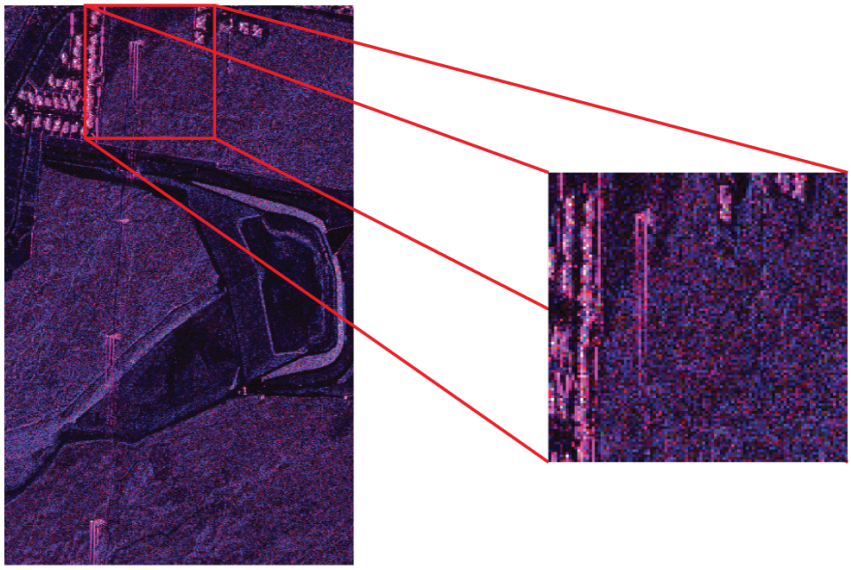
We illustrate this property on the UAVSAR data already tested in the previous sections. In Figure 3.9, we have computed and plotted the mean of the eigenvalues of the selected area of Scene 2. In this simulation, the data size is equal to $p = 12$ with four spectro-angular diversities and three polarimetric channels. We note that the eigenvalues decrease significantly, and we conclude that the data lie in a low-dimensional subspace. In the following, the rank R is assumed to be known or pre-estimated (either globally or locally) by using rank estimation methods from the literature, such as information theory-based ones (Stoica and Selen 2004).

The detectors derived in the previous sections assume unstructured covariance matrices. We propose in this section to derive Gaussian and non-Gaussian detectors when the covariance matrix of the data has an LR structure. We only focus on the GLRT approach even it is of course possible to extend the other approaches presented in the previous section 3.4.

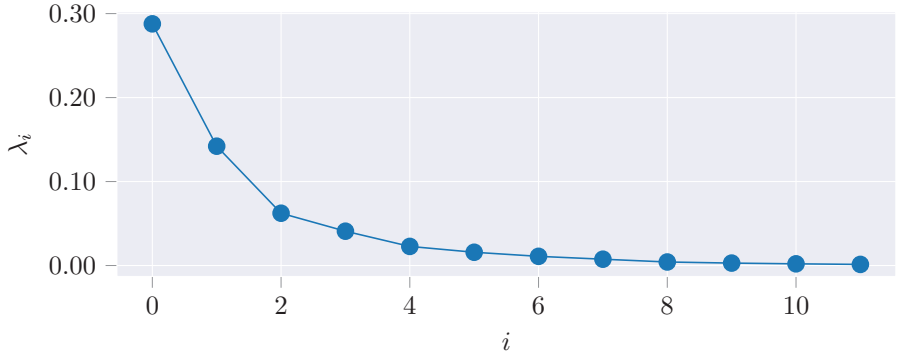
3.5.1. *Low-rank Gaussian change detector*

We recall that the dataset $\{\mathbf{x}_k\}$, where $\mathbf{x}_k \sim \mathcal{CN}(\mathbf{0}_p, \mathbf{\Sigma})$, has the following likelihood function

$$\mathcal{L}_{\mathcal{G}}\left(\{\mathbf{x}_k\}_{k=1}^{N} \mid \mathbf{\Sigma}\right) \propto \prod_{k=1}^{K} |\mathbf{\Sigma}|^{-1} e^{(} - \mathbf{x}_k^H \, \mathbf{\Sigma}^{-1} \, \mathbf{x}_k) \qquad [3.38]$$



(a) UAVSAR data (Courtesy NASA/JPL-Caltech). Selection of the area on scene 2.



(b) Spectrum of UAVSAR data (wavelets+polarimetry). $p = 12$.

Figure 3.9. *Low-rank properties of UAVSAR data with polarimetric and spectro-angular diversity. For a color version of this figure, see www.iste.co.uk/atto/change1.zip*

Since $\boldsymbol{\Sigma}$ has an LR structure, it is written as follows:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_R + \sigma^2 \mathbf{I} \qquad [3.39]$$

where $\boldsymbol{\Sigma}_R$ belongs to the set of Hermitian positive semi-definite matrices of rank R, denoted $\mathcal{H}^+_{p,R}$. A Gaussian GLRT that accounts for this prior knowledge denoted

$\hat{\Lambda}_{\mathrm{LRG}}$ (LRG: low-rank Gaussian) can be formulated according to [3.10] with the following distribution parameters:

$$\begin{array}{ll} \text{Model:} & \mathbf{x}_k^t \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}_R^t + \sigma_t^2\,\mathbf{I}\right) \\ & \text{Likelihood in [3.38]} \\ \text{Param.:} & \mathrm{H}_0 : \boldsymbol{\theta}_0 = \left\{\mathbf{\Sigma}_R^0, \sigma_t^2\right\} \\ & \mathrm{H}_1 : \{\boldsymbol{\theta}_t\}_{t=1}^T = \left\{\mathbf{\Sigma}_R^t, \sigma_t^2\right\}_{t=1}^T \end{array} \quad [3.40]$$

The corresponding detector when we assume that the white Gaussian noise level is known has been derived in Abdallah *et al.* (2019). To derive the corresponding GLRT, we have $\boldsymbol{\theta}_t = \left\{\mathbf{\Sigma}_R^t, \sigma_t^2\right\}$ & $\mathbf{\Phi}_t = \varnothing$. Finally, we obtain for the expression of $\hat{\Lambda}_{\mathrm{LRG}}$:

$$\hat{\Lambda}_{\mathrm{LRG}} = \frac{\mathcal{L}\left(\left\{\{\mathbf{x}_k^t\}_{k=1}^N\right\}_{t=1}^T / \mathrm{H}_1; \mathcal{T}_R\left\{\hat{\mathbf{\Sigma}}_1\right\}, \ldots, \mathcal{T}_R\left\{\hat{\mathbf{\Sigma}}_T\right\}\right)}{\mathcal{L}\left(\left\{\{\mathbf{x}_k^t\}_{k=1}^N\right\}_{t=1}^T / \mathrm{H}_0; \mathcal{T}_R\left\{\hat{\mathbf{\Sigma}}_0\right\}\right)} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda. \quad [3.41]$$

where $\hat{\mathbf{\Sigma}}_t = \dfrac{1}{N}\sum_{k=1}^{N} \mathbf{x}_k^t \left(\mathbf{x}_k^t\right)^H$. From $\mathbf{\Sigma} \overset{\mathrm{EVD}}{=} \mathbf{U}\,\mathrm{diag}\left(\mathbf{d}\right)\,\mathbf{U}^H$, we obtain $\mathcal{T}_R\left\{\mathbf{\Sigma}\right\}$:

$$\mathcal{T}_R\left\{\mathbf{\Sigma}\right\} = \mathbf{U}\,\mathrm{diag}\left(\tilde{\mathbf{d}}\right)\,\mathbf{U}^H$$

where $\tilde{\mathbf{d}} = \left[d_1, \ldots, d_R, \hat{\sigma}_t^2, \ldots, \hat{\sigma}_t^2\right]$ and $\hat{\sigma}_t^2 = \displaystyle\sum_{r=R+1}^{p} \frac{d_r}{p-R}$.

In the next section, we make the same derivation when the data follow the compound Gaussian model.

3.5.2. *Low-rank compound Gaussian change detector*

We recall that the dataset $\{\mathbf{x}_k\}$, where $\mathbf{x}_k \mid \tau_k \sim \mathcal{CN}(\mathbf{0}_p, \tau_k\mathbf{\Sigma})$, has the following likelihood function

$$\mathcal{L}_{\mathcal{CG}}\left(\{\mathbf{x}_k\}_{k=1}^N \left|\tau_k\,\mathbf{\Sigma}\right.\right) \propto \prod_{k=1}^{K} \left|\tau_k\,\mathbf{\Sigma}\right|^{-1}\, e^{(}-\mathbf{x}_k^H\left(\tau_k\,\mathbf{\Sigma}\right)^{-1}\mathbf{x}_k) \quad [3.42]$$

where like for the Gaussian version, $\mathbf{\Sigma} = \mathbf{\Sigma}_R + \sigma^2\,\mathbf{I}$.

The corresponding GLRT for CD, denoted $\hat{\Lambda}_{\mathrm{LRCG}}$ (LRCG: low-rank compound Gaussian), corresponds to equation [3.10] with the following distribution/parameters:

$$\begin{aligned}
&\text{Model: } \mathbf{x}_k^t \sim \mathcal{CN}\left(\mathbf{0}_p, \tau_k^t \left(\boldsymbol{\Sigma}_R^t + \sigma_t^2 \mathbf{I}\right)\right) \\
&\qquad \text{Likelihood in [3.42]} \\
&\text{Param.: } \mathrm{H}_0 : \boldsymbol{\theta}_0 = \left\{\boldsymbol{\Sigma}_R^0, \sigma_0^2, \left\{\tau_k^0\right\}_{k=1}^N\right\} \\
&\qquad \mathrm{H}_1 : \{\boldsymbol{\theta}_t\}_{t=1}^T = \left\{\boldsymbol{\Sigma}_R^t, \sigma_t^2, \{\tau_k^t\}_{k=1}^N\right\}_{t=1}^T
\end{aligned} \tag{3.43}$$

Here, the test accounts for a possible change of both the covariance matrix and the textures between acquisitions, as it was shown to be the most relevant approach for SAR-ITS in the previous section 3.4 and in Mian *et al*. (2019a).

Evaluating Λ_{LRCG} according to the generic equation [3.11] requires us to compute:

$$\hat{\Lambda}_{\mathrm{LRCG}} = \frac{\mathcal{L}_{LRCG}^{\mathrm{H}_1}\left(\left\{\{\mathbf{x}_k^t\}_{k=1}^N\right\}_{t=1}^T \mid \hat{\boldsymbol{\theta}}_{\mathrm{LRCG}}^{\mathrm{H}_1}\right)}{\mathcal{L}_{LRCG}^{\mathrm{H}_0}\left(\left\{\{\mathbf{x}_k^t\}_{k=1}^N\right\}_{t=1}^T \mid \hat{\boldsymbol{\theta}}_{\mathrm{LRCG}}^{\mathrm{H}_0}\right)}, \tag{3.44}$$

where $\mathcal{L}_{LRCG}^{\mathrm{H}_0}$ and $\mathcal{L}_{LRCG}^{\mathrm{H}_1}$ are the likelihood (derived from equation [3.42]) under H_0 and H_1, respectively, and where

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{\mathrm{LRCG}}^{\mathrm{H}_0} &= \left\{\hat{\boldsymbol{\Sigma}}_R^0, \hat{\sigma}_0^2, \left\{\hat{\tau}_k^0\right\}_{k=1}^N\right\}, \\
\hat{\boldsymbol{\theta}}_{\mathrm{LRCG}}^{\mathrm{H}_1} &= \left\{\hat{\boldsymbol{\Sigma}}_R^t, \hat{\sigma}_t^2, \{\hat{\tau}_k^t\}_{k=1}^N\right\}_{t=1}^T
\end{aligned} \tag{3.45}$$

are the maximum likelihood estimators under H_0 and H_1, respectively. As for the Gaussian version, we do not have any closed form for this CD.

In the following, we present the method to obtain $\hat{\boldsymbol{\theta}}_{\mathrm{LRCG}}^{\mathrm{H}_1}$. The approach to have $\hat{\boldsymbol{\theta}}_{\mathrm{LRCG}}^{\mathrm{H}_0}$ are very close and only the differences will be given.

3.5.2.1. *Maximum likelihood under* H_1

Under H_1, the likelihood is separable in t. For a fixed t, the maximum likelihood estimation of the unknown parameters $\left\{\hat{\boldsymbol{\Sigma}}_R^t, \hat{\sigma}_t^2, \{\hat{\tau}_k^t\}_{k=1}^N\right\}$ consists of solving the problem

$$\begin{aligned}
\underset{\left\{\tau_k^t, \boldsymbol{\Sigma}_k^t\right\}_{k=1}^N, \boldsymbol{\Sigma}_R^t, \sigma_t^2}{\text{maximize}} \quad & \sum_{k=1}^N \log\left(\mathcal{L}_{\mathrm{H}_1}^t\left(\mathbf{x}_k^t \mid \boldsymbol{\Sigma}_k^t\right)\right) \\
\text{subject to} \quad & \boldsymbol{\Sigma}_k^t = \tau_k^t \left(\boldsymbol{\Sigma}_R^t + \sigma_t^2 \mathbf{I}\right) \\
& \boldsymbol{\Sigma}_R^t \succcurlyeq \mathbf{0}, \text{ and } \operatorname{rank}(\boldsymbol{\Sigma}_R^t) = R \\
& \sigma_t^2 > 0, \text{ and } \tau_k^t > 0, \forall k
\end{aligned} \tag{3.46}$$

Algorithm 3.1.– MLE under H_1 and fixed t

Input: R, t, $\{\mathbf{x}_k^t\}_{k=1}^N$
while *(convergence criterion not met)* **do**
  Update textures with equation [3.47] Update covariance matrix parameters with equations [3.49]–[3.50]
return $\hat{\mathbf{\Sigma}}_R^t, \hat{\sigma}_t^2, \{\hat{\tau}_k^t\}_{k=1}^N$

where $\mathcal{L}_{\mathrm{H}_1}^t\left(\mathbf{x}_k^t|\mathbf{\Sigma}_k\right)$ reads directly from equation [3.42]. The solution to this problem cannot be obtained in closed form, but the following sections derive practical block-coordinate descent algorithm in order to evaluate it. The algorithm is summed up in the box of Algorithm 3.1.

– Update of the textures (H_1): Assuming a fixed covariance matrix $\mathbf{\Sigma}_R^t + \sigma_t^2\,\mathbf{I}$, the maximum likelihood estimator of the texture parameters is obtained in closed form (Pascal *et al*. 2008), with

$$\hat{\tau}_k^t = \frac{1}{p}\left(\mathbf{x}_k^t\right)^H \left(\mathbf{\Sigma}_R^t + \sigma_t^2\,\mathbf{I}\right)^{-1}\mathbf{x}_k^t \qquad [3.47]$$

– Update of the covariance matrix parameters (H_1): Assuming fixed textures $\{\tau_k^t\}_{k=1}^N$, equation [3.46] can be re-expressed as:

$$\begin{array}{ll}\underset{\mathbf{\Sigma}_t,\mathbf{\Sigma}_R^t,\sigma_t^2}{\text{minimize}} & \log|\mathbf{\Sigma}_t| + \mathrm{Tr}\left\{\tilde{\mathbf{S}}_t\,\mathbf{\Sigma}_t^{-1}\right\} \\ \text{subject to} & \mathbf{\Sigma}_t = \mathbf{\Sigma}_R^t + \sigma_t^2\,\mathbf{I} \\ & \mathbf{\Sigma}_R^t \succcurlyeq \mathbf{0},\ \text{and rank}\left(\mathbf{\Sigma}_R^t\right) = R \\ & \sigma_t^2 > 0\end{array} \qquad [3.48]$$

with $\tilde{\mathbf{S}}_t = \dfrac{1}{N}\displaystyle\sum_{k=1}^{N}\frac{\mathbf{x}_k^t\left(\mathbf{x}_k^t\right)^H}{\tau_k^t}$. The solution is given in Tipping and Bishop (1999) and leads to the update

$$\mathbf{\Sigma}_t = \mathbf{U}\,\mathrm{diag}\left(\tilde{\mathbf{d}}\right)\mathbf{U}^H \triangleq \mathcal{T}_{\mathrm{R}}\left(\tilde{\mathbf{S}}_t\right) \qquad [3.49]$$

defined through the operator $\mathcal{T}_{\mathrm{R}}$, with

$$\begin{array}{l}\tilde{\mathbf{S}}_t \overset{\mathrm{EVD}}{=} \mathbf{U}\,\mathrm{diag}\left(\mathbf{d}\right)\mathbf{U}^H \\ \mathbf{d} = \left[d_1,\ldots,d_R,d_{R+1},\ldots,d_p\right] \\ \tilde{\mathbf{d}} = \left[d_1,\ldots,d_R,\hat{\sigma}_t^2,\ldots,\hat{\sigma}_t^2\right] \\ \hat{\sigma}_t^2 = \dfrac{1}{p-R}\displaystyle\sum_{r=R+1}^{p} d_r.\end{array} \qquad [3.50]$$

Algorithm 3.2.– MLE under H_0

Input: R, $\left\{\{\mathbf{x}_k^t\}_{k=1}^N\right\}_{t=1}^T$
while *(convergence criterion not met)* **do**
  Update textures with equation [3.52] Update covariance matrix parameters with equations [3.49]–[3.50] applied to equation [3.53]
return $\hat{\mathbf{\Sigma}}_R^0, \hat{\sigma}_0^2, \{\hat{\tau}_k^0\}_{k=1}^N$

3.5.2.2. *Maximum likelihood under* H_0

$$\begin{array}{cl} \underset{\{\tau_k^0, \mathbf{\Sigma}_k^0\}_{k=1}^N, \mathbf{\Sigma}_R^0, \sigma_0^2}{\text{maximize}} & \displaystyle\sum_{t=1}^{T}\sum_{k=1}^{N} \log\left(\mathcal{L}_{\mathrm{H}_0}^t\left(\mathbf{x}_k^t \mid \mathbf{\Sigma}_k^0\right)\right) \\ \text{subject to} & \mathbf{\Sigma}_k^0 = \tau_k^0\left(\mathbf{\Sigma}_R^0 + \sigma_0^2\,\mathbf{I}\right) \\ & \mathbf{\Sigma}_R^0 \succcurlyeq \mathbf{0}, \text{ and } \operatorname{rank}\left(\mathbf{\Sigma}_R^0\right) = R \\ & \sigma_0^2 > 0, \text{ and } \tau_k^0 > 0, \forall k \end{array} \quad [3.51]$$

This problem can be solved as the one in equation [3.46], with some modifications due to the likelihood function. The differences are detailed below and summed up in the box of Algorithm 3.2.

– Update of the textures (H_0): Assuming a fixed covariance matrix $\mathbf{\Sigma}^0 = \mathbf{\Sigma}_R^0 + \sigma_0^2\,\mathbf{I}$, the maximum likelihood estimator of the texture parameters is obtained in closed form (Mian *et al.* 2019a), with

$$\hat{\tau}_k^0 = \frac{1}{T\,p}\sum_{t=1}^{T}\left(\mathbf{x}_k^t\right)^H\left(\mathbf{\Sigma}_R^0 + \sigma_0^2\,\mathbf{I}\right)^{-1}\mathbf{x}_k^t \quad [3.52]$$

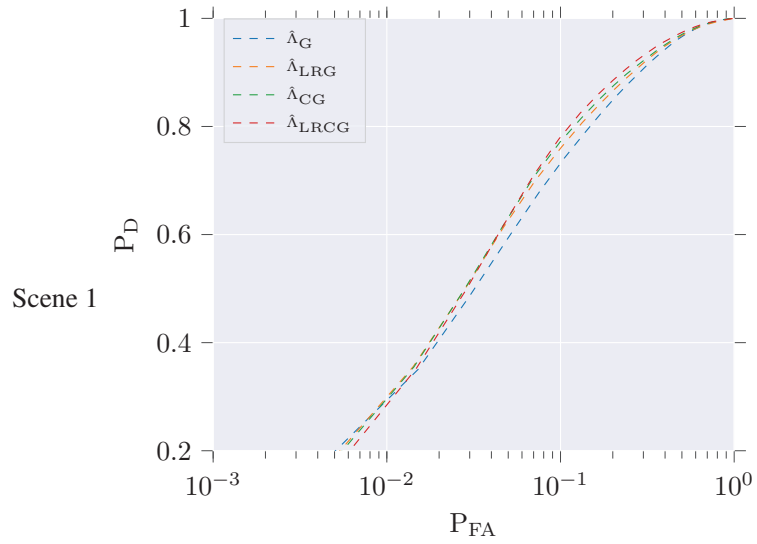
– Update of the covariance matrix parameters (H_0): Assuming fixed textures, the update problem can be reformulated as in [3.48], using

$$\tilde{\mathbf{S}}_0 = \frac{1}{N\,T}\sum_{t=1}^{T}\sum_{k=1}^{N}\frac{\mathbf{x}_k^t\left(\mathbf{x}_k^t\right)^H}{\tau_k^0} \quad [3.53]$$

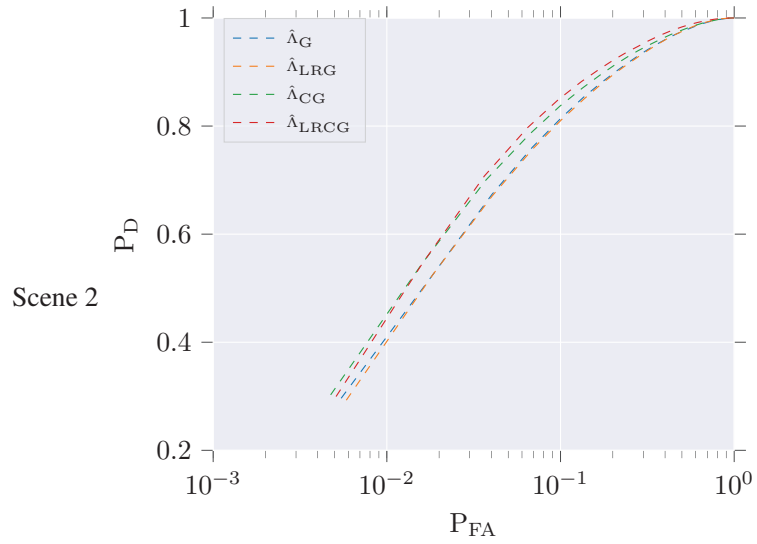
instead of $\tilde{\mathbf{S}}_t$. The solution is then obtained as in equations [3.49] and [3.50], and yields $\mathbf{\Sigma}_0 \triangleq \mathcal{T}_{\mathrm{R}}(\tilde{\mathbf{S}}_0)$.

3.5.3. *Results of low-rank change detectors on the UAVSAR dataset*

We recall that the dataset and the used scenes are described in section 3.2. In this experimentation, Scene 1 with four dates and Scene 2 are tested.



(a)



(b)

Figure 3.10. *Comparison between four methods: Gaussian, low-rank Gaussian, compound Gaussian (CG) and low-rank compound Gaussian (LRCG).* $p = 12$*, rank is fixed as 3, the window size is* 7×7 *and* σ^2 *is assumed unknown for both low-rank models. For a color version of this figure, see www.iste.co.uk/atto/change1.zip*

The performance of the change detectors Λ_{LRG} and Λ_{LRCG} are tested on a SAR ITS dataset and assessed with ROC curves (displaying the probability of detection versus the false alarm rate). As a mean to assess the effectiveness of using the LR structure, they are compared to the following detectors: i) the classic Gaussian detector proposed in Conradsen *et al.* (2003) and Ciuonzo *et al.* (2017) and given in [3.14], denoted $\hat{\Lambda}_G$; ii) the compound Gaussian detector proposed in Mian *et al.* (2019a) and given in [3.20], denoted $\hat{\Lambda}_{CG}$.

The rank R is chosen according to Figure 3.9, which has displayed the eigenvalues of the total sample covariance matrix. For this dataset, $R = 3$ appears to be an interesting value to separate signal from noise components. Notably, this rank gathers 81% of the total variance.

Figure 3.10 displays the ROC curves for the four detectors applied to Scenes 1 and 2. It assesses that the proposed method achieves the best performance in terms of probability of detection versus false alarm rate.

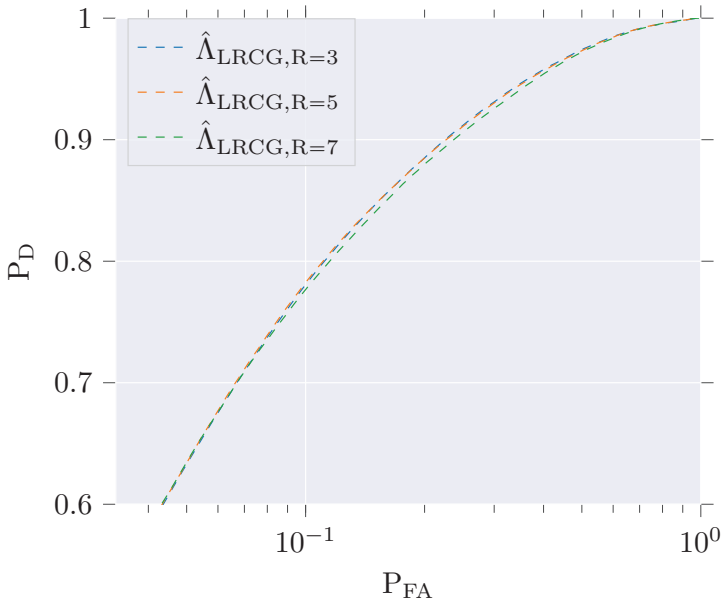


Figure 3.11. *Comparison for Scene 1 of the LRCG (low-rank compound Gaussian) for different rank values. $p = 12$, the window size fixed at 7×7 and σ^2 assumed unknown. For a color version of this figure, see www.iste.co.uk/atto/change1.zip*

Figure 3.11 displays the ROC curves of Λ_{LRCG} on Scene 1 for three different values of rank R. It is interesting to note that these curves do not vary significantly with respect to this parameter. Therefore, we can expect that a slight error in the rank estimation will not lead to a significant drop in CD performance.

3.6. Conclusion

The local covariance matrix of pixel patches appears to be relevant feature to analyze multivariate image time series (especially when dealing with SAR images). In this scope, this chapter presented an overview of covariance-based change detection algorithms.

Most of the standard methods were initially driven from the Gaussian distribution assumption and can be divided into two groups. The first one consists of two-step approaches, where plug-in estimates of the local covariance matrices are used to evaluate a distance, assessing for the dissimilarity (i.e. change) between the patches. Multiple theoretical frameworks motivated the use of various matrix distances (each with their own merits), notably information theory, Riemannian geometry and optimal transport. The second approach consists of one-step procedures, based on the derivation of a test statistic suited to the binary hypothesis test ("change" or "no change" in the local covariance matrix). The most common statistic for this problem is the GLRT, but many other exist, such as the t_1 and the Wald statistics.

When dealing with high-resolution SAR images, it is well known that the Gaussian assumption is no longer valid. Conversely, the compound Gaussian modeling was shown to be more suited to reflect the empirical distribution of the data. This family of distributions notably drove robust covariance estimation methods, leading to various matrix estimates (notably, M-estimators) that can be used in plug-in approaches. Moreover, compound Gaussian distributions can serve as a basis to derive robust test statistics such as the GLRT. Empirical experiments were conducted on various UAVSAR datasets, which illustrated the practical interest of the compound Gaussian modeling. However, it is to be noted that the gain in detection performance is obtained at the cost of a slight increase in computational complexity.

In some cases, this spectro-angular diversity of the scatterers (obtained through wavelet-transform of the polarimetric data) can be useful to enhance the change detection performance. This transformation generally increases the size of the data ($p > 3$), which implies a need for larger patches in order to correctly estimate the local covariance matrix. This dimensionality issue can be mitigated by exploiting some structure exhibited by the data, such as a low-rank one observed in the covariance matrix. This chapter concluded by presenting a generalization of the GLRT approaches accounting for this structure and illustrating the interest of the approach.

3.7. References

Abdallah, R.B., Mian, A., Breloy, A., Korso, M.N.E., Lautru, D. (2019). Detection methods based on structured covariance matrices for multivariate SAR images processing. *IEEE Geoscience and Remote Sensing Letters*, 16(7), 1160–1164.

Achim, A., Tsakalides, P., Bezerianos, A. (2003). SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 41(8), 1773–1784.

Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A., Zoppetti, C. (2013). Nonparametric change detection in multitemporal SAR images based on mean-shift clustering. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4), 2022–2031.

Akbari, V., Anfinsen, S.N., Doulgeris, A.P., Eltoft, T., Moser, G., Serpico, S. B. (2016). Polarimetric SAR change detection with the complex Hotelling–Lawley trace statistic. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7), 3953–3966.

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

Beaulieu, J. and Touzi, R. (2004). Segmentation of textured polarimetric SAR scenes by likelihood approximation. *IEEE Transactions on Geoscience and Remote Sensing*, 42(10), 2063–2072.

Breloy, A., Ginolhac, G., Renaux, A., Bouchard, F. (2019). Intrinsic Cramér–Rao bounds for scatter and shape matrices estimation in CES distributions. *IEEE Signal Processing Letters*, 26(2), 262–266.

Bruzzone, L. and Prieto, D.F. (2000). Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3), 1171–1182.

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

Ciuonzo, D., Carotenuto, V., Maio, A.D. (2017). On multiple covariance equality testing with application to SAR change detection. *IEEE Transactions on Signal Processing*, 65(19), 5078–5091.

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

Conradsen, K., Nielsen, A.A., Schou, J., Skriver, H. (2003). A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(1), 4–19.

Foucher, S., Lopez-Martinez, C. (2014). Analysis, evaluation, and comparison of polarimetric SAR speckle filtering techniques. *IEEE Transactions on Image Processing*, 23(4), 1751–1764.

Freitas, C.C., Frery, A.C., Correia, A.H. (2005). The polarimetric G distribution for SAR data analysis. *Environmetrics*, 16(1), 13–31. https://onlinelibrary.wiley.com/doi/abs/10.1002/env.658.

Frery, A.C., Nascimento, A.D.C., Cintra, R.J. (2014). Analytic expressions for stochastic distances between relaxed complex Wishart distributions. *IEEE Transactions on Geoscience and Remote Sensing*, 52, 1213–1226.

Gao, G. (2010). Statistical modeling of SAR images: A survey. *Sensors*, 10(1), 775–795.

Gelbrich, M. (1990). On a formula for the L2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten*, 147(1), 185–203.

Ghaffari, N. and Walker, S. (2018). On multivariate optimal transportation. arXiv preprint arXiv:1801.03516.

Gong, M., Zhao, J., Liu, J., Miao, Q., Jiao, L. (2016). Change detection in synthetic aperture radar images based on deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 27(1), 125–138.

Greco, M.S. and Gini, F. (2007). Statistical analysis of high-resolution SAR ground clutter data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3), 566–575.

Hecheltjen, A., Thonfeld, F., Menz, G. (2014). *Recent Advances in Remote Sensing Change Detection – A Review*. Springer Netherlands, Dordrecht, pp. 145–178.

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

Inglada, J. and Mercier, G. (2007). 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*, 45(5), 1432–1445.

Kervrann, C. and Boulanger, J. (2006). Optimal spatial adaptation for patch-based image denoising. *IEEE Transactions on Image Processing*, 15(10), 2866–2878.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88, 365–411.

Liu, M., Zhang, H., Wang, C., Wu, F. (2014). Change detection of multilook polarimetric SAR images using heterogeneous clutter models. *IEEE Transactions on Geoscience and Remote Sensing*, 52(12), 7483–7494.

Maronna, R.A. (1976). Robust *M*-estimators of multivariate location and scatter. *Annals of Statistics*, 4(1), 51–67.

Maronna, R.A., Martin, D.R., Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*. Wiley Series in Probability and Statistics, 1st edition, Wiley.

Martin, B. and Pierre, D. (2000). Robust estimation of the SUR model. *Canadian Journal of Statistics*, 28(2), 277–288. https://onlinelibrary.wiley.com/doi/abs/10.2307/3315978.

Mian, A., Ovarlez, J.-P., Ginolhac, G., Atto, A.M. (2017). Multivariate change detection on high resolution monovariate SAR image using linear time-frequency analysis. *2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 1942–1946.

Mian, A., Ginolhac, G., Ovarlez, J.-P., Atto, A.M. (2019a). New robust statistics for change detection in time series of multivariate SAR images. *IEEE Transactions on Signal Processing*, 67(2), 520–534.

Mian, A., Ovarlez, J.-P., Atto, A.M., Ginolhac, G. (2019b). Design of new wavelet packets adapted to high-resolution SAR images with an application to target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6), 3919–3932.

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

Nascimento, A.D.C., Frery, A.C., Cintra, R.J. (2019). Detecting changes in fully polarimetric SAR imagery with statistical information theory. *IEEE Transactions on Geoscience and Remote Sensing*, 57(3), 1380–1392.

Ollila, E. and Tyler, D. (2014). Regularized m -estimators of scatter matrix. *IEEE Transactions on Signal Processing*, 62(22), 6059–6070.

Ollila, E., Tyler, D.E., Koivunen, V., Poor, H.V. (2012a). Complex elliptically symmetric distributions: Survey, new results and applications. *IEEE Transactions on Signal Processing*, 60(11), 5597–5625.

Ollila, E., Tyler, D.E., Koivunen, V., Poor, H.V. (2012b). Compound-Gaussian clutter modeling with an inverse Gaussian texture distribution. *IEEE Signal Processing Letters*, 19(12), 876–879.

Pascal, F., Chitour, Y., Ovarlez, J.-P., Forster, P., Larzabal, P. (2008). Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis. *IEEE Transactions on Signal Processing*, 56(1), 34–48.

Pascal, F., Chitour, Y., Quek, Y. (2014). Generalized robust shrinkage estimator and its application to STAP detection problem. *IEEE Transactions on Signal Processing*, 62(21), 5640–5651.

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

Ratha, D., De, S., Celik, T., Bhattacharya, A. (2017). Change detection in polarimetric SAR images using a geodesic distance between scattering mechanisms. *IEEE Geoscience and Remote Sensing Letters*, 14(7), 1066–1070.

Reed, I., Mallett, J., Brennan, L. (1974). Rapid convergence rate in adaptive arrays. *IEEE Transactions on Aerospace and Electronic Systems*, AES-10(6), 853–863.

AQ2 Skovgaard, L.T. (1984). A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, 11, pp. 211–223.

Smith, S.T. (2005). Covariance, subspace, and intrinsic Cramér–Rao bounds. *IEEE Transactions on Signal Processing*, 53(5), 1610–1630.

Stoica, P. and Selen, Y. (2004). Model-order selection: A review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4), 36–47.

Tipping, M. and Bishop, C. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3), 611–622.

Wang, F., Wu, Y., Zhang, Q., Zhang, P., Li, M., Lu, Y. (2013). Unsupervised change detection on SAR images using triplet Markov field model. *IEEE Geoscience and Remote Sensing Letters*, 10(4), 697–701.

Yohai, V.J. (1974). Robust estimation in the linear model. *Annals of Statististics*, 2(3), 562–567. https://doi.org/10.1214/aos/1176342717.

Yousif, O. and Ban, Y. (2013). Improving urban change detection from multitemporal SAR images using PCA-NLM. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4), 2032–2041.

Yueh, S., Kong, J., Jao, J., Shin, R., Novak, L. (1989). K-distribution and polarimetric terrain radar clutter. *Journal of Electromagnetic Waves and Applications*, 3(8), 747–768. https://doi.org/10.1163/156939389X00412.

Zhou, J. and Said, S. (2019). Fast, asymptotically efficient, recursive estimation in a Riemannian manifold. *Entropy*, 21(10). AQ3

Zoubir, A.M., Koivunen, V., Ollila, E., Muma, M. (2018). *Robust Statistics for Signal Processing*. Cambridge University Press.