ASPECTS OF MULTIVARIATE STATISTICAL THEORY WITH THE APPLICATION TO CHANGE DETECTION

Esra Erten[1,2], *Andreas Reigber*[1] *and Olaf Hellwich*[2]

[1]Microwaves and Radar Institute, German Aerospace Centre (DLR), 82234 Oberpfaffenhofen, Germany.
[2]Computer Vision and Remote Sensing, Technical University of Berlin, 10587, Berlin, Germany.

1. ABSTRACT

This paper proposes a new method for change detection measurement independent from system configuration in a set of multi-temporal-multidimensional SAR images. The method is based on the *Kullback-Leibler* (KL-divergence) test, known as *Mutual Information*. In order to develop an algorithm independent from the system configuration, firstly the joint distribution of PolInSAR data set, based on the second order statistics has been derived. Such a derivation accounts for the whole multi-temporal system configurations as interferometric and partial-PolInSAR data sets. Then the KL-divergence test is used to measure the difference between the joint density of multi-temporal PolSAR data set and their marginal density known as complex *Wishart distribution*. A comparison between the proposed and the other well-known change detection (e.g. cross correlation and the maximum likelihood ratio test) techniques is shown, describing the advantages due to the fact that the proposed change detector involve almost every facet of applied change detection.

2. INTRODUCTION

The experimental results presented in [1], [2] and [3] encourage to use of the approach of KL-divergence test with the aim of detecting changes in temporal SAR images. [1] is not direct application of the KL-divergence test, it is the implementation of bivariate Gamma distribution into image registration and change detection. The KL-divergence test has been used to overcome the problem of estimating the correlation between two temporal intensity images. However, the technique proposed in [2] evaluates the local statistics of interferometric temporal data set through KL-divergence test. Instead of using fixed PDF, cumulant-based series expansion which approximates the PDF from samples has been used to calculate the KL-divergence scalar. *Morio et al* [3] did a very interesting work by analyzing the KL-divergence test for contrast measurement for multi-channel SAR data set. They show that the KL-divergence test for contrast measurement allows to precisely characterize the contribution of each channel for different system configurations, including intensity, polarimetric, and interferometric images. Here, the idea of change detection with the use of the KL-divergence is to extend the previous applications into the second order statistics of temporal data sets.

The section 3 introduces the joint distribution of multi-temporal polarimetric data set. The next section will define the KL-divergence test measuring the difference between two probability distributions and its implementation to temporal polarimetric dataset to detect the change, while it also reviews the other well-known polarimetric change detection techniques to make a comparison between the proposed technique and the well-known technique regarding accuracy assessment.

3. THE JOINT DISTRIBUTION OF TEMPORAL POLARIMETRIC SAR IMAGES

Let the temporal target vector $k = [k_1\ k_2]^T$ be a complex target vector distributed as a multicomponent circular Gaussian $\mathcal{N}^C(0,\Sigma)$ that consist of two target vectors $k_1 \sim \mathcal{N}^C(0,\Sigma_{11})$ and $k_2 \sim \mathcal{N}^C(0,\Sigma_{22})$ obtained from multi-temporal SAR images at time t_1 and t_2. Thus, these two observations are a correlated or uncorrelated process over time depending on the monitored objects. To not to make any assumption concerning their independence, Σ may be used to characterize the behavior of the temporal multi-channel data. The number of elements in one of the target vector k_i at the time t_i is represented by m, and hence the temporal target vector k has the dimension of $q = 2 \times m$.

Σ as well as its estimation n samples covariance matrix $A = \frac{1}{n}\sum_{j=1}^{n} k_j k_j^\dagger$ can be portioned as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{1}$$

which summaries the whole (joint and marginal) information from temporal multi-channel SAR systems. The matrices A_{11} and A_{22} are the standard n samples (n-look in the case of SAR images) $m \times m$ covariance matrices from separate temporal images. $A_{12} = A_{21}^{\dagger}$ is a $m \times m$ cross correlation matrix between k_1 and k_2. The properties of the hermitian matrix A permit to define the joint and the conditional probability of the temporal images. The joint density of element A_{22} conditioned on A_{11} follows the complex Wishart distribution with $n-m$ degrees of freedom $p(A_{11}|A_{22}) = \mathcal{W}^C(n-m, \Sigma_{11|22})$ [4] where $\Sigma_{11|22} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is independent from A_{12} and A_{22}. Moreover, as the sample covariance matrix of A_{22} follows a complex Wishart density function with n degrees of freedom $\mathcal{W}^C(n, \Sigma_{22})$, the conditional probability density of A_{12} given A_{22} is a complex normal function $p(A_{12}|A_{22}) = \mathcal{N}^{\mathcal{C}}(\Sigma_{12}\Sigma_{22}^{-1}A_{22}, \Sigma_{11|22} \otimes A_{22})$ [5]. Using that any linear transformation of a normal vector has a normal distribution and multiplying $(p(A_{12}|A_{22}))$ with $A_{12}^{-1/2}$ results in $p(A_{12}A_{12}^{-1}|A_{22}) = \mathcal{N}^{\mathcal{C}}(\Sigma_{12}\Sigma_{22}^{-1}A_{22}^{1/2}, \Sigma_{11|22} \otimes I_m)$. Then, re-formulating $A_{12}A_{22}^{-1}A_{21}$ as $A_{12}A_{22}^{-1/2}(A_{12}A_{22}^{-1/2})^{\dagger}$, we can write $p(A_{12}A_{22}^{-1}A_{21}|A_{22})$ as a complex *Non-central Wishart* distribution with the help of [6, Definition II]:

$$p(A_{12}A_{22}^{-1}A_{21}|A_{22}) = \mathcal{W}^{\mathcal{C}}(m, \Sigma_{11|22}, \Sigma_{12}\Sigma_{22}^{-1}A_{22}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11|22}). \tag{2}$$

Let $A_{12}A_{22}^{-1}A_{21} = \mathcal{D}$, then the density function of $p(A_{11|22}, A_{22}, \mathcal{D}|A_{22})$ can be written as

$$\begin{aligned} p(A_{11|22}, A_{22}, \mathcal{D}|A_{22}) &= p(A_{11|22})p(A_{22})p(\mathcal{D}|A_{22}) \\ &= \frac{n^{m(n-m)}|A_{11|22}|^{n-2m}\exp(-n\mathrm{tr}\Sigma_{11|22}^{-1}A_{11|22})}{|\Sigma_{11|22}|^{n-m}\tilde{\Gamma}_m(n-m)} \times {}_0\tilde{F}_1(m, n^2\Sigma_{11|22}^{-1}\Sigma_{12}\Sigma_{22}^{-1}A_{22}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11|22}^{-1}D) \\ &\times \frac{\exp(-n\mathrm{tr}\Sigma_{11|22}^{-1}D)exp(-n\mathrm{tr}\Sigma_{11|22}^{-1}\Sigma_{12}\Sigma_{22}^{-1}A_{22}\Sigma_{22}^{-1}\Sigma_{21})}{|\Sigma_{11|22}|^m\tilde{\Gamma}_m(m)} \times \frac{n^{mn}|A_{22}|^{n-m}\exp(-n\mathrm{tr}\Sigma_{22}^{-1}A_{22})}{|\Sigma_{22}|^n\tilde{\Gamma}_m(n)}. \end{aligned} \tag{3}$$

Here, ${}_0\tilde{F}_1(n, M)$ is the *complex hypergeometric function* of matrix M and closely related to *Bessel* functions. This function can be calculated with the help of positive eigenvalues of the Hermitian matrix M by [7].

Substituting $A_{11} = A_{11|22} + \mathcal{D}$, $R^2 = A_{11}^{-1/2}A_{12}A_{22}^{-1}A_{11}^{-1/2}$ into (3) and applying the change of variable [4, Teorem 2.1.5], $R^2 = A_{11}^{-1/2}\mathcal{D}A_{11}^{-1/2} \Longrightarrow dR^2 = |A_{11}|^m\mathrm{d}\mathcal{D}$, we can write $p(A_{11}, A_{22}, R^2)$ as

$$\begin{aligned} p(A_{11}, A_{22}, R^2) = &\frac{n^{-qn}|A_{22}|^{n-m}\mathrm{etr}(-n\Sigma_{22|11}^{-1}A_{22})}{|\Sigma_{22}|^n\tilde{\Gamma}_m(n)\tilde{\Gamma}_m(n-m)} \frac{|I-R^2|^{n-q}|A_{11}|^{n-m}\exp(-n\mathrm{tr}\Sigma_{11|22}^{-1}A_{11})}{|\Sigma_{11}|^n|I_m-P^2|^n\tilde{\Gamma}_m(m)} \\ &\times {}_0\tilde{F}_1(m, n^2A_{11}^{1/2}\Sigma_{11|22}^{-1}\Sigma_{12}\Sigma_{22}^{-1}A_{22}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11|22}^{-1}A_{11}^{1/2}R^2) \end{aligned} \tag{4}$$

where $P^2 = \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{11}^{-1/2}$. It is clear that (4) is valid under the condition of $A_{11}, A_{22} > 0$ and $0 < R^2 < I_m$ which means that A_{11}, A_{22}, R^2 and $I_m - R^2$ are positive definitive matrix[1]. Integrating R^2 over (4), the joint distribution of temporal multi-channel SAR systems is given by

$$p(\mathbf{A}_{11}, \mathbf{A}_{22}) = \mathrm{e}^{\left(\frac{n\mathbf{\Sigma}_{22}^{-1}\mathbf{A}_{22}+\mathbf{\Sigma}_{11}^{-1}\mathbf{A}_{11}}{-(\mathbf{I}-\mathbf{P}^2)}\right)} \frac{n^{2mn}|\mathbf{\Sigma}_{11}\mathbf{\Sigma}_{22}|^{-n}|\mathbf{A}_{11}\mathbf{A}_{22}|^{n-m}}{|I-P^2|^n\tilde{\Gamma}_m(n)\tilde{\Gamma}_m(n)} {}_0\tilde{F}_1\left(n, n^2\mathbf{A}_{11}^{1/2}\mathbf{\Sigma}_{11.2}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{A}_{22}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\Sigma_{11.2}^{-1}\mathbf{A}_{11}^{1/2}\right). \tag{5}$$

Fig.1 plots the comparison between the theoretical bivariate distribution (5) and 3D bivariate histogram from Monte Carlo simulations in the case of $m = 1$. The 3D bivariate histogram represents the occurrence as a series of 3D bars. It can be considered to be a conjunction of two simple (i.e., univariate) histograms, combined such that the frequencies of co-occurrences of values on the two analyzed variables can be examined.

4. KL-DIVERGENCE AND CHANGE DETECTION

In probability theory and information theory, the *KL-divergence* test is a non-commutative measure of the difference between two probability distributions $p_X(x)$ and $p_Y(y)$ of the random variables X and Y, respectively. KL-divergence from Y to X is given by

$$\begin{aligned} D_{KL} &= \int log\left(\frac{p_X(x)}{p_Y(y)}\right)p_X(x)\mathrm{d}x \\ &= \mathcal{H}\left(p_X(x), p_Y(y)\right) \end{aligned} \tag{6}$$

[1]The real case of this distribution can be found in [8] and the key theorems in the complex case, concerning the derivation of $p(A_{11}, A_{22})$, can be found in [9].

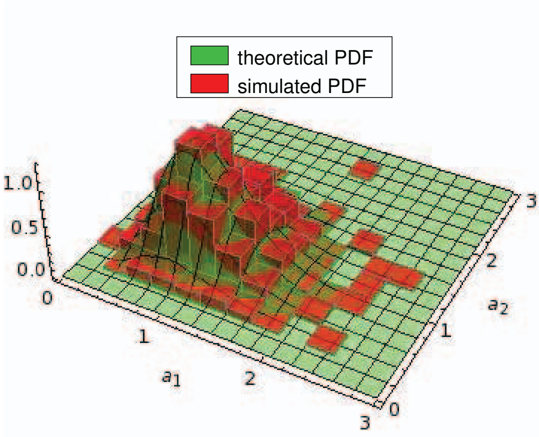


Fig. 1. Comparison between theoretical bivariate distribution (5) and 3D bivariate histogram from simulated data in the case of $m = 1$. In both case, the powers are given by $\sigma_1 = \sigma_2 = 1$ and correlation by $\rho = 0.1$. When $n \to \infty$, $a_1 = \sigma_1 = 1$ and $a_2 = \sigma_2 = 1$.

where $\mathcal{H}(p_X(x), p_Y(y))$ is called the relative entropy. It can be seen from (6) that the KL-divergence test is based on a relative entropy of continuous random variables, which does not change under invertible transformations.

Therefore, the KL-divergence can be interpreted as the measure of the discrimination between the hypothesis $\mathcal{H}_X$ and $\mathcal{H}_Y$, if Hypothesis $\mathcal{H}_X$ is associated with the PDF $p_X(x)$ and $\mathcal{H}_Y$ with that of $p_Y(y)$.

To investigate the temporal behavior of multidimensional SAR systems including whole the system configurations, i.e., $m = 1$ interferometric pair and $m = 3$ polarimetric-interferometric pair, the joint density functions of polarimetric multi-temporal images derived in (5) and the statistical similarity measurement of two densities explained in (6) will be combined. The basic concept behind the definition of the KL-divergence test is to define a convenient scalar change detection parameter which is independent of the number of channel and to extend the previous applications into the second order statistics of multi-temporal data set. Since $p(A_{11}, A_{22})$ characterize the joint behavior of the second order statistics of the temporal data set, $p(A_{11}, A_{22}) = p(A_{11})p(A_{22})$ if and only if the multi-temporal data set are independent. Due to this, the KL-divergence scalar with the aim of change detection can be written as

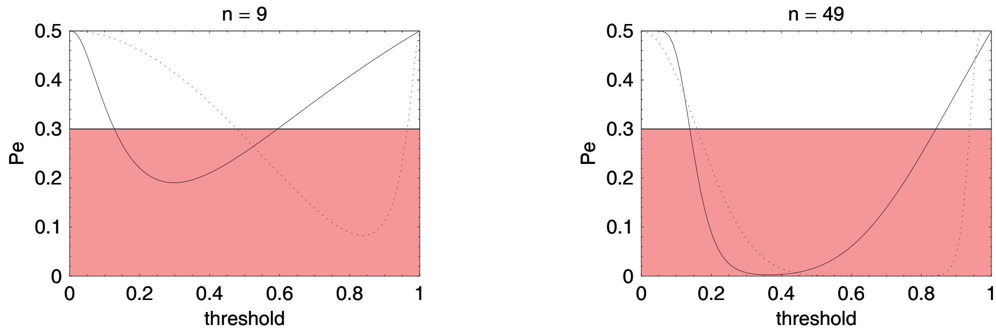
$$\mathcal{D}_n = \int log\left[\frac{p(A_{11}, A_{22})}{p(A_{11})p(A_{22})}\right] p(A_{11}, A_{22}) d\overrightarrow{A} \tag{7}$$

where $p(\mathbf{A}_{11})$ and $p(\mathbf{A}_{22})$ are marginal densities of the m^2 complex element vector obtained by stacking the columns of $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ respectively and A is the vector including $2m^2$ elements obtained by stacking the A_{11} and A_{22} consecutively. Here, n indicates the number of samples using in the estimation of the covariance matrices. Substituting the joint density (5) and the marginal densities known as Wishart distribution into (7) and writing $\Sigma_{11.2} = \Sigma_{11}^{-1}(I_m - P^2)$ and $\Sigma_{22.1} = \Sigma_{22}^{-1}(I_m - P^2)$ imply the following result

$$\mathcal{D}_n = E\left\{\log\left({}_0\tilde{F}_1(n, n^2 A_{11}^{1/2}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}A_{22}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}A_{11}^{1/2})\right)\right\} - \mathrm{tr}\left(\frac{2nP^2}{I_m - P^2}\right) - n\log(|I_m - P^2|) \tag{8}$$

At this point, it can be seen from (8) that the KL-divergence decision statistic $\mathcal{D}_n$ and the cross correlation R are related to each other. Thus the final goal of this change detection study is to highlight the advantage of using the KL-divergence compared to correlation. We will evaluate the performance of KL-divergence detector with correlation in two levels of samples $n = 9$ and $n = 49$. In particular, the average *Probability of Errors* (PE) analysis will be performed using the *Receiver Operating Characteristic* (ROC) curves. The averaged PE is a function of threshold $\mathcal{T}$ including the sum of the rejection of a null hypothesis $(1 - PD)$ and the fail of rejecting a alternative hypothesis (PFA), i.e., $PE = (1/2)(1 - PD + PFA)$.

Assuming that we would like to make the change detection with the condition of $PE \leq 0.3$ where $PE = 0.5$ is the highest error. In that case it is clearly seen from Fig.2 that there are more thresholds values supplying this condition by KL-divergence instead of correlation paremater. Fig.2 summaries that improvements for detection problem are obtainable by increasing the number of samples. Similar results have been obtained in [1] with the system configuration $m = 1$. Note that Pe in the below curves would change if H_0 and H_1 were chosen to be some other values.



(a) $m = 3$ (b) $m = 3$

Fig. 2. The plots of probability of errors versus threshold for KL-divergence (the dashed line) and correlation (the solid line) between second order statistics estimators at two level of samples, $n = 9$ and $n = 49$. It has to be noted that the correlation in the case of $m = 1$ is not the interferometric coherence, it is the correlation between two intensity images.

5. CONCLUSION

A new joint distribution and a change detection decision statistic based on the second order statistic has been proposed for multi-channel temporal SAR Images. A new algorithm for change detection, which is based on the KL-divergence test, is independent of system configuration. The proposed detector has been compared to the classical change detector and has been shown to have a more robust behavior than the classical algorithms based on simulated and real data.

6. REFERENCES

[1] F. Chatelain, J. Y. Tourneret, J. Inglada, and A. Ferrari, "Bivariate gamma distributions for image registration and change detection," *IEEE Trans. on Image Processing*, vol. 16, no. 7, pp. 1796–1806, Jul. 2007.

[2] J. Inglada, G. Mercier, and T. CNES, "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 Part 2, pp. 1432–1445, 2007.

[3] J. Morio, P. Réfrégier, F. Goudail, P. C. Dubois-Fernandez, and X. Dupuis, "Information Theory-Based Approach for Contrast Analysis in Polarimetric and/or Interferometric SAR Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2185–2196, 2008.

[4] R. J. Muirhead, *Aspects of Multivariate Statistic Theory*, 1982.

[5] L. Ferro-Famil, E. Pottier, and J. S. Lee, "Unsupervised classification of multifrequency and fullypolarimetric SAR images based on the H/A/Alpha-Wishart classifier," *Geoscience and Remote Sensing, IEEE Transaction on*, vol. 39, no. 11, pp. 2332–2342, Nov. 2001.

[6] MR McKay and IB Collings, "Statistical properties of complex noncentral wishart matrices and MIMO capacity," in *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, 2005, pp. 785–789.

[7] P. J. Smith and L. M. Garth, "Distribution and characteristic functions for correlated complex Wishart matrices," *Journal of Multivariate Analysis*, vol. 98, no. 4, pp. 661–677, 2007.

[8] G. Iliopoulos, "UMVU estimation of the ratio of powers of normal generalized variances under correlation," *Journal of Multivariate Analysis*, vol. 99, no. 6, pp. 1051–1069, 2008.

[9] E. Erten, *Information Theory Based Multi-Channel SAR Image Analysis*, Ph.D. thesis, Technische Universitt́ Berlin, not published yet.