



Contributions to SAR Image Time Series Analysis

Ammar Mian

► To cite this version:

| Ammar Mian. Contributions to SAR Image Time Series Analysis. Signal and Image processing. Université Paris-Saclay, 2019. English. NNT : 2019SACL074 . tel-02464840

HAL Id: tel-02464840

<https://tel.archives-ouvertes.fr/tel-02464840>

Submitted on 3 Feb 2020

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

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

Contributions to SAR Image Time Series Analysis

Thèse de doctorat de l'Université Paris-Saclay
préparée à Centrale-Supélec

École doctorale n°580 : Sciences et Technologies de l'Information
et de la Communication (STIC)
Spécialité de doctorat : Traitement du signal et des images

Thèse présentée et soutenue à Gif-sur-Yvette le 26 septembre 2019 par

Ammar Mian

Composition du jury :

André Ferrari	
Professeur des Universités, Université de Nice (Laboratoire J.L. Lagrange)	Rapporteur
Jean-Yves Tourneret	
Professeur, INP-ENSEIHT	Rapporteur
Maria Sabrina Greco	
Professeur des Universités, Università di Pisa (Dipartimento di Ingegneria dell'Informazione)	Examinateur
Frédéric Pascal	
Professeur, CentraleSupélec (Laboratoire des Signaux et Systèmes)	Président
Abdourahmane M. Atto	
Maître de Conférences, Université Savoie Mont Blanc (Laboratoire LISTIC)	Encadrant
Guillaume Ginolhac	
Professeur des Universités, Université Savoie Mont Blanc (Laboratoire LISTIC)	Co-Directeur de thèse
Jean-Philippe Ovarlez	
Maître de Recherche 2, ONERA Palaiseau (DEMR) et CentraleSupélec (SONDRA)	Directeur de thèse

YI<RFMF\$NF<M

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisors who helped me and encouraged me through those whole 3 years. I could not have imagined having better advisors and mentors for my Ph.D study. Guillaume, despite the geographic distance, you were very engaged in the advancement of my work, taught me a lot about science and helped me achieve the results I have obtained. I also thank you and Dominique for your care during my several trips to Annecy. Jean-Philippe, I thank you for the kindness you have shown me during the thesis. Your patience, motivation, and immense knowledge have been a huge help in all the time of research and writing of this thesis. Abdou, I thank you for your kindness and your great insight that you have taught me about scientific writing and the importance of taking your time.

I am grateful to my Ph.D defense committee who accepted to evaluate my research work. I thank Jean-Yves Tourneret and Andre Ferrari for carefully reading my dissertation and providing me with corrections in order to improve it. I thank again Maria Greco for coming all the way to Saclay for the defense and Frederic Pascal for the very pertinent remarks about the work I have presented.

I have spent three years at the SONDRA laboratory where I had the chance to meet and interact with wonderful people. I first thank Virginie Bouvier who was very kind to me and helped me so much despite all the work I made her do. I am deeply grateful to Chengfang who has helped tremendously with technical aspects more than once. I have been blessed to have shared my time with Thierry, Israel, Laeticia, Regis. I also thank my fellow Ph.D and post-docs: the team Pedro, Uyhour, Orian, Eugénie and Samir who helped me feel welcome when I first arrived. Then Bruno, Thibault, Giovanni, Marie-José, Dihia, Augustin who came after to brighten the office. I also thank the three directors of SONDRA for the opportunities they have provided me with: Marc, Sylvain, and Stéphane as well as Chin and Uma from Singapore. My thanks to Emmanuel from LISTIC who has been very nice to me when I came on several occasions.

I thank my fellow researchers from the Signal Processing community with whom I had the chance to collaborate: Arnaud, Lucien, Alex, Nabil, Florent, and Jialun. My sincere thanks to Eric, Joana, Jordi, Gordana, Gilles, Virginie, Rémi, Yannick, Salem, Barbara, Charles, and Clémence.

Finally, I would like to express my sincere gratitude to my family who has supported me through this Ph.D. To my brothers Shamas, Zain and my sister Madhia: thank you. I would like to thank my niece Mahnoor who has learned much more than me in those past three years and I hope you'll continue to learn many things when you grow up. My words could not describe the amount of gratitude I have towards my parents who have brought me to France when I was a child in the hope that I achieve a better education. You have sacrificed a lot in order to raise us and that is why this manuscript is dedicated to you, with love.

INTRODUCTION

Remote sensing data from Synthetic Aperture Radar (SAR) sensors offer a unique opportunity to record, to analyze, to predict the evolution of the Earth. SAR are moving radar systems capable of producing high-quality images of the Earth's surface. They consist in emitting an electromagnetic wave which is reflected on the Earth surface. The backscattered signal from the scatterers illuminated surface is then processed to build an image of the scene. SAR systems are known for their usability in all weather, illumination conditions. They are thus capable to monitor all kinds of areas of interest. Recent years have seen an increase in the number of SAR systems.

In the last decade, numerous satellite remote sensing missions have been launched (Sentinel-1, UAVSAR, TerraSAR X, etc.). This resulted in a dramatic improvement in the Earth image acquisition capability, accessibility. The growing number of observation systems allows now to build high temporal/spatial-resolution Earth surface images datasets.

This new scenario significantly raises the interest in time-series processing to monitor changes occurring over large areas. On the other hand, developing new algorithms to process such a huge volume of data represents a current challenge. Notably, the modern trend of deep-learning approaches shows its limits since most of this data is not annotated, corrupted by a problematic speckle noise (inherent to SAR images).

In this context, the present thesis aims at developing methodologies for change detection in high-resolution time series of images. These series raise some notable challenges that we will tackle in this thesis.

- Standard statistical methods rely on multivariate data corresponding to a diversity of some kind in order to infer a result which is often superior to a one of a monovariate approach. When it concerns SAR images, the data is often released as monovariate in which case, there is no diversity to exploit. In this case, if the resolution of the images is high enough, it is possible to build multivariate images corresponding to a physical diversity using time frequency analysis. We will thus develop wavelet tools to tackle this issue.
- The improvement in terms of resolution obtained from the latest generation of sensors comes with an increased heterogeneity of the data obtained. For this setup, the standard Gaussian assumption used to develop change detection methodologies is no longer valid. In order to tackle this issue, we consider the family of elliptical distributions which accurately describe the data, develop new change detection methodologies.
- In order to further analyse the image time series, it is of interest to estimate the time at which these changes occur in the time series. For this problem, an iterative procedure based on robust detection schemes will be proposed.

Since this methodology relies on the principle of the spatial sliding windows, the number of samples used is directly linked to the spatial resolution of detection. In

order to analyze the impact of the window size, we will consider the development of a lower bound on the error of estimated changes.

- We will finally explore, as an opening of this work, the potential use of Riemannian geometry to analyze changes in the image time series.

By tackling these issues, the present work has allowed to achieve promising results in order to analyse changes in time series of high resolution SAR images.

List of publications:

Journals :

- **A. Mian**, G. Ginolhac, J. Ovarlez and A. M. Atto, "New Robust Statistics for Change Detection in Time Series of Multivariate SAR Images", in IEEE Transactions on Signal Processing, vol. 67, no. 2, pp. 520-534, 15 Jan, 2019.
- **A. Mian**, J. Ovarlez, A. M. Atto and G. Ginolhac, "Design of New Wavelet Packets Adapted to High-Resolution SAR Images With an Application to Target Detection," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 6, pp. 3919-3932, June 2019.
- R. Ben Abdallah, **A. Mian**, A. Breloy, A. Taylor, M. N. El Korso and D. Lautru, "Detection Methods Based on Structured Covariance Matrices for Multivariate SAR Images Processing," in IEEE Geoscience and Remote Sensing Letters.

Conferences proceedings :

- **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," 2017 25th European Signal Processing Conference (EUSIPCO), Kos, 2017, pp. 1942-1946.
- **A. Mian**, J. P. Ovarlez, G. Ginolhac, A. M. Atto, Détection de changement sur images SAR monovariées par analyse temps-fréquence linéaire, GRETSI 2017 XXVIème colloque, 2017, Juan-les-Pins.
- **A. Mian**, J. Ovarlez, G. Ginolhac and A. M. Atto, "A Robust Change Detector for Highly Heterogeneous Multivariate Images," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, 2018, pp. 3429-3433.
- **A. Mian**, J. Ovarlez, G. Ginolhac and A. M. Atto, "Robust Detection and Estimation of Change-Points in a Time Series of Multivariate Images," 2018 26th European Signal Processing Conference (EUSIPCO), Rome, 2018, pp. 1097-1101.
- **A. Mian**, L. Bacharach, G. Ginolhac, A. Renaux, M. N. El Korso and J. -. Ovarlez, "Designing Sar Images Change-point Estimation Strategies Using an Mse Lower Bound," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 5312-5316.
- **A. Mian**, A.Breloy, G. Ginolhac, J.-P. Ovarlez, Robust Low-rank Change Detection for SAR Image Time Series, 2019 International Geoscience, Remote Sensing Symposium (IGARSS), Yokohama, Japan, 2019.

CONTENTS

Acknowledgements	5
Introduction	7
Contents	9
List of Figures	12
List of Tables	17
Nomenclature	19
Acronyms	21
1 SAR Image Time Series issues	23
1.1 Specificities of SAR data	24
1.2 The principle of diversity: exploiting multivariate data	25
1.2.1 Interest of multivariate data	25
1.2.2 Sources of diversity in SAR images	26
1.2.3 Spectro-angular diversity using wavelet analysis of SAR images	28
1.3 Models considered for the distribution of multivariate SAR images	30
1.3.1 The data	30
1.3.2 The Gaussian model	31
1.3.3 Non-Gaussianity of high-resolution images	31
1.3.4 The Complex Elliptical model	32
1.3.5 The deterministic compound-Gaussian model	35
1.4 Change detection in SAR image time series	36
1.4.1 An overview of change detection techniques for SAR images	36
1.4.2 Defining the problem of covariance equality testing	40
1.4.3 On the hypothesis problem and test statistics	40
1.4.4 On the GLRT technique and known alternatives	42
1.4.5 A note on the false discovery rate and familywise error rate	43
1.4.6 Results on the equality of covariance in Gaussian context	44
1.4.7 An extension to the Elliptical case using local Normal Asymptotic theory	46
1.4.8 Some attempts of equality testing under deterministic compound-Gaussian model	48
1.5 Change-point estimation in time series	49
1.5.1 Some techniques for estimating the change-point	50
1.5.2 Joint detection and estimation of change-points for SAR images . . .	51
1.6 Summary	53

2 Wavelet packets for SAR analysis	55
2.1 Reminding the principle of SAR image reconstruction	56
2.2 Adapting Shannon wavelet packets to SAR geometry	57
2.2.1 Shannon M-band wavelets theory	58
2.2.2 Adaptation of Shannon wavelets to SAR geometry	61
2.3 Design of a new wavelet packets to reduce sidelobes	63
2.4 Application to real SAR images	68
2.4.1 Datasets considered	68
2.4.2 Analysis of high-resolution SAR images decomposition	68
2.4.3 Preliminary results in target detection applications	68
2.5 Conclusions	79
2.A Appendix	80
2.A.1 Proof of proposition 2.2.1 at p. 59	80
2.A.2 Proof of proposition 2.2.2 at p. 62	80
2.A.3 Proof of proposition 2.2.3 at p. 62	81
2.A.4 Proof of proposition 2.2.4 at p. 63	81
2.A.5 Proof of proposition 2.3.2 at p. 65	81
3 Robust Change Detection	83
3.1 Description of the problem	84
3.1.1 Definitions	84
3.1.2 Defining the problem of covariance equality testing	84
3.2 Adapting robust M-estimation and Tyler's theory to covariance testing	85
3.2.1 GLRT under Elliptical model	86
3.2.2 GLRT under CAE model	86
3.2.3 GLRT under deterministic compound-Gaussian model	88
3.2.4 Convergence considerations	90
3.2.5 Statistical properties	93
3.2.6 Performance study of new statistics	97
3.3 Application of robust statistics to SAR change detection	101
3.3.1 Data description	101
3.3.2 First analysis	103
3.4 Conclusions	108
3.A Appendix	109
3.A.1 Proof of proposition 3.2.1 at p. 86	109
3.A.2 Proof of proposition 3.2.2 at p. 88	111
3.A.3 Proof of proposition 3.2.3 at p. 89	111
3.A.4 Proof of proposition 3.2.4 at p. 89	115
3.A.5 Proof of proposition 3.2.5 at p. 90	117
3.A.6 Proof of proposition 3.2.6 at p. 91	120
3.A.7 Proof of proposition 3.2.7 at p. 91	121
3.A.8 Proof of proposition 3.2.8 at p. 92	122
3.A.9 Proof of proposition 3.2.10 at p. 95	124
3.A.10 Proof of proposition 3.2.12 at p. 95	125
3.A.11 Proof of proposition 3.2.13 at p. 96	125
3.A.12 Proof of proposition 3.2.14 at p. 96	127
3.B Detection Methods Based on Structured Covariance Matrices for Multivariate SAR Images Processing	127
3.C Robust Low-Rank Change Detection for SAR Image Time Series	133

4 Change-point detection and estimation	139
4.1 Derivation of statistics under robust model adapted to this strategy	140
4.2 Simulation on synthetic dataset	141
4.2.1 Results on synthetic time-series	141
4.2.2 Results on synthetic images	144
4.2.3 Application to SAR image time series	149
4.3 A Bayesian MSE Lower-Bound for the tuning	152
4.3.1 Data model	153
4.3.2 Covariance inequality	154
4.3.3 Hybrid bound for the change-point model	155
4.3.4 Derivation of $\mathbf{F}(\Sigma)$, $\rho(s)$ and $\phi_{\sigma_j, \ell}(s)$	155
4.3.5 Computation of the tightest bound	156
4.3.6 Simulations	157
4.4 Conclusions	159
4.A Appendix	161
4.A.1 Proof of proposition 4.1.1 at p. 140	161
4.A.2 Proof of proposition 4.1.2 at p. 141	164
4.A.3 Proof of proposition 4.3.4 at p. 156	166
5 Riemannian geometry	169
5.1 Some elements of Riemannian geometry	171
5.1.1 Riemannian geometry basics	171
5.1.2 Riemannian geometry over covariance matrices in Gaussian context .	172
5.1.3 The case of shape matrices	173
5.1.4 The case of texture parameters	174
5.2 Development of Riemannian geometry adapted to deterministic compound-Gaussian model	175
5.3 Recursive change-point detection	176
5.3.1 Online implementation in Gaussian context	176
5.3.2 Problems relative to an online implementation in robust case	177
5.3.3 Recursive Stochastic optimization in $\mathcal{M}_{p,N}$	178
5.3.4 Application to recursive change-point detection	179
5.3.5 Numerical illustrations	179
5.4 Clustering SAR image time series	182
5.4.1 Description of the problem	182
5.4.2 K-mean algorithm for clustering	183
5.4.3 Geometric mean on $\mathcal{M}_{p,N}$	183
5.4.4 Application to real SAR images	185
5.5 Conclusions	191
5.A Proof of proposition 5.2.1 at p. 169	192
5.B Proof of proposition 5.3.1 at p. 179	192
Conclusions	195
Résumé en français	199
Bibliography	203

LIST OF FIGURES

1.1	Illustration of SAR image acquisition principle: a scatterer (building) is seen at two different acquisition angles (θ_1 and θ_2) during the processing time.	24
1.2	Comparison between a SAR and an optical image of the same scene (Los Angeles, Californie).	25
1.3	Illustration of data dimensionality	26
1.4	Illustration of monovariate versus multivariate points comparison.	26
1.5	Illustration of polarized electromagnetic waves. Depending on the direction of the magnetic field \mathbf{E} , the polarization is either horizontal or vertical.	27
1.6	Illustration of a repeat-pass interferometry acquisition system.	27
1.7	An ONERA RGB color-coded SAR image acquired for three consecutive frequency bands.	28
1.8	Illustration of an image with isotropy and non-dispersity hypothesis respected (Left) and not respected (Right). Credit: Michael Duquenoy.	29
1.9	Left: SAR image amplitude (dB). Right: Coefficient amplitude of a 3×3 decomposition using STFT (dB).	29
1.10	Illustration of local data selection ($p = 3, N = 9$) for detection test. The gray area corresponds to $\mathbb{W}_{1,T}$ and the central pixel (\mathbf{x}_5^t) is the test pixel.	30
1.11	Selection of a subset for analysing the empirical distribution of real data.	31
1.12	Empirical distribution of data on the selected subset.	32
1.13	Gaussian fitting of the empirical distribution of data on the selected subset.	32
1.14	Generalized Gaussian fitting of the empirical distribution of data on the selected subset.	33
1.15	General procedure for a change detection methodology.	37
1.16	Summary of various change detection methodologies.	39
1.17	Illustration of false alarm/good detection probabilities for a given threshold λ . The dark area corresponds to the probability of false alarms while the line filled area correspond to the probability of good detection.	41
1.18	Illustration of CFAR property: if the distribution varies when the parameter of the distribution changes, it is impossible to guarantee a false alarm rate.	41
1.19	Representation of Gaussian GLRT for $p = 2, T = 2$ when matrices are different. Left: Ellipsoids representing the covariance matrices. Right: Ellipsoids representing the arithmetic and geometric mean.	45
1.20	Representation of Gaussian GLRT for $p = 2, T = 2$ when matrices are equal. Left: Ellipsoids representing the covariance matrices. Right: Ellipsoids representing the arithmetic and geometric mean.	45
1.21	Experimental validation (10^4 Monte Carlo trials) of P_{FA} approximation of eq. (1.27) with $p = 15, T = 5$. Top-Left: non-asymptotic regime ($N = 20$). Top-right: non-asymptotic regime ($N = 50$). Bottom-left: asymptotic regime ($N = 200$). Bottom-right: asymptotic regime ($N = 1500$).	47

2.1	SAR acquisition geometry. A reflector is viewed at two different angles of illumination θ_1 and θ_2 in a given frequency. This information is summarized through the wave vectors $\mathbf{k}_1, \mathbf{k}_2$	56
2.2	Illustration of SAR spectra partitioning	57
2.3	Shannon 2×3 multi-band decomposition tree associated with $j = 2$. The positive part of $\Delta_{j,\bullet}^{1,+}$ of $\Delta_{j,\bullet}^1$ are given in the upper tree whereas $\Delta_{j,\bullet}^{2,+}$, positive part of $\Delta_{j,\bullet}^2$, are given at the bottom tree. The frequency tiles associated with the decomposition are the intervals $\Delta_{j,\bullet}^1 \times \Delta_{j,\bullet}^2$ for every fixed j : the whole tree involves all combination of nodes given at a fixed resolution level j	58
2.4	Peacock image - Its Cartesian Fourier magnitudes - Shannon $M = 4$ band frequency sub-selections and the corresponding 4-band wavelet subimages.	60
2.5	Peacock image - Its Polar Fourier magnitudes - Shannon $M = 4$ band frequency sub-selections and the corresponding 4-band wavelet subimages.	61
2.6	Cartesian product \otimes of spectral and angular multi-resolution trees associated with radial ‘2 bands - 2 resolutions’ and angular ‘3 looks - 2 resolutions’ splitting $\Psi_{j,[m,n]}^{S,\angle}(\mathfrak{K}, \theta)$ defined by Eq. (2.16) when considering $[f_0 - B/2, f_0 + B/2] \times [-\theta_B, \theta_B] = [1GHz, 5GHz] \times [-45 \text{ deg}, 45 \text{ deg}]$. The intervals represented are given for illustration.	62
2.7	Example of bright point decomposition. Left: Image I . Right: $\mathbf{C}_{1,[1,1]}^{S,\angle}$ with $R = L = 2$	64
2.8	Bell-shaped function with width $a = 3$, center $c = 0$ and different slope parameters $b \in \{1, 3, 10, 50\}$	64
2.9	Real Part (\mathcal{R}), Imaginary Part (\mathcal{I}) and Fourier transform (\mathcal{F}) shapes of wavelet $\Phi_{1,[\bullet,\bullet]}^{[d_1,d_2],\angle}$ functions, for $d_1 = d_2 \in \{0.1, 0.5, 1, 10\}$	66
2.10	Frequencies selectivity of wavelet packet for $f_0 = 9.6\text{GHz}$, $B = 640\text{MHz}$, $\theta_B = 0.25$ rad, $j=1$, $R=L=2$	67
2.11	Coefficients for Shannon and Bell-Shaped wavelets on SDMS Image with $R = 2$, $L = 2$. The improvement in terms of linear patterns (side lobes of bright points) for $d_1 = d_2 = 3$ are highlighted using dashed boxes.	69
2.12	Coefficients for Shannon and Bell-Shaped wavelets on SANDIA image with $R = 3$, $L = 3$	70
2.13	P_{FA} - λ curves ($j = 1$, $R = 5$, $L = 5$, $K = 88$). $d_1 = d_2 = \infty$. Left: AMF, Right: ANMF. Top: SANDIA dataset, Bottom: SDMS dataset.	74
2.14	A target near a bright point. Dataset is SANDIA, $R = L = 5$. The target has an amplitude of -60 dB when compared to the bright point	75
2.15	Results at $P_{FA} = 10^{-3}$. Top: $d_1 = d_2 = \infty$. Bottom: $d_1 = d_2 = 10$. Dataset is SANDIA, $R = L = 5$. The target has an amplitude of -60 dB when compared to the bright point.	75
2.16	Pd-SNR plots for several values of $d_1 = d_2$. Top: AMF Detector. Bottom: ANMF Detector.	78
3.1	Illustration of local data selection ($N_1 = N_2 = p = 3$) for detection test. The gray area corresponds to $\mathbb{W}_{1,T}$ and the central pixel (\mathbf{x}_5^t) is the test pixel.	85
3.2	Example of normalization effect when the shape matrix is the same for two K-distributed group of samples (blue and red dots). Left: The original bivariate data. Right: Normalized data in polar representation. In polar representation both histograms coincide.	87

3.3	Example of normalization effect when the shape matrix is the different for two K-distributed group of samples (blue and red dots). Left: The original bivariate data. Right: Normalized data in polar representation. In polar representation the two histograms do not coincide.	87
3.4	Illustration of g-convexity concept: For any set $(q_0, q_1) \in \mathcal{M}^2$, the function is convex when following the geodesic path between these two points.	91
3.5	Convergence property of estimates. Left: Same matrices at each date ($p = 3$, $N = 25$, $T = 3$, 1000 Trials, $\rho_0 = 0.70$, $\alpha = 0.30$, $\beta = 0.10$). Right: Different matrices at each date ($p = 3$, $N = 25$, $T = 3$, 1000 Trials, $\rho_1 = 0.08$, $\rho_2 = 0.90$, $\rho_3 = 0.10$, $\alpha = 0.30$, $\beta = 0.10$). The textures are different at each date for both settings.	93
3.6	Distribution of statistic under small sample size ($p = 3$, $T = 5$, $N = 9$, 10^5 Monte-Carlo Trials)	94
3.7	Distribution of statistic under large sample size ($p = 3$, $T = 5$, $N = 100$, 10^5 Monte-Carlo Trials)	94
3.8	Texture CFAR behaviour. Top-left: $\hat{\Lambda}_G$. Top-center: $\hat{\Lambda}_{t_1}$. Top-right: $\hat{\Lambda}_{Wald}$. Bottom-left: $\hat{\Lambda}_{MT}$. Bottom-center: $\hat{\Lambda}_{Mat}$. Bottom-Right: $\hat{\Lambda}_{Tex}$. $\rho = 0.3$ at each date for all the curves.	95
3.9	Matrix CFAR behaviour. Top-left: $\hat{\Lambda}_G$. Top-center: $\hat{\Lambda}_{t_1}$. Top-right: $\hat{\Lambda}_{Wald}$. Bottom-left: $\hat{\Lambda}_{MT}$. Bottom-center: $\hat{\Lambda}_{Mat}$. Bottom-Right: $\hat{\Lambda}_{Tex}$	96
3.10	Asymptotic consistency of estimates. Left: $p = 3$, $T = 3$, 1000 Trials, $\rho_0 = 0.70$, $\alpha = 0.3$, $\beta = 0.1$, same textures for each date. Right: $p = 3$, $T = 3$, 1000 Trials, $\rho_1 = 0.29$, $\rho_2 = 0.56$, $\rho_3 = 0.58$, $\alpha = 0.3$, $\beta = 0.1$, same textures for each date.	97
3.11	P_D as a function of $d_B(\Sigma_0, \Sigma_1)$ for a Gaussian case. Left: Change in Shape only. Right: Change in scale only. Parameters: $p = 5$, $N = 15$, $T = 10$, $P_{FA} = 10^{-3}$. Curves are computed with 600 Monte Carlo trials.	98
3.12	P_D as a function of $d_B(\Sigma_0, \Sigma_1)$ for a Student-t with degrees of freedom $\nu = 2$. Left: Change in Shape only. Right: Change in scale only. Parameters: $p = 5$, $N = 15$, $T = 10$, $P_{FA} = 10^{-3}$. Curves are computed with 600 Monte Carlo trials.	99
3.13	P_D as a function of $d_B(\Sigma_0, \Sigma_1)$ for a K-distributed data with shape parameter $\nu = 0.1$. Left: Change in Shape only. Right: Change in scale only. Parameters: $p = 5$, $N = 15$, $T = 10$, $P_{FA} = 10^{-3}$. Curves are computed with 600 Monte Carlo trials.	99
3.14	ROC curves obtained on synthetic data (10^4 Monte-Carlo trials). Top-left: Problem (3.9). Top-right: Problem (3.10). Bottom-left: Problem (3.11). Bottom-right: Gaussian setting.	101
3.15	UAVSAR Dataset in Pauli representation. Left: April 23, 2009. Middle: May 15, 2011. Right: Ground Truth. Top: Scene 1. Bottom: Scene 2.	102
3.16	SDMS Dataset. Top-left: FP0120. Top-middle: FP0121 . Top-right: FP0124. Bottom-left: Ground Truth FP0120-FP0121-FP0124. Bottom-right: Ground Truth FP0121-FP0124.	102
3.17	Value of the different statistics for SDMS FP0120-FP0121-FP0124. $p = 3$, $N_1 = N_2 = 11$	103
3.18	Value of the different statistics for UAVSAR Scene 1. $p = 3$, $N_1 = N_2 = 11$	104
3.19	Value of the different statistics for UAVSAR Scene 2. $p = 3$, $N_1 = N_2 = 11$	104
3.20	P_D versus P_{FA} on real data. Top-left: SDMS FP0121-FP0124. Top-right: FP0120-FP0121-FP0124. Bottom-left: UAVSAR Scene1. Bottom-right: UAVSAR Scene 2. For all images, $p = 3$, $N_1 = N_2 = 11$	105

3.21	P_D as a function of window size at $P_{FA} = 10^{-2}$. Top-left: SDMS FP0121-FP0124. Top-right: FP0120-FP0121-FP0124. Bottom-left: UAVSAR Scene1. Bottom-right: UAVSAR Scene2. For all images, $p = 3$.	107
3.22	P_D versus P_{FA} on SDMS FP0121-FP0124 with wavelet decomposition. $N_1 = N_2 = 11$.	107
4.1	$P_D = f(\text{SNR})$ for a Gaussian model. $p = 10$, $N = 25$, $T = 10$, $t_C = 5$, $\rho_{t < t_C} = 0.01$. The false alarm rate is fixed at $P_{FA} = 10^{-3}$ and each point of the curve is plotted using 4800 Monte-Carlo trials.	142
4.2	$P_D = f(\text{SNR})$ for an Elliptical model. $p = 10$, $N = 25$, $T = 10$, $t_C = 5$, $\rho_{t < t_C} = 0.01$, $\alpha = 0.1$, $\beta = 0.3$. The false alarm rate is fixed at $P_{FA} = 10^{-3}$ and each point of the curve is plotted using 4800 Monte-Carlo trials.	143
4.3	$P_D = f(\text{SNR})$ for an elliptical model with texture equality constraint. $p = 10$, $N = 25$, $T = 10$, $t_C = 5$, $\rho_{t < t_C} = 0.01$, $\alpha = 0.1$, $\beta = 0.3$. The false alarm rate is fixed at $P_{FA} = 10^{-3}$ and each point of the curve is plotted using 4800 Monte-Carlo trials.	143
4.4	Description of synthetic images parameters. Parameters α and β only apply for elliptical model.	144
4.5	Estimation/Detection results at $P_{FA} = 10^{-3}$ for synthetic Gaussian data ($p = 5$, $N = 25$, $T = 5$). A change-point detected and estimated corresponds to non-zero value on the resulting image.	146
4.6	Estimation/Detection results at $P_{FA} = 10^{-3}$ for synthetic K-distributed data ($p = 5$, $N = 25$, $T = 5$). A change-point detected and estimated corresponds to non-zero value on the resulting image.	147
4.7	Estimation/Detection results at $P_{FA} = 10^{-3}$ for synthetic K-distributed data with texture equality constraint between dates ($p = 5$, $N = 25$, $T = 5$). A change-point detected and estimated corresponds to non-zero value on the resulting image.	148
4.8	Estimation/Detection on SDMS dataset results at $P_{FA} = 10^{-4}$	150
4.9	Estimation/Detection on SDMS dataset results at $P_{FA} = 10^{-4}$	151
4.10	Illustration of the problem	157
4.11	MSE on the change-point for $p = 3$, $T = 20$, $\alpha_0 = 0.1$, $\alpha_1 = 0.5$. The estimators curves have been computed with 4800 Monte Carlo trials.	158
4.12	Evolution of $\log_{10} \sqrt{(\text{HCRWWB})_{M,M}}$ for several parameters p and N , $T = 100$, $\alpha_0 = 0.1$ and $\alpha_1 = 0.3$. The dashed line corresponds to the region where $\sqrt{(\text{HCRWWB})_{M,M}} = 10^{-2}$.	159
5.1	Example of manifold: Boy's surface in a 3-dimensional space. Credit: Maksim.	171
5.2	Illustration of Riemannian geometry concepts.	172
5.3	Top-left: Natural distance to shape matrix. Top-Right: Natural distance to texture parameters. Bottom-left: Natural compound-Gaussian distance. Bottom-Right: Mean time of computation over the Monte-Carlo trials. Parameters: $p = 10$, $N = 49$, $\rho = 0.71 + i * 0.71$, $\alpha = 0.01$, $\beta = 100.00$, $T = 400$, 4800 Monte-Carlo Trials.	181
5.4	$H - \alpha$ clustering. Z1: High Entropy Multiple Scattering, Z2: High Entropy Vegetation Scattering, Z3: (Not a Feasible Region), Z4: Medium Entropy Multiple Scattering, Z5: Medium Entropy Vegetation Scattering, Z6: Medium Entropy Surface Scattering, Z7: Low Entropy Multiple Scattering, Z8: Low Entropy Dipole Scattering, Z9: Low Entropy Surface Scattering.	184

5.5	Left: Span of EMISAR SAR image (Span) over Foulom region in Denmark. Right: Optical image for comparison (@Google)	185
5.6	$H - \alpha$ initialisation	186
5.7	Results of clustering on EMISAR image. Top-left: Wishart classifier. Top-right: Wishart classifier with geodesic mean. Bottom-left: Robust shape clustering. Bottom-right: Shape and texture Riemannian clustering (proposed).	187
5.8	UAVSAR Snjoaq dataset in Pauli representation	188
5.9	Output of $\hat{\Lambda}_{MT}$ statistic with a window size of 7×7	189
5.10	Results of Snjoaq temporal clustering with $K = 4$ classes. Top: Wishart classifier. Middle-Top: Wishart classifier with Riemannian mean. Middle-Bottom: Robust shape clustering. Bottom: Proposed approach.	190

LIST OF TABLES

2.1	Description of Dataset	68
2.2	Results of detection when randomising steering vector. SNR=0 dB, 100 different signatures have been tested for 100 different target positions on each image.	77
3.1	Description of SAR data used	103
4.1	Time-consumption in seconds.	159

NOMENCLATURE

Probability and Statistics

ξ	Shape matrix parameter: $p\boldsymbol{\Sigma}/\text{Tr}(\boldsymbol{\Sigma})$
\mathbb{CAE}	Complex angular elliptical distribution
\mathbb{CCG}	Complex Compound Gaussian distribution
\mathbb{CE}	Complex elliptical distribution
χ^2	Chi-squared distribution
\mathbb{CN}	Complex normal distribution
$\boldsymbol{\Sigma}$	Covariance matrix parameter
\mathbb{CW}	Complex Wishart distribution
$\hat{\Lambda}$	Statistic of decision
$\mathbb{E}\{\bullet\}$	Esperance operator
$\hat{\theta}$	Estimator of parameter θ
λ	Threshold of decision
\mathbf{x}	Random vector
θ	Parameter of a parametric model
τ	Scale parameter: $\text{Tr}(\boldsymbol{\Sigma})/p$

Functions

$\delta_{i,j}$	Kronecker function equal to one if $i = j$ and 0 otherwhise
Γ	Gamma function
$\Gamma_p(N)$	Multivariate gamma function
$\mathbb{1}_{\mathbb{K}}$	Indicator function over set \mathbb{K}
$\mathbf{x}^H \boldsymbol{\xi}^{-1} \mathbf{x}$	

Operators

$ \bullet $	Determinant operator
$\text{diag}(\bullet)$	Diagonal operator: over a matrix return its diagonal elements, over a vector return a matrix with diagonal elements as elements of the vector
\bullet^H	Transpose conjugate operator
$\inf_{\theta} f$	Infimum of function f subjected to parameter θ
\otimes	Kronecker product
$\mathcal{I}(\bullet)$	Imaginary part of a complex vector

$\mathcal{R}(\bullet)$	Real part of a complex vector
$\max_{\theta} f$	Maximum of function f subjected to parameter θ
$\min_{\theta} f$	Minimum of function f subjected to parameter θ
\odot	Hadamard product
$\bar{\bullet}$	Conjugate operator
$\partial f / \partial \mathbf{x}$	Partial derivative of f with regards to vector \mathbf{x}
$\sup_{\theta} f$	Supremum of function f subjected to parameter θ
$\tau_{[p,q]} g(x, y)$	$g(x-p, y-q)$
$\text{Tr}(\bullet)$	Trace operator
\bullet^T	Transpose operator
$\text{vech}^{-1}(\bullet)$	Obtain a symmetric matrix by rearranging the elements of the vector
$\text{vech}(\bullet)$	Vectorisation operator over a symmetric matrix by omitting duplicate terms
$\text{vec}^{-1}(\bullet)$	Obtain a matrix by rearranging the elements of the vector
$\text{vec}(\bullet)$	Vectorisation operator over a matrix or a set of values

Physical constants

c	Speed of light in a vacuum inertial system
-----	--

Riemannian Geometry

$\exp_{\mathcal{M}}(\bullet)$	Exponential mapping for manifold \mathcal{M}
$\log_{\mathcal{M}}(\bullet)$	Logarithm mapping for manifold \mathcal{M}
\mathcal{M}	Arbitrary manifold
$\mathcal{M}_{p,N}$	Manifold of elements in $\mathbb{S}_{\mathbb{H}, \bullet }^p \times (\mathbb{R}^+)^N$
$\mathcal{P}_{\eta}\mathcal{M}$	Orthogonal projection on $T_{\eta}\mathcal{M}$
$D\phi[\xi]$	Directional derivative of function ϕ in the direction ξ
$T_{\eta}\mathcal{M}$	Tangent space at point η to the manifold \mathcal{M}
$\nabla^{\mathcal{M}}$	Levi-Civita connection
$\mathbb{S}_{\mathbb{H}, \bullet }^p$	Sub manifold of matrices in $\mathbb{S}_{\mathbb{H}}^p$ whose determinant is equal to one

Sets

\mathbb{B}	Set $\{0, 1\}$
\mathbb{C}	Set of Complex Numbers
\mathbb{D}^p	Set of diagonal real matrices of size $p \times p$
$[\![1, N]\!]$	Set $\{1, \dots, N\}$
\mathbb{N}	Set of Natural numbers
$\mathbb{S}_{\mathbb{H}}^p$	Set of positive definite Hermitian matrices of size $p \times p$
\mathbb{S}_{++}^p	Set of positive definite symmetric matrices of size $p \times p$
\mathbb{R}	Set of Real Numbers
\mathbb{Z}	Set of signed Natural numbers
$L^n(\mathbb{C}^2)$	Set of the functions having values in \mathbb{C}^2 for which the n -th power of the absolute value is integrable

ACRONYMS

3FT	Fractional Fast Fourier Transform.
AMF	Adaptive Matched Filter.
ANMF	Adaptive Normalized Matched Filter.
BB	Barankin Bound.
CCD	Coherent Change Detection.
CCG	Complex Compound Gaussian.
CES	Complex Elliptical Symmetric.
CFAR	Constant False Alarm Rate.
CRB	Cramer-Rao Bound.
CUSUM	Cumulative Sum Control Chart.
EVD	EigenValue Decomposition.
FIM	Fisher Information Matrix.
FWER	FamilyWise Error Rate.
GLRT	Generalized Likelihood Ratio Test.
LRT	Likelihood Ratio Test.
LTFD	Linear Time Frequency Distributions.
MC	Monte Carlo.
MSE	Mean Square Error.
PCA	Principal Component Analysis.
PDH	Positive Definite Hermitian.
PW	Paley-Wiener.
RMA	Range Migration Algorithm.
SAR	Synthetic A��rture Radar.
SCM	Sample Covariance Matrix.
SIRV	Spherically Invariant Random Vectors.
SLC	Single Look Complex.
SVD	Single Value Decomposition.
SVM	Support Vector Machine.
TE	Tyler's Estimator.
UMP	Uniformly Most Powerful.
WWB	Weiss-Weinstein Bound.

CHAPTER 1

AN OVERVIEW OF SAR IMAGE TIME SERIES ISSUES

Contents:

1.1	Specificities of SAR data	24
1.2	The principle of diversity: exploiting multivariate data	25
1.2.1	Interest of multivariate data	25
1.2.2	Sources of diversity in SAR images	26
1.2.3	Spectro-angular diversity using wavelet analysis of SAR images	28
1.3	Models considered for the distribution of multivariate SAR images	30
1.3.1	The data	30
1.3.2	The Gaussian model	31
1.3.3	Non-Gaussianity of high-resolution images	31
1.3.4	The Complex Elliptical model	32
1.3.5	The deterministic compound-Gaussian model	35
1.4	Change detection in SAR image time series	36
1.4.1	An overview of change detection techniques for SAR images	36
1.4.2	Defining the problem of covariance equality testing	40
1.4.3	On the hypothesis problem and test statistics	40
1.4.4	On the GLRT technique and known alternatives	42
1.4.5	A note on the false discovery rate and familywise error rate	43
1.4.6	Results on the equality of covariance in Gaussian context	44
1.4.7	An extension to the Elliptical case using local Normal Asymptotic theory	46
1.4.8	Some attempts of equality testing under deterministic compound-Gaussian model	48
1.5	Change-point estimation in time series	49
1.5.1	Some techniques for estimating the change-point	50
1.5.2	Joint detection and estimation of change-points for SAR images	51
1.6	Summary	53

This first introductory chapter aims at giving a brief overview of issues, concerning the analysis of changes in a SAR image time series that we will consider in the context of this thesis. To that end, we will describe shortly the specificities of acquisition of SAR images which differ from other images acquisition approaches. We will detail the notion of diversity

which allows considering multivariate data. Then we will discuss how the traditional Gaussian model does not accurately describe the data for high-resolution SAR images. From that point, we will give a brief overview of the change detection problem associated with statistical methodologies and focus peculiarly on covariance equality testing approach. Finally, we will consider the alternative problem of change-point detection and estimation for which we will describe an interesting algorithm from the literature.

1.1 SPECIFICITIES OF SAR DATA

Synthetic Aperture Radar (SAR) is a technology developed in the 1950s as a military reconnaissance tool with all-weather and 24-hour aerial remote surveillance capabilities. It was first introduced by Carl A. Wiley in 1951 under the terminology “Doppler radar beam sharpening” and patented under patent number 3,196,436 dated July 20, 1965, [Love, 1985].

The idea behind this technology was to solve the resolution limitations of radars of this time. Indeed, the resolution of an antenna is inversely proportional to its size, thus improving resolution means obtaining antenna of a size which was impractical on airborne or spaceborne sensors. SAR sensors tackle this problem by taking advantage of the movement of the sensors to create an equivalent larger synthetic antenna. As presented in Figure 1.1, since the swath of the antenna is large, the same objects are illuminated several times during the movement of the sensor meaning that it is possible to coherently sum the backscattering over time to improve the resolution in the direction of the moving radar.

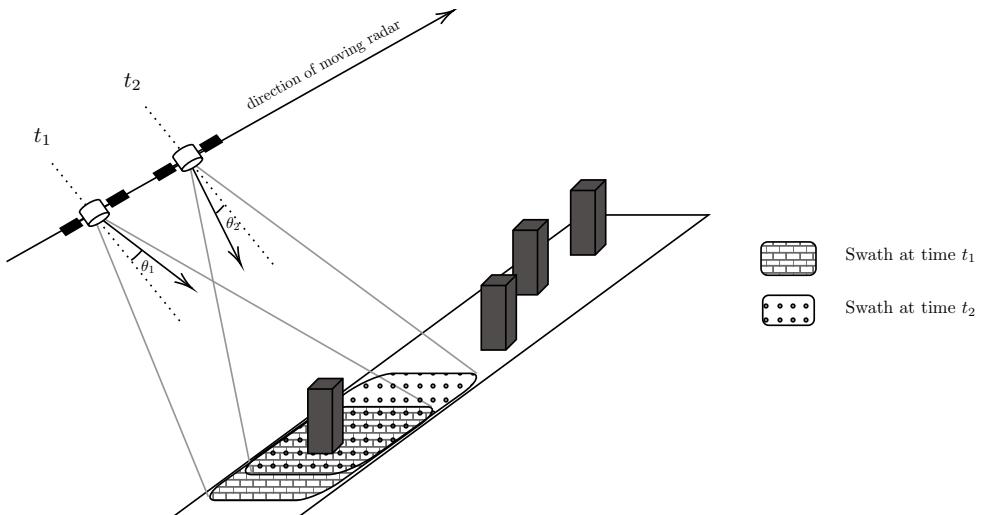


Figure 1.1: Illustration of SAR image acquisition principle: a scatterer (building) is seen at two different acquisition angles (θ_1 and θ_2) during the processing time.

As a result of this technology, azimuth integration has improved allowing us to obtain images with resolution lower than 1m. The drawback lies however in the presence of a typically multiplicative noise called speckle which is due to the coherent processing of various scattered signals [Goodman, 1976]. The speckling in an image reduces the ability of a human observer to resolve fine details within the image as well as making typical processing techniques based on additive noise unadapted. As an example, Figure 1.2 shows a comparison between an optical and a SAR image of the same scene. The SAR image presents a high

level of noise while comparatively, the optical image does not present such level of noise. The speckle noise must then be taken into account in order to develop well-adapted image processing methodologies.



Figure 1.2: Comparison between a SAR and an optical image of the same scene (Los Angeles, California).

Another distinction between optical images and SAR images lies in the nature of the data which is real-valued in optical images but complex-valued for SAR images. The complex nature of the images comes from the SAR reconstruction techniques which work on modulated signals to generate the images [Carrara et al., 1995]. As a consequence, the literature on SAR processing literature usually separates into two categories depending on the nature of the data considered. A high number of works considers the amplitude of the complex value while discarding the phase, in order to develop methodologies based on the backscattered amplitude. On the other hand, techniques such as interferometry require the phase value for the fine measurement of displacements [Gens and Genderen, 1996].

In the context of this thesis, we will consider complex SAR images in order to take advantage of all the information (amplitude and phase) available.

1.2 THE PRINCIPLE OF DIVERSITY: EXPLOITING MULTIVARIATE DATA

In the context of this thesis, we consider a time series of multivariate images as illustrated in Figure 1.3. For each pixel we assume that a vector of data is available and we have the evolution of those vectors over time.

1.2.1 Interest of multivariate data

Intuitively, having more information about a system allows us to better infer about its properties. When it concerns the comparison between datasets, as illustrated in Figure 1.4, multivariate data can lead to better discrimination compared to a monovariate dataset. It is often desirable to obtain the most of information possible about the objects of interest.

Multivariate data is however not always informative. Indeed, when the dimension is high, the proportion of channels bringing discriminative information about an object can

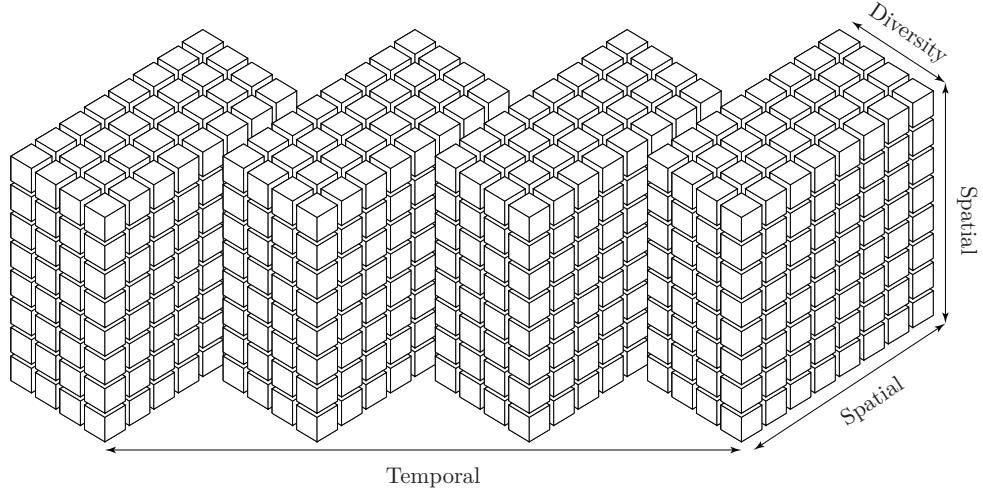


Figure 1.3: Illustration of data dimensionality

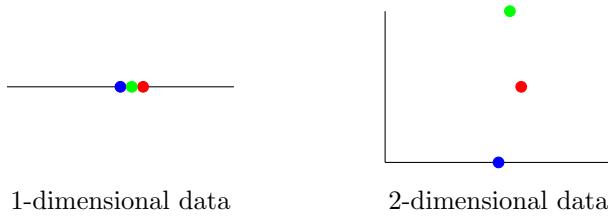


Figure 1.4: Illustration of monovariate versus multivariate points comparison.

decrease. In such cases, considering all dimensions can be seen as counterproductive with respect to complexity. Moreover, in statistical adaptive schemes, when the dimension of the parameter space increases, it is well known that the number of samples required for inference grows fast as well. This is why many high-dimensional processing methodologies rely on low-rank structure¹ assumption or dimension reduction techniques such as Principal Component Analysis (PCA) [F.R.S., 1901].

1.2.2 Sources of diversity in SAR images

As explained earlier, exploiting diversity allows considering advanced processing methodologies to improve results in classical target detection and change detection schemes. The literature on the diversity concerning SAR images is vast as many kinds of diversities can be exploited.

First of all, the polarimetric diversity is the most obvious one when it concerns SAR images. Since SAR sensors are based on electromagnetic waves, it is possible to control their polarization. As illustrated in Figure 1.5, the direction of the electric field lies in the plane perpendicular to the direction of propagation and defines the polarization of the wave.

¹See for example the Ph.D. thesis [Breloy, 2015].

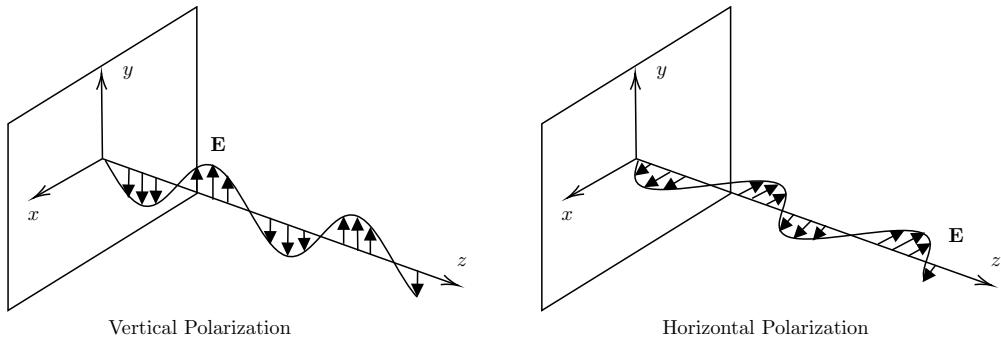


Figure 1.5: Illustration of polarized electromagnetic waves. Depending on the direction of the magnetic field \mathbf{E} , the polarization is either horizontal or vertical.

Polarimetric SAR sensors are able to receive and transmit waves in Horizontal (H) and Vertical (V) polarization modes on alternate pulses. Hence, depending on the setup, the signal obtained is either HH (transmit and receive in H mode), VV (transmit and receive in V mode), HV (transmit in H mode and receive in V mode) or VH (transmit in V mode and receive in H mode). The amount of returned signal for different polarizations depends on the physics of the interaction of microwaves with the surface. This allows to characterize the different scatterers present in a scene.

Note that for most situations, the canals HV and VH are identical yielding a diversity vector of dimension $p = 3$.

Another source of diversity can be obtained through interferometry sensors. In this case, the scene is seen at two different close looking-angles as illustrated in Figure 1.6. The difference of phase between the two acquisitions is usually used for altitude measurements.

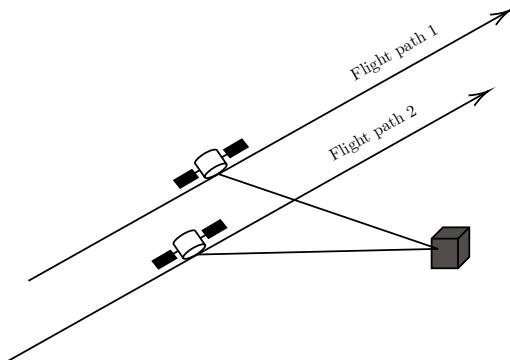


Figure 1.6: Illustration of a repeat-pass interferometry acquisition system.

However, in many cases, these diversities are not available due to the fact that they require specific sensors. Thus, the most common available data corresponds to the commonly referred to as Single Look Complex (SLC) data. On the other hand, polarimetric or interferometric diversities may not be sufficient to discriminate subtle changes.

For such situations, it has been shown that if the resolution is high enough, it is possible

to obtain a diversity corresponding to the physical behavior of the scatterers present in the scene. This point will be detailed hereafter.

1.2.3 Spectro-angular diversity using wavelet analysis of SAR images

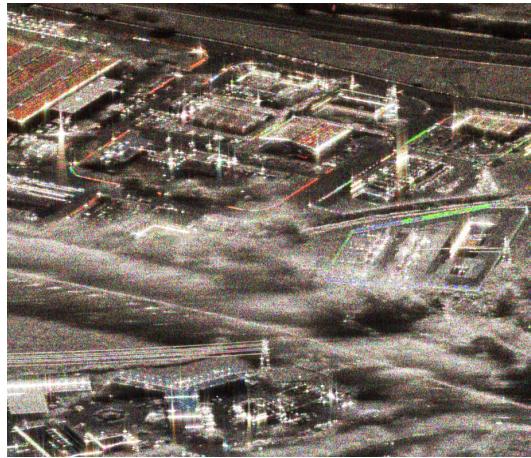


Figure 1.7: An ONERA RGB color-coded SAR image acquired for three consecutive frequency bands.

Figure 1.7 presents an ONERA SAR image in X-band. The responses of the scene relative to three consecutive frequency bands have been coded in RGB color-coding. The three consecutive frequency bands correspond to the frequencies emitted by the antenna of the moving radar system. This image corresponds to a campaign measurement of the same scene where 3 different bands of frequency have been used by splitting the total bandwidth of the antenna in 3. Then the amplitude of each image obtained with the different band is coded to the RGB color scheme. Red points are responding only on the first band, green ones on the second band and the blue ones on the third band. They are called colored scatterers. Gray points are called white scatterers as they are responding equivalently in the three sub-bands. This image perfectly illustrates how some scatterers have different behavior given the band used. Similar results can be achieved when looking at the scene at different ranges of angles.

In HR SAR Images, the hypothesis of isotropy and non-dispersivity of the scatterers is no longer obvious. When a target is illuminated using a large bandwidth and a large range of angles, it is more reasonable to assume that its response is dependent on the wave vector. Recent studies of the spectral and angular behavior of the scatterers have shown the variation of the scatterers' response for several angles of illumination and several frequencies [Bertrand and Bertrand, 1996, Duquenoy et al., 2010, Tria, 2005] as illustrated in Figure 1.8. This diversity is of interest as it can allow to discriminate between objects when polarimetric diversity is neither present or sufficient.

Several works have investigated methods for retrieving the spectro-angular diversity. For example, approaches such as steerable pyramids [Simoncelli et al., 1992, Unser et al., 2011], curvelets [Candès et al., 2006] or subspaces [Durand et al., 2009, Brigui et al., 2014] are possible. However, they are usually heavy methods. For example, subspace methods have high-computational cost and are not adaptive. They also assume the knowledge of a physical model, which makes them specific to an application. Linear Time-Frequency Distributions

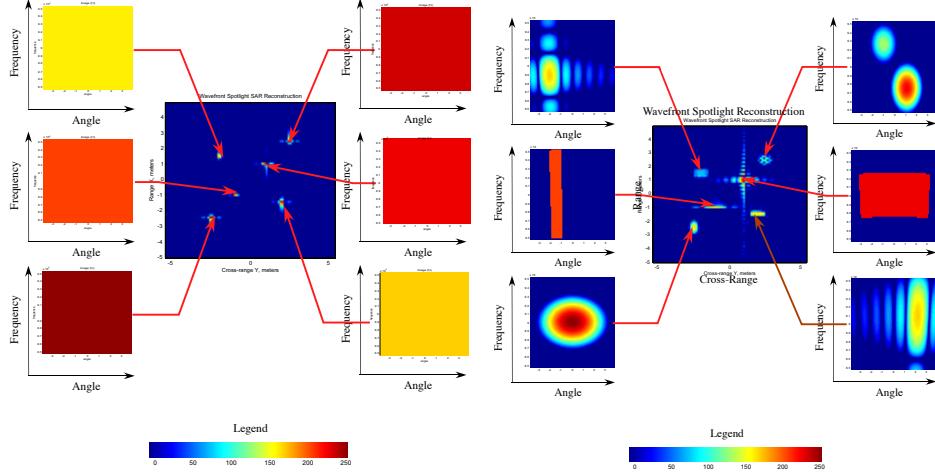


Figure 1.8: Illustration of an image with isotropy and non-dispersity hypothesis respected (Left) and not respected (Right). Credit: Michael Duquenoy.

(LTFD) is a simpler approach that allows analyzing SAR data to retrieve non-stationary information such as spectral and angular behaviors.

Retrieval of spectro-angular diversity using LTFD have been investigated in work such as [Tria et al., 2007, Ovarlez et al., 2003]. Time-Frequency has been used in [Brekke et al., 2013], where the azimuth bandwidth has been separated into two sub-bands for ship detection. In those works, the spectro-angular information has presented promising results. The Short Time Fourier Transform (STFT) has been used for target detection applications in [Ovarlez et al., 2017] or for change detection in [Mian et al., 2017]. However, the decomposition induced side lobes on the sub-images which may decrease the performance.

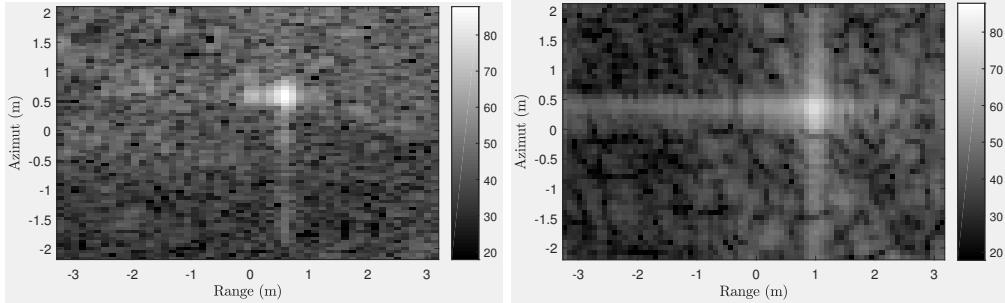


Figure 1.9: Left: SAR image amplitude (dB). Right: Coefficient amplitude of a 3×3 decomposition using STFT (dB).

Indeed, the works presented here relied on the STFT which is known in Time-frequency literature for its high level of side-lobes with regards to other wavelet alternatives. As an example, if we consider the decomposition in Figure 1.9, an undesired linear pattern appears at the bright point location. It is well known that wavelet can be used in order to tackle

problems of sidelobes.

Wavelet decomposition of SAR images have been studied for many applications. In [De Grandi et al., 2007, De Grandi et al., 2009], wavelet frames have been used in order to derive a measure of polarimetric texture used in segmentation and target detection schemes. Wavelet transforms on SAR images have been used in [Zecchetto and Biasio, 2008] to retrieve wind fields. In [Lopez-Martinez and Fabregas, 2002], wavelets have been used to reduce speckle noise in interferometric SAR images. Fusion techniques on wavelet coefficients have been used in [Ma et al., 2012] in order to compute a change detection map. In [Subotic et al., 1994, Irving et al., 1997], multi-resolution information is used for target detection schemes. However, those works did not consider specifically the sidelobe problem. This point will be tackled in Chapter 2 where new wavelet packets will be introduced to reduce the level of sidelobes.

Once the diversity needed has been obtained (either naturally present or thanks to time-frequency tools), each pixel is represented by a vector. Since SAR images are noisy, the pixels can be modelled as the realization of a random variable in order to derive adapted processing techniques. We propose, in the next section, to consider multivariate statistical distributions adapted to model the values of the pixel vectors.

1.3 MODELS CONSIDERED FOR THE DISTRIBUTION OF MULTIVARIATE SAR IMAGES

1.3.1 The data

Denote by $\mathbb{W}_{1,T} = \{\mathbf{X}_1, \dots, \mathbf{X}_T\}$ a collection of T mutually independent groups of p -dimensional i.i.d complex vectors: $\mathbf{X}_t = [\mathbf{x}_1^t, \dots, \mathbf{x}_N^t] \in \mathbb{C}^{p \times N}$. With regards to SAR images these sets correspond to the local observations on a spatially sliding windows as illustrated in Figure 1.10. The subscript k correspond to a spatial index while the superscript t corresponds to a time index.

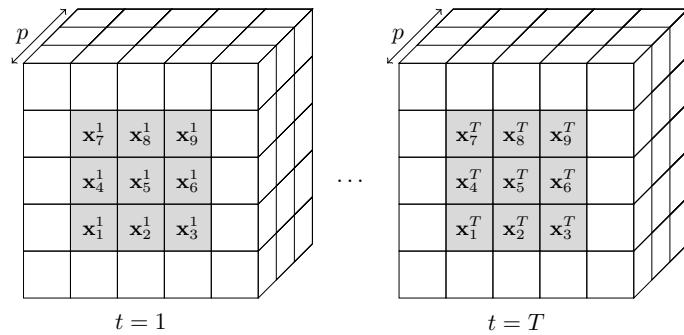


Figure 1.10: Illustration of local data selection ($p = 3, N = 9$) for detection test. The gray area corresponds to $\mathbb{W}_{1,T}$ and the central pixel (\mathbf{x}_5^t) is the test pixel.

Following SAR clutter analysis, we assume $\forall(k, t) \in \llbracket 1, N \rrbracket \times \llbracket 1, T \rrbracket$, $\mathbb{E}\{\mathbf{x}_k^t\} = \mathbf{0}_p$. This assumption is indeed well-respected in complex SAR images and is a classic assumption in SAR data processing literature [Deng et al., 2017].

Given $t \in [1, T]$, we denote for any $k \in [1, N]$, $\Sigma_t = \mathbb{E}\{\mathbf{x}_k^t \mathbf{x}_k^{t^H}\}$ the shared covariance matrix among the elements of the group \mathbf{X}_t .

1.3.2 The Gaussian model

The centered complex Gaussian distribution is the most encountered one in the SAR literature with regards to analysis of covariance matrices. Its sole parameters are the set of covariance matrices $\{\Sigma_t | t \in [1, T]\}$ such that:

$$\mathbf{x}_k^t \sim \mathcal{CN}(\mathbf{0}_p, \Sigma_t). \quad (1.1)$$

The probability density function (p.d.f) of $x \sim \mathcal{CN}(\mathbf{0}_p, \Sigma)$ is given by:

$$p_{\mathbf{x}}^{\mathcal{CN}}(\mathbf{x}; \Sigma) = \frac{1}{\pi^p |\Sigma|} \exp(-\mathbf{x}^H \Sigma^{-1} \mathbf{x}). \quad (1.2)$$

Note: While it is defined as a complex model, the distribution is assumed to be circular (i.e. $\mathbb{E}\{\mathbf{x}_k^t (\mathbf{x}_k^t)^T\} = \mathbf{0}_{p \times p}$) meaning that it is equivalent to a real-valued Gaussian distribution of a the stacked vector of real and imaginary parts.

1.3.3 Non-Gaussianity of high-resolution images

The Gaussian model has been widely popular to model the empirical distribution of the data for standard resolution images [Conradsen et al., 2001]. 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.

However, as described in [Greco and Gini, 2007], [Gao, 2010] or in [Ollila et al., 2012b], while the Gaussian hypothesis is the most popular one, it fails to accurately describe the heterogeneity observed in very high-resolution images. Indeed, in those images, the number of scatterers in each resolution cell 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 [Yueh et al., 1989, Muller, 1994], the Weibull distribution in [Buccarelli et al., 1995] or inverse Generalized Gaussian distribution in [Freitas et al., 2005]. These various models belong to the family of complex elliptical distributions which generalizes them as discussed in [Ollila et al., 2012a].

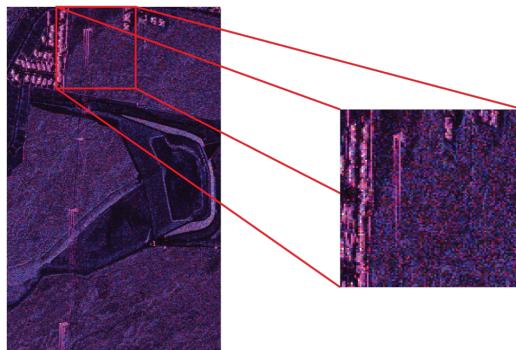


Figure 1.11: Selection of a subset for analysing the empirical distribution of real data.

As an example, let us consider a high-resolution image obtained from the UAVSAR dataset². From this image, we consider the HH polarization where we consider a small subset as shown in Figure 1.11.

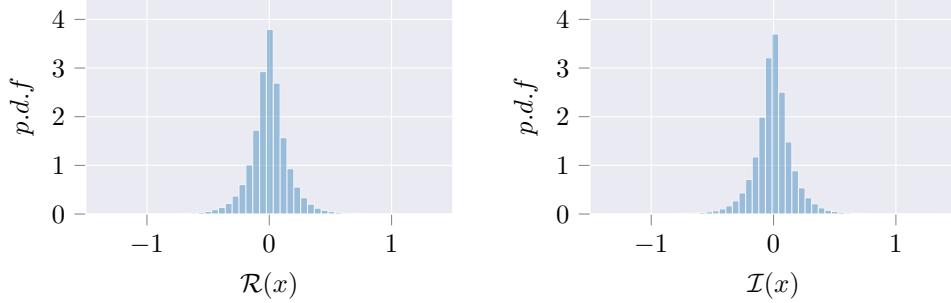


Figure 1.12: Empirical distribution of data on the selected subset.

The histogram of both real and imaginary parts, plotted in Figure 1.12, shows an impulsive behaviour with a heavy tail which does not correspond typically to a Gaussian distribution. Fitting for Gaussian model is shown in Figure 1.13 and generalized Gaussian model is shown in Figure 1.14. From these results it is clear that the Gaussian model fails at describing the empirical distribution while the Generalized Gaussian has an overall better description. This distribution is part of the more general family of distributions known as the elliptical distributions which will be considered hereafter.

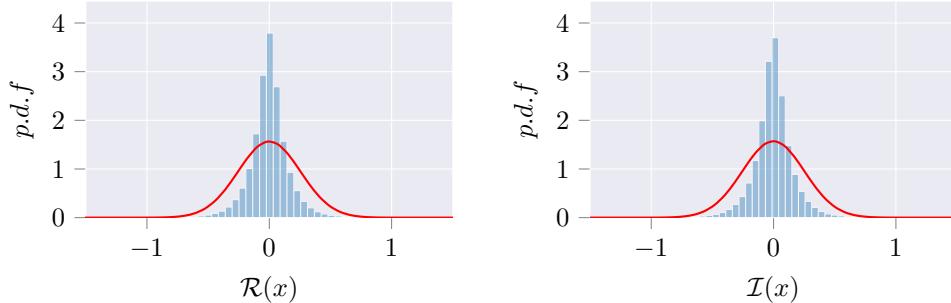


Figure 1.13: Gaussian fitting of the empirical distribution of data on the selected subset.

1.3.4 The Complex Elliptical model

This model is a extension of the Gaussian distribution, depending on a function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ called density generator that satisfies the finite moment condition $\mathfrak{m}_{p,g} = \int_{\mathbb{R}^+} t^{p-1} g(t) dt < \infty$ and a set of scatter matrices $\{\boldsymbol{\Xi}_t \in \mathbb{S}_{\mathbb{H}}^p | t \in [1, T]\}$. The complex elliptical model will be written as follows:

$$\mathbf{x}_k^t \sim \mathbb{C}\mathcal{E}(\mathbf{0}_p, g, \boldsymbol{\Xi}_t). \quad (1.3)$$

²Available at <https://uavstar.jpl.nasa.gov>.

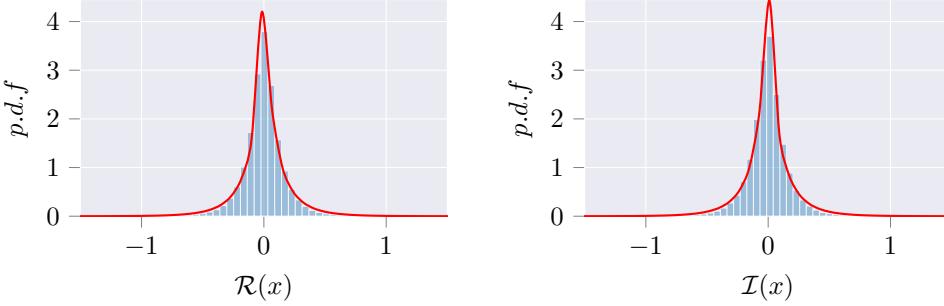


Figure 1.14: Generalized Gaussian fitting of the empirical distribution of data on the selected subset.

The p.d.f of $\mathbf{x} \sim \mathbb{CE}(\mathbf{0}_p, g, \Sigma)$ is given by:

$$p_{\mathbf{x}}^{\mathbb{CE}}(\mathbf{x}; \Xi, g) = \mathfrak{C}_{p,g} |\Xi|^{-1} g(\mathbf{x}^H \Xi^{-1} \mathbf{x}), \quad (1.4)$$

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

$$\mathfrak{C}_{p,g} \triangleq 2(\mathfrak{S}_p m_{p,g})^{-1}, \quad (1.5)$$

where $\mathfrak{S}_p = 2\pi^p / \Gamma(p)$ is the surface area of the complex unit p -sphere $\mathbb{CS}^p = \{\mathbf{s} \in \mathbb{C}^p; \|\mathbf{s}\|_2 = 1\}$ and Γ is the gamma function.

In practice, as shown in [Ollila et al., 2012a], the scatter matrix is equal to the covariance up to a scale factor which only depends on the dimension of the matrix and the density generator function. It means that, if we assume that the samples share the same density generator function, testing an equality of scatter matrices or covariance matrices is equivalent. Thus as a simplification, we will consider the equality of scatter matrices when we refer to covariance equality in the elliptical case.

Note: A classic parametrisation of the covariance matrix is of the following form:

$$\Sigma_t = \tau_t \xi_t, \quad (1.6)$$

where:

- $\tau_t \in \mathbb{R}^+$ is the *scale parameter*³ which is equivariant under multiplication by a positive constant.
- $\xi_t \in \mathbb{S}_{\mathbb{H}}^p$ is called the *shape matrix* which is invariant under multiplication by a positive constant. It accounts for the local structure between the elements of the vectors without consideration about the norm of the matrix.

To ensure identifiability, ξ_t is assumed to be normalized either by the first diagonal element [Randles, 2000] ($\xi_t = \Sigma_t / (\Sigma_t)_{11}$), the trace [Ollila et al., 2012a] ($\xi_t = p \Sigma_t / \text{Tr}(\Sigma_t)$) or the determinant [Tatsuoka and Tyler, 2000] ($\xi_t = \Sigma_t / |\Sigma_t|^{1/p}$).

In practice, all choices are essentially equivalent with regards to covariance equality testing problem under mild regularity conditions as studied in [Hallin and Paindaveine,

³This is often referred as the texture parameter in the radar literature. We will use alternatively both terms when referring to this parameter.

2007]. We will consider the normalization by the trace in the chapters 3 and 4 in order to keep consistent with [Ollila et al., 2012a] which is a reference in the robust statistics literature: $\boldsymbol{\xi}_t = p\boldsymbol{\Sigma}_t/\text{Tr}(\boldsymbol{\Sigma}_t)$.

In Chapter 5, we will see that the determinant constraint yields some simplification in Riemannian geometry, which explains why we will shift to the determinant normalization.

Some properties of the elliptical distributions are provided below:

- **Stochastic representation:** if $\mathbf{x} \sim \mathbb{CE}(\mathbf{0}_p, g, \boldsymbol{\Sigma})$, we can write \mathbf{x} as:

$$\mathbf{x} \stackrel{d}{=} \mathcal{R}\boldsymbol{\Sigma}^{1/2}\mathbf{u}, \quad (1.7)$$

where $\mathcal{R} \triangleq \sqrt{Q}$ is a non-negative real random variable called the modular variate which is independent of the complex random vector \mathbf{u} having a uniform distribution on \mathbb{CS} denoted as $\mathcal{U}(\mathbb{CS}^p)$. The p.d.f of the modular variate is given by:

$$p_Q(Q) = \mathfrak{m}_{p,g}^{-1} Q^{p-1} g(Q), \quad (1.8)$$

which is also the p.d.f of the quadratic form $\mathbf{x}^H \boldsymbol{\Sigma}^{-1} \mathbf{x}$.

This representation can be related to the sub-family of distributions called Complex Compound-Gaussian (CCG) or Spherically Invariant Random Vectors (SIRV) [Yao, 1973]. They are families of distributions of the form:

$$\mathbf{x} \sim \sqrt{\tau} \mathbf{z}, \quad (1.9)$$

where $\sqrt{\tau}$, called the texture, follows a distribution on \mathbb{R}^+ and \mathbf{z} is random vector distributed as a complex Gaussian distribution. Depending on the model assigned to the texture, a different elliptical distribution ensue.

- **Projection on \mathbb{CS}^p :** Let $\mathbf{x} \sim \mathbb{CE}(\mathbf{0}_p, g, \boldsymbol{\Sigma})$ and define

$$\mathbf{z} = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}. \quad (1.10)$$

The self-normalized vector \mathbf{z} is said to have a Complex Angular Elliptical distribution which is denoted as $\mathbb{CAE}(\mathbf{0}_p, \boldsymbol{\Sigma})$. Its p.d.f is given by:

$$p_{\mathbf{z}}^{\mathbb{CAE}}(\mathbf{z}; \boldsymbol{\xi}) = \mathfrak{S}_p^{-1} |\boldsymbol{\xi}|^{-1} (\mathbf{z}^H \boldsymbol{\xi}^{-1} \mathbf{z})^{-p}. \quad (1.11)$$

Note that this p.d.f is defined for vectors on \mathbb{CS}^p and not the standard set \mathbb{C}^p which means that the corresponding measure is not the standard Lebesgue measure but one on the unit sphere. Since we normalized the observations, the matrix $\boldsymbol{\xi}$ consists in solely the shape matrix while the scale is lost.

Examples of Elliptical distributions:

- **The complex t -Distribution:** A vector \mathbf{x} is said to have the complex multivariate t -distribution with ν degrees of freedom ($0 < \nu < \infty$) if it follows an elliptical distribution with density generator:

$$g(t) = (1 + 2t/\nu)^{-(2p+\nu)/2}. \quad (1.12)$$

The CCG representation is obtained for $\tau =_{\text{d}} \nu/x$ where $x \sim \chi_{\nu}^2$ and χ_{ν}^2 is the Chi-squared distribution with ν degrees of freedom.

A limit case is the complex Cauchy distribution, obtained using $\nu = 1$, for which the second order moment does not exist. The complex Gaussian distribution is a limit case when $\nu \rightarrow \infty$.

- **The K-Distribution:** A vector \mathbf{x} is said to have the complex multivariate K-distribution with texture shape $\nu > 0$ if it follows an elliptical distribution with density generator:

$$g(t) = t^{(\nu-p)/2} K_{\nu-p}(2\sqrt{\nu t}), \quad (1.13)$$

where $K_{\ell}(\cdot)$ denotes the modified Bessel function of the second kind of order ℓ . The CCG representation is obtained when τ follows a Γ -Distribution with shape parameter ν and scale parameter $\theta = 1/\nu$. The complex Gaussian distribution is obtained as a limit case when $\nu \rightarrow \infty$.

- **The generalized Gaussian Distribution:** A vector \mathbf{x} is said to have the complex multivariate generalized distribution with exponent $s > 0$ and scale $b > 0$ if it follows an elliptical distribution with density generator:

$$g(t) = \exp(-t^s/b). \quad (1.14)$$

The CCG representation is obtained when τ follows a Γ -Distribution with shape parameter $\nu = p/s$ and scale parameter $\theta = b$. The complex Gaussian distribution is obtained when $s = 1$ and for $s = 1/2$ we obtain the Laplace distribution.

Note: The elliptical and CCG distributions are probability models which are in the scope of the robust statistical literature initiated by works such as [Maronna, 1976a, Yohai, 1974, Martin and Pierre,]. 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 the context of this thesis, when its statistical properties are independent of the density generator function.

1.3.5 The deterministic compound-Gaussian model

Following the compound-Gaussian model of (1.9) defined earlier, it is possible to consider the texture as a deterministic value which varies at each observation. This model, which we call deterministic compound-Gaussian model, assumes a homogeneity of the shape matrix and a heterogeneity of the scale for each observation on the local spatial windows. Its parameters include deterministic scale parameters $\mathbf{T} = [\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_T]$ and set of shape matrices⁴ $\{\boldsymbol{\xi}_t | t \in [1, T]\}$. The model is written as follows:

$$\mathbf{x}_k^t \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, \tau_k^t \boldsymbol{\xi}_t), \quad (1.15)$$

where τ_k^t is the k -th element of vector $\boldsymbol{\tau}_t$.

Note: This model does not strictly follow the parametrisation defined in eq. (1.6) since here each observation has a different scale parameter while in the parametrisation only one for each date would be defined. However the same constraints assigned to the different set of parameters the set of parameters. Namely, $\forall (k, t) \in [1, N] \times [1, T], \tau_k^t \in \mathbb{R}^+$ and

⁴In the sense that they are normalized by the trace.

$\forall t \in [1, T]$, $\text{Tr}(\boldsymbol{\xi}_t) = p$. That is why, we will refer them as scale and shape parameters as well.

The interest of choosing deterministic scale parameters in this model is to be able to generalize the Gaussian model without adding any prior on the distribution of the scale parameter.

1.4 CHANGE DETECTION IN SAR IMAGE TIME SERIES

In this section, we briefly give an overview of change detection techniques relative to the SAR image time series. We hereby present a broad overview of recent techniques observed in the SAR literature. Among them, it is worth detailing some methodologies relying on a statistical framework since they consist in an interesting approach to tackle the noise inherent to SAR data. Then we focus on change detection based on covariance equality testing approach, which has been the main interest of the current thesis.

1.4.1 An overview of change detection techniques for SAR images

Change detection has been a popular subject of study for recent decades. Remote sensing, in peculiar, has attracted a plethora of scholars due to the various applications to military (activity monitoring) or civil (geophysics, the study of global warming, etc) applications. Thus literature on the subject is dense and a variety of methodologies are available⁵.

For example, IEEEExplore database registers more than 1,000 papers on the sole subject of SAR change detection. It is then impossible to discuss all methodologies while keeping a clear picture. Indeed, these various approaches are based on different modalities, considers various input data and output different kind of results. In this case, we will choose to classify general families of methodologies while detailing more precisely the approaches that are closely related to the work undertaken in the present thesis.

General process

Broadly speaking, as illustrated in Figure 1.15, 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. Depending on the nature of the data (single sensor, multiple sensors for example), various methodologies exist [Scheiber and Moreira, 2000, Bentoutou et al., 2005, Sansosti et al., 2006]. This step is critical in the whole process since a misaligned time series will result in many false alarms.

Various methodologies also consider a denoising step in which the speckle noise is reduced thanks to filtering techniques [Achim et al., 2003, Foucher and Lopez-Martinez, 2014].

Lastly, features can be selected from the images in order to obtain a more concise or informative representation for the change detection purpose. As discussed earlier, the choice of spectro-angular features can yield a better representation than the raw monovariate data. Other possibilities include Markov fields [Wang et al., 2013] or PCA [Yousif and Ban, 2013] among others.

⁵Notably, see [Hussain et al., 2013] or [Hecheltjen et al., 2014] for an overview.

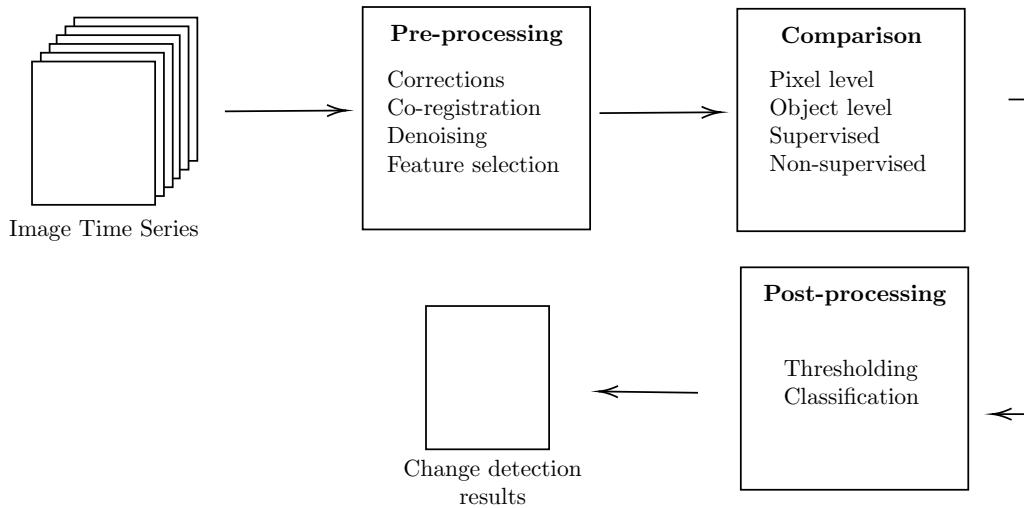


Figure 1.15: General procedure for a change detection methodology.

- A comparison step in which the features at each date are compared among themselves. Many techniques exist based on various principles. We will delay the overview of those methods to the next subsection in order to give a more comprehensive description since this step has been the main subject of study of the present thesis. At this point several outputs can be obtained: a value corresponding to the amplitude of change or a list of labels corresponding to various classes.
- A post-processing step which varies depending on the methodology used for the comparison phase. It can either correspond to a thresholding [Bruzzone and Prieto, 2000, Kervrann and Boulanger, 2006] or involves machine learning classification algorithms [Gong et al., 2016].

Comparison step methodologies

In order to distinguish the various approaches [Hussain et al., 2013] proposed to differentiate pixel level methodologies from object level ones⁶:

- Object level methodologies consider the object⁷ as the basic unit. As such, the detection can be done as a function of spatial coordinates (sub-pixel level).
- Pixel-level methodologies consider that the pixel is the atomic analytical unit for the detection mostly without considering the spatial context. The work presently done in this thesis considers the pixel level approach.

For pixel level methodologies, two types of methodologies exist:

⁶Another distinction made by [Hussain et al., 2013] consists of data mining methods such as [Petitjean et al., 2010] which considers pattern matching methods from computer science literature in order to analyze the time series. But the methodology involves computer science literature on database searching which does not enter the field of signal processing methodologies.

⁷An object is an abstraction which can be constituted of various concepts such as texture, geometric shape, relationships with neighbors among others. An overview of those methods can be found in [Chen et al., 2012].

- Supervised methodologies: in this case, machine learning techniques such as Support Vector Machine (SVM) [Jia et al., 2014] or deep learning [Gong et al., 2016] are used on an extensive database of labeled samples in order to obtain a decision rule to apply for other non labeled datasets.
- Non-supervised methodologies: in this case, no secondary labeled data is assumed to be available which is often the case in SAR problems. Indeed, since these types of images usually span a very large area, it is very difficult to obtain reliable ground truth. As a consequence, the unsupervised methodologies are more attractive from a practical standpoint. We considered this paradigm.

Finally, non-supervised methods at pixel level decompose into:

- Model-free methodologies: no probability model is assigned to the data for change detection. one of the most popular approaches consists in the Coherent Change Detection (CCD) [Preiss and Stacy, 2006] where a coherence measure is computed between pair of complex images in order to highlight zones with dissimilarities⁸. This technique is popular since it considers a difference in phase which allows detecting very fine changes in the scene. There is however a limitation due to the well-known coherence loss over time. Moreover, since it relies mostly on the phase, the methodology yields a significant number of false alarms for zones such as forests. Another popular approach is the log-ratio operator [Gong et al., 2012, Garzelli and Zoppetti, 2017] on local neighborhoods around the test pixel for a pair of two amplitude only images. The idea behind these methods is to reduce the effect of speckle noise by transforming into an additive noise through the logarithm function. Other approaches have considered PCA [Yousif and Ban, 2013] or wavelet analysis [Atto et al., 2016]. The advantage of those approaches is that they do not depend on a probability model which can be inaccurate sometimes. The main drawback, however, concerns the choice of a threshold value for the detection which cannot be based on the statistic of the distance function.
- Statistical based methodologies: in this case, the methodology relies on a probability model in order to infer about changes⁹. Statistical methods can rely on a model with associated parameters in which case they are referred to as parametric methodologies. Non-parametric methods consider statistical tools without considering parameters [Aiazzi et al., 2013] or Bayesian models [Prendes et al., 2015] with a prior on the distribution of those parameters.

Note: An alternative approach called post-classification scheme, consists in detecting changes by first classifying the time series and then comparing the labels over time to detect the changes. This methodology can be at pixel or object level, involves supervised training and can be based or not on statistical approaches. It was not mentioned up to this point in order to avoid unnecessary complication since the literature of SAR classification is also prolific.

Statistical change detection based on a parametric model

We will consider here parametric methods which rely on a probability model on the data associated with parameters. In such setup, the change detection can be seen as obtaining a

⁸There also exist variant of this approach based on a Gaussian model to derive a distance [Barber, 2015, Novak, 2005].

⁹Some works also considered possibilities models based on fuzzy logic theory to design distances adapted to SAR images [Carincotte et al., 2006, Lesniewska-Choquet et al., 2017].

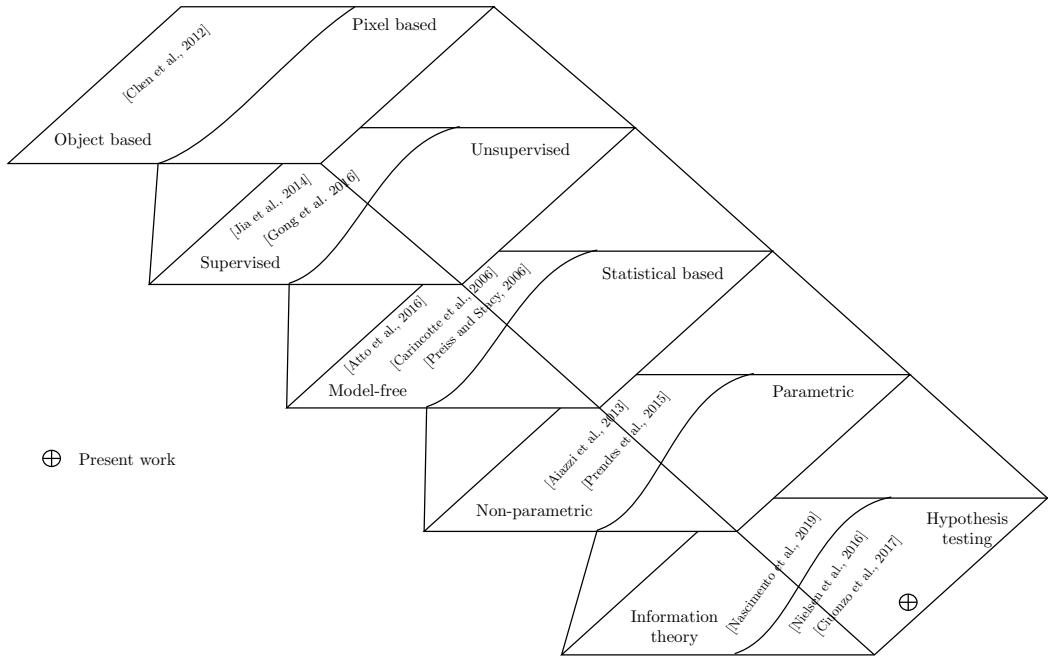


Figure 1.16: Summary of various change detection methodologies.

distance between the distributions of the data over time such that if the distributions over time coincide, the value is low and high otherwise. These methodologies have been widely popular in recent years thanks to the preliminary work of [Conradsen et al., 2003] which showed good results for analyzing changes for polarimetric SAR data.

Literature on the subject consists in two distinct approaches:

- Information-theoretic measures [Atto et al., 2013, Inglada and Mercier, 2007, Nascimento et al., 2010, Nascimento et al., 2019, Ratha et al., 2017]: the idea behind this approach is to consider distances from information theory to compare the data over time. Suppose that the local data at date $t = 1$ is modelled by the distribution $p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_1)$ and at date $t = 2$ by $p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_2)$, it is possible to compare the similarity between these two models thanks to well-known distances such as the Kullback-Leibler distance (which is a symmetrized version of the Kullback-Leibler divergence):

$$\delta_{KL} = \frac{1}{2} \left(\int \log \frac{p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_1)}{p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_2)} p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_1) d\mathbf{x} + \int \log \frac{p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_2)}{p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_1)} p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_2) d\mathbf{x} \right). \quad (1.16)$$

In practice, since the parameters are unknown they are estimated using, for example, the Maximum Likelihood Estimator (MLE). The limitation of this methodology lies in the fact that there is no general result allowing to obtain a decision threshold for any pair of distributions $(p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_1), p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_2))$. Indeed, although [Frery et al., 2014] showed that information theoretical distances on covariances do have an asymptotic χ^2 distribution under the null hypothesis, there is no result such as the Constant False Alarm Rate (CFAR) property at finite distance.

- Hypothesis testing schemes [Nielsen et al., 2016, Barber, 2015, Carotenuto et al., 2015, Ciuonzo et al., 2017]: the idea is to define a detection problem for choosing

between two alternatives in which either the parameter stays constant or change over time. For multivariate SAR images, the test can be done on the covariance parameter. This methodology, which we will detail later, has interesting advantages with regards to the previous one: on the one hand, it allows to consider the whole time series while in the previous case, the methodologies apply on a pair of images. On the other hand, literature on hypothesis testing is very prolific and many theoretical results, notably on thresholding can be applied here. As a consequence, we will consider this approach in the development of this thesis.

A summary of the related litterature in the form of an illustration can be found in Figure 1.16.

1.4.2 Defining the problem of covariance equality testing

The test of equality of covariances, sometimes referred as homogeneity of covariances, can be written as a binary hypothesis testing problem of the following form:

$$\begin{cases} H_0 : \theta_1 = \dots = \theta_T = \theta_0 & \& \Phi_1 \neq \dots \neq \Phi_T, \\ H_1 : \exists (t, t') \in [1, T]^2, \theta_t \neq \theta_{t'} & \& \Phi_1 \neq \dots \neq \Phi_T \end{cases} . \quad (1.17)$$

where θ is either a covariance matrix, a scale parameter or a shape matrix. The nuisance parameters which we do not want to test are gathered in the set Φ . These nuisances parameters are either the scale parameter, the shape matrix or the empty space, depending on the definition of θ .

Notes:

- When testing homogeneity of the shape matrix in the Gaussian context, the test can be seen as equivalent to the proportionality testing which is written in the following form (for $T = 2$):

$$\begin{cases} H_0 : (\Sigma_1 \propto \Sigma_2) \equiv (\Sigma_1 = \alpha \Sigma_2) \\ H_1 : (\Sigma_1 \not\propto \Sigma_2) \equiv (\Sigma_1 \neq \alpha \Sigma_2) \end{cases} . \quad (1.18)$$

This problem has been considered in works such as [Federer, 1951, Liu et al., 2014a] or more recently in [Taylor et al., 2017] for radar applications.

- The test of covariance homogeneity is well-known for its use in the Multivariate Analysis Of Variance (MANOVA) technique found in numerous statistical references [Wilks, 1932, Roy, 1946]. This technique relies on the homogeneity of covariances among group of datasets in order to check the statistical significance of comparing the means of those datasets.

1.4.3 On the hypothesis problem and test statistics

In order to decide between the two alternatives (1.17), it is necessary to obtain a test statistic compared to a threshold value.

A statistic $\hat{\Lambda}$ is a function of the input data $\mathbb{W}_{1,T} = \{\mathbf{X}_1, \dots, \mathbf{X}_T\}$ which is expected to be high when hypothesis H_1 is true while being low under H_0 hypothesis. In this way, its value gives an insight about the probability of H_1 to be true.

Since it is a function of random variables, this test statistic is a random variable as well. It is then possible to consider the distribution of the statistic in both regimes as illustrated in Figure 1.17. Given the distribution under H_0 , it is possible to select a threshold λ in

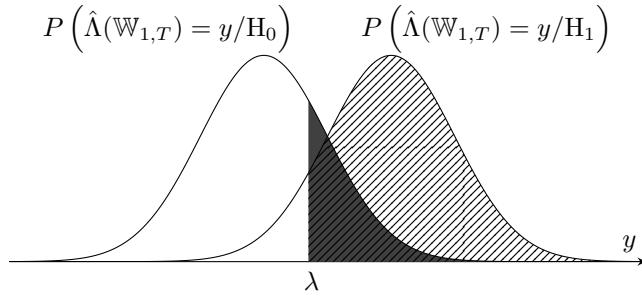


Figure 1.17: Illustration of false alarm/good detection probabilities for a given threshold λ . The dark area corresponds to the probability of false alarms while the line filled area correspond to the probability of good detection.

such a way that the probability of the statistic being superior to the threshold is fixed. This probability is usually referred as the probability of false alarms¹⁰ (or type I errors) which are a situation where the statistic is greater than the threshold is not due to the data being in H_1 hypothesis. Conversely, the threshold is also linked to the probability of false negatives (or type II errors) that are events in which a H_1 regime is falsely rejected. The trade-off between these two types of errors is dependent on the threshold value used. This paradigm of detection is, however, not applicable when the distribution of the statistics under null hypothesis varies with the parameters.

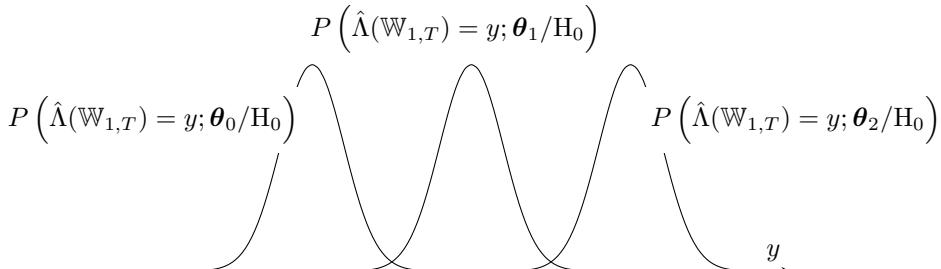


Figure 1.18: Illustration of CFAR property: if the distribution varies when the parameter of the distribution changes, it is impossible to guarantee a false alarm rate.

A critical concept to guarantee a false alarm rate is the Constant False Alarm Rate (CFAR) property. This property is obtained when the distribution of the statistic under the null hypothesis is independent of the parameters of the problem. As illustrated in Figure 1.18, if the distribution of the statistic varies along with the parameter of the problem, the threshold becomes a function of this parameter. Since in many detection applications the parameter under null hypothesis is not known, it is not possible to guarantee a false alarm rate.

There are many possibilities to define a test statistic for a given problem, as will be discussed in subsection 1.4.4. Under some conditions, it is possible to obtain the Uniformly Most Powerful (UMP) test statistic which maximizes the probability of detection¹¹ at fixed

¹⁰It is often referred as the significance level in the statistical literature.

¹¹Also known as the power of the test in the statistical literature.

false alarm rate. Unfortunately as discussed in [Lehmann, 2009] and [Ciuonzo et al., 2017], this statistic does not exist for the covariance equality testing problem.

Nonetheless, various test statistics can be obtained thanks to standard derivation techniques, which we will detail in the next subsection.

1.4.4 On the GLRT technique and known alternatives

Since we will rely extensively on the GLRT technique to obtain test statistics, we remind here its core principle and interesting statistical properties. Then we present briefly some other alternative techniques and explain why they were not considered in our work.

Given a hypothesis testing problem of the form presented in (1.17), the GLRT is formulated as follows:

$$\hat{\Lambda} = \frac{\max_{\theta_1, \dots, \theta_T, \Phi_1, \dots, \Phi_T} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}; \theta_1, \dots, \theta_T, \Phi_1, \dots, \Phi_T)}{\max_{\theta_0, \Phi_1, \dots, \Phi_T} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}; \theta_0, \Phi_1, \dots, \Phi_T)} \stackrel{H_1}{\gtrless} \lambda. \quad (1.19)$$

This test differs from the standard likelihood ratio test [Kay, 1998] in the fact that the parameters being tested are assumed to be unknown. These unknown parameters are thus estimated using the prior that the model is either under H_0 or H_1 hypothesis. Note that this also differs from a two-step methodology which would consist in plugging estimates of the parameters in the test statistic obtained assuming the assumption of known parameters¹². This technique is very popular due to interesting statistical properties such as invariance of the statistic with regards to the parameter space [Kay and Gabriel, 2003]. The most notable result which is of interest is known as the Wilks Theorem and is given in the following proposition:

Proposition 1.4.1. (Wilks Theorem) *Assume that the joint distribution of $\{\mathbf{X}_t | 1 \leq t \leq T\}$ depends on K unknown parameters under H_1 and that, under H_0 , the joint distribution depends on K_0 unknown parameters. Let $\nu = K - K_0$. Then, under some regularity conditions ($K > K_0$), when the null hypothesis is true, the distribution of the statistic $2 \log(\hat{\Lambda})$ converges to a χ^2_ν distribution as the sample size $N \rightarrow \infty$, i.e when H_0 is true and N is large, we have:*

$$2 \log(\hat{\Lambda}) \underset{\text{approx}}{\sim} \chi^2_\nu \quad (1.20)$$

Proof. First demonstrated in [Wilks, 1938]. The result is a consequence of the asymptotic consistency property of the maximum likelihood estimation. \square

This proposition allows to obtain a confidence region to guarantee a false alarm rate without having to resort to extensive Monte-Carlo simulations, which is interesting. We must note however that this is only an asymptotic result which means that if the sample size N is small, the result will no longer apply.

Many alternatives exist to the GLRT¹³, from which the most well known are the Rao statistic [Radhakrishna Rao, 1948], Terrell gradient statistic [Terrell, 2002] and the Wald

¹²In target detection schemes, the well known Adaptive Normalized Matched Filter (ANMF) is an example of a two-step statistic.

¹³See [Ciuonzo et al., 2017] for an exhaustive list in the context of covariance testing in Gaussian context.

statistic [Wald, 1943]. These tests also enjoy asymptotic χ^2 distributions under null hypothesis when the sample size is large and are thus interesting alternatives to the GLRT¹⁴. These tests require however most of the time heavier derivations. For instance, the Rao and Wald statistics require the inversion of the Fisher Information Matrix (FIM) which is obtained from the computation of the Hessian of the likelihood function while the GLRT only need to compute the gradient in order to optimize when the likelihood function is convex. Thus, since they share similar properties, we will rely on the GLRT technique which is the simplest one.

Note: It is also possible to obtain a statistic using the principle of the maximal invariant statistic as done in [Ciuonzo et al., 2017]. The idea is to find a group of transformations for the parameter space which leaves the detection problem invariant. A maximal invariant statistic is a statistic which is invariant by the same group of transformation and yields the least data compression (due to the transformation) compared to all other invariant statistics. This principle allows us to derive statistics by analyzing the invariances of the problem and results in a statistic with Constant False Alarm Rate (CFAR) property while the property is not guaranteed for other tests. In our detection problem, we did not consider such a technique since it requires a fine analysis of the complex non-Gaussian detection problem in order to find a suitable group of transformations and the maximal invariant statistic. Moreover, there is no general result such as Wilks theorem in this case as well.

Once a statistic of decision with its associated null hypothesis distribution, it is possible to test a single hypothesis. When a decision is based on multiple hypothesis testing schemes, some additional issues have to be considered, as we will see in the next subsection.

1.4.5 A note on the false discovery rate and familywise error rate

When considering a large number of multiple hypothesis tests, it is necessary to consider the problems associated with false alarms. Indeed, it is well known that when a decision is made according to n hypothesis tests, the significance level of the procedure does not equal that of the tests. For instance, it is necessary to consider the family wise rate error which is defined as:

$$\alpha_{\text{FWER}} = 1 - (1 - \alpha)^n, \quad (1.21)$$

where α is the chosen significance level for the n tests. From this, it is clear that the significance level grows as the number of tests done increases. In order to tackle this issue, the well known Bonferroni correction [Dunn, 1961] or the false discovery rate [Benjamini and Hochberg, 1995] help to reduce the number of false alarms in such setups.

In our case, we have to consider two issues: we consider a test on each pixel which considers several data over time and we consider as many tests as the number of pixels of the image.

For the first case, our detection is based on a single test of detection as opposed to multiple detection tests between pair of images. This approach considers the two alternatives as either no change at all in the series or any possible change (either one or more). Thus since only one test is done, there is no need to consider the temporal aspect.

For the second case, the decision is taken independently for each pixel of the image. However, depending on the image size a high number of tests are done. For example, say that an image is constituted of 1000×1000 pixels, 10^6 tests are done in a pixelwise approach.

¹⁴It has also been shown that in [De Maio et al., 2010], that under mild regularity conditions, the GLRT, Rao and Wald statistics are asymptotically equivalent

If we fix a false alarm rate to 10^{-3} , there is expected to be at least 1000 pixels corresponding to a false alarm which is arguably a high number. To reduce these false alarms, it is possible to correct the detection threshold according to the false discovery rate as presented in [Efron, 2007] which provides a correction based on the number of false discoveries among the number of total H_1 outcomes. In that regard, the Benjamini–Hochberg procedure [Benjamini and Hochberg, 1995] can be implemented to decide for each pixel if it has to be counted as a detection. This comes however at a cost of detection performance. Thus the decision to use the correction depends on the type of data considered: if we expect a few changes in a globally non-changing image, adapting the threshold is a necessity while in a situation where many pixels can change, the actual detection performance is important.

1.4.6 Results on the equality of covariance in Gaussian context

Many works have considered the testing of covariance homogeneity problem [Nagao, 1973, Schott, 2001, Anderson, 2003, Hallin and Paindaveine, 2009] and many statistics have been proposed. The case for $T = 2$ has been especially studied for change detection purposes [Conradsen et al., 2001, Carotenuto et al., 2015]. Recently, [Ciuonzo et al., 2017] did a comparative study and showed that many test statistics for the covariance homogeneity are statistically equivalent and reduce to the following decision rules:

- the GLRT statistic:

$$\hat{\Lambda}_G = \frac{\left| \hat{\Sigma}_0^{\text{SCM}} \right|^{TN}}{\prod_{t=1}^T \left| \hat{\Sigma}_t^{\text{SCM}} \right|^N} \stackrel{H_1}{\gtrless} \lambda, \quad (1.22)$$

where:

$$\forall t, \hat{\Sigma}_t^{\text{SCM}} = \frac{1}{N} \sum_{k=1}^N \mathbf{x}_k^t \mathbf{x}_k^{t \text{ H}} \text{ and } \hat{\Sigma}_0^{\text{SCM}} = \frac{1}{T} \sum_{t=1}^T \hat{\Sigma}_t^{\text{SCM}}. \quad (1.23)$$

- the t_1 statistic which is obtained from Terrell or Rao tests:

$$\hat{\Lambda}_{t_1} = \frac{1}{T} \sum_{t=1}^T \text{Tr} \left[\left(\left(\hat{\Sigma}_0^{\text{SCM}} \right)^{-1} \hat{\Sigma}_t^{\text{SCM}} \right)^2 \right] \stackrel{H_1}{\gtrless} \lambda. \quad (1.24)$$

- the Wald statistic:

$$\begin{aligned} \hat{\Lambda}_{\text{Wald}} = & N \sum_{t=2}^T \text{Tr} \left[\left(\mathbf{I}_p - \hat{\Sigma}_1^{\text{SCM}} (\hat{\Sigma}_t^{\text{SCM}})^{-1} \right)^2 \right] \\ & - q \left(N \sum_{t=1}^T (\hat{\Sigma}_t^{\text{SCM}})^{-T} \otimes (\hat{\Sigma}_t^{\text{SCM}})^{-1}, \text{vec} \left(\sum_{t=2}^T \boldsymbol{\Upsilon}_t \right) \right) \stackrel{H_1}{\gtrless} \lambda, \end{aligned} \quad (1.25)$$

where

$$\boldsymbol{\Upsilon}_t = N \left((\hat{\Sigma}_t^{\text{SCM}})^{-1} - (\hat{\Sigma}_t^{\text{SCM}})^{-1} \hat{\Sigma}_1^{\text{SCM}} (\hat{\Sigma}_t^{\text{SCM}})^{-1} \right). \quad (1.26)$$

The Gaussian GLRT can be interpreted easily as a ratio of volumes. Indeed, it is well known that the determinant of Positive Definite Hermitian (PDH) matrix is equal to the volume of the ellipsoid whose axes correspond to the eigenvalues of the matrix as illustrated in Figures 1.19 and 1.20. Then, the ratio between the determinants of respectively

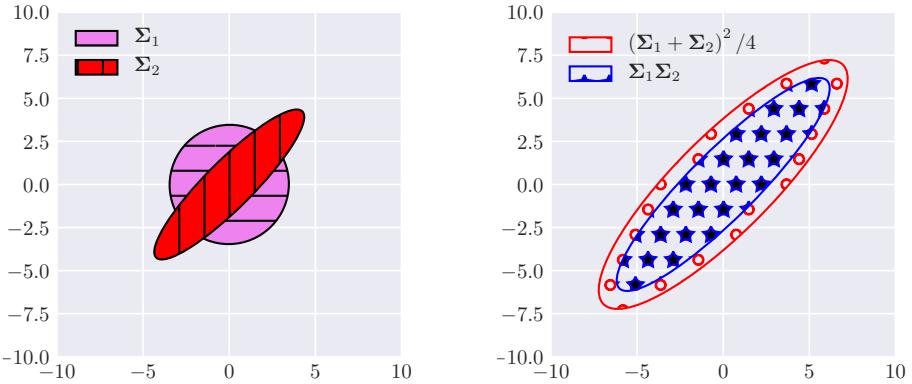


Figure 1.19: Representation of Gaussian GLRT for $p = 2$, $T = 2$ when matrices are different. Left: Ellipsoids representing the covariance matrices. Right: Ellipsoids representing the arithmetic and geometric mean.

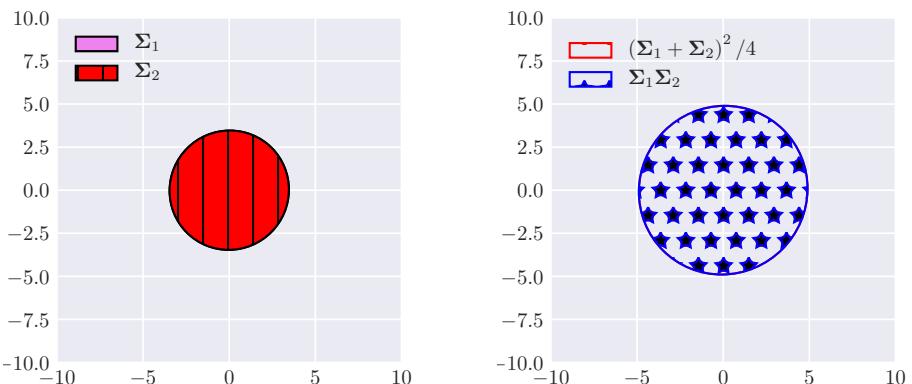


Figure 1.20: Representation of Gaussian GLRT for $p = 2$, $T = 2$ when matrices are equal. Left: Ellipsoids representing the covariance matrices. Right: Ellipsoids representing the arithmetic and geometric mean.

the arithmetic and the geometric means of the matrices can be seen as the ratio of their respective ellipsoid volume. The ratio is expected to be one when the matrices are equal and greater than one otherwise due to the well known arithmetic/geometric mean inequality. Concerning the other two statistics, there is no trivial interpretations such as this one.

It can be noted that in practice the Wald statistic is numerically challenging for high-dimensional vectors. Indeed, since it involves the inverse of a Kronecker product, the computational complexity grows much faster compared to the other two methods.

Concerning theoretical analysis, [Anderson, 2003] has considered the distribution of the GLRT for the real case under the null hypothesis based on a theory of asymptotic expansion. Later [Conradsen et al., 2003] adapted the result to work in the complex case:

Proposition 1.4.2. *Under hypothesis H_0 , the probability of false alarms of the GLRT for the covariance homogeneity under Gaussian model with $\boldsymbol{\theta} = \boldsymbol{\Sigma}$ and $\boldsymbol{\Phi} = \emptyset$ can be obtained*

using the following formula:

$$\begin{aligned} P\left\{2\rho \log(\hat{\Lambda}_G) \leq z\right\} &\approx P\left\{\chi_{f^2}^2 \leq z\right\} + \omega_2 [P\left\{\chi_{f^2+4}^2 \leq z\right\} - P\left\{\chi_{f^2}^2 \leq z\right\}] \\ f &= (T-1)p^2, \rho = 1 - \frac{(2p^2-1)}{6(T-1)p}\left(\frac{T}{N}-\frac{1}{NT}\right), \\ \omega_2 &= \frac{p^2(p^2-1)}{24\rho^2}\left(\frac{T}{N^2}-\frac{1}{(NT)^2}\right) - \frac{p^2(T-1)}{4}\left(1-\frac{1}{\rho}\right)^2. \end{aligned} \quad (1.27)$$

This approximation allows one to obtain detection threshold without having to resort to expensive Monte-Carlo simulations. Although the result is asymptotic, the relevance of the approximation is obtained faster than the one obtained using the Wilks theorem as shown in Figure 1.21. This result is a better approximation than the one obtained through Wilks theorem. The other test statistics mentioned in (1.24) and (1.25), have χ^2 approximations under the null hypothesis resulting from asymptotic properties of Terell and Wald methodologies.

1.4.7 An extension to the Elliptical case using local Normal Asymptotic theory

Study of the covariance homogeneity test under non-Gaussian context has been an active area of research for the past decades and many developments have been carried to bootstrap the Gaussian GLRT in order to make it more robust to non-Gaussianity. A classic reference about this subject consists of [Muirhead and Waternaux, 1980] in which the authors provide an in-depth study of the problem of turning standard Gaussian tests about covariance matrices into pseudo-Gaussian ones remaining valid under elliptical densities. In this reference, the authors conclude that turning the Gaussian GLRT into a pseudo-Gaussian test is an arduous task for which no solution had been found yet.

Later, under elliptical distributions assumption, [Yanagihara et al., 2005] studied the distribution of the Bartlett test defined as:

$$\hat{\Lambda}_B = \frac{\left|\hat{\Sigma}_0^{SCM}\right|^{T(N-1)}}{\prod_{t=1}^T \left|\hat{\Sigma}_t^{SCM}\right|^{N-1}} \stackrel{H_1}{\gtrless} \lambda, \quad (1.28)$$

which is a variant of the Gaussian GLRT [Bar, 1937]. The following result has been established:

Proposition 1.4.3. *Assuming an elliptical model with an equality of kurtosis between the T groups of observations (homokurticity), the asymptotic distribution of $\hat{\Lambda}_B$ for the real case:*

$$(1 + \kappa_p) \left\{ \left[1 + \frac{p\kappa_p}{2(1 + \kappa_p)} \right] \chi_{T-1}^2 + \chi_{(T-1)(p-1)\frac{p+2}{2}}^2 \right\}, \quad (1.29)$$

where κ_p is the common kurtosis of the T groups of samples.

This result generalizes the χ^2 approximation of the test statistic since we obtain the approximation for a Gaussian context (the kurtosis are null). From this result, several attempts have been made to correct the Bartlett statistic to improve robustness to non-Gaussianity[Zhang and Boos, 1992, Goodnight and Schwartz, 1997, Zhu et al., 2002]. These

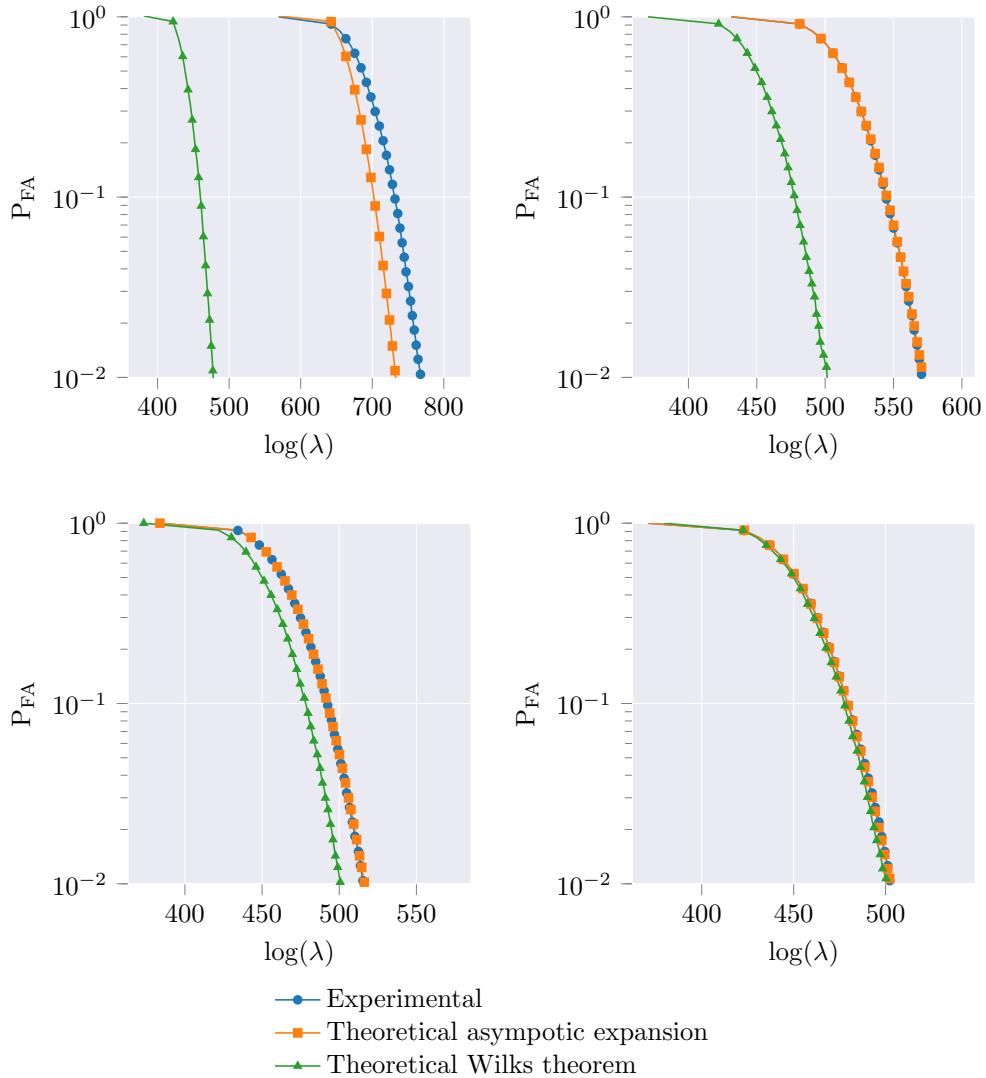


Figure 1.21: Experimental validation (10^4 Monte Carlo trials) of P_{FA} approximation of eq. (1.27) with $p = 15$, $T = 5$. Top-Left: non-asymptotic regime ($N = 20$). Top-right: non-asymptotic regime ($N = 50$). Bottom-left: asymptotic regime ($N = 200$). Bottom-right: asymptotic regime ($N = 1500$).

attempts aim at correcting the statistic so that it keeps the CFAR property under the context of elliptical distributions.

On the other hand, based on Wald statistic [Schott, 2001] proposed a robustification to handle non-Gaussianity. Following this approach, Hallin and Paindaveine proposed in [Hallin and Paindaveine, 2009] to use the Local Asymptotic Normality theory of Le Cam [LeCam, 1960] in order to design pseudo-Gaussian tests which share the same asymptotic results than the Wald and Gaussian GLRT while being robust and optimal (in Le Cam sense) to non-Normality.

As such the optimal test obtained is given by:

Proposition 1.4.4. *Pseudo-Gaussian test (homokurtic real case)*

$$\begin{aligned} \mathcal{Q}_{\mathcal{N}} &= 2 \sum_{1 \leq t < t' \leq T} \mathcal{Q}_{\mathcal{N};t,t'}, \text{ where} \\ \mathcal{Q}_{\mathcal{N};t,t'} &= \frac{1}{4(1 + \hat{\kappa}_p)} \left\{ \text{Tr} \left[(\hat{\Sigma}_0^{\text{SCM}})^{-1} (\hat{\Sigma}_t^{\text{SCM}} - \hat{\Sigma}_{t'}^{\text{SCM}})^2 \right] - \right. \\ &\quad \left. \frac{\hat{\kappa}_p}{(p+2)\hat{\kappa}_p + 2} \text{Tr}^2 \left[(\hat{\Sigma}_0^{\text{SCM}})^{-1} (\hat{\Sigma}_t^{\text{SCM}} - \hat{\Sigma}_{t'}^{\text{SCM}}) \right] \right\}, \\ \hat{\kappa}_p &= p(p+1)/2 \sum_{t=1}^T \sum_{k=1}^N d^4(\mathbf{x}_k^{(t)}, \hat{\Sigma}_0^{\text{SCM}}) - 1 \quad \text{and} \quad d(\mathbf{x}, \Sigma) = \|\Sigma^{-1/2}\mathbf{x}\|. \\ \text{Under } H_0, \text{ the statistic is distributed as } &\chi_{\frac{(T-1)p(p+1)}{2}}^2. \end{aligned} \quad (1.30)$$

Other tests having less restrictive assumptions about homokurticity have also been proposed in the same paper. The authors have also considered the case where the kurtosis is not a finite moment and proposed a rank-based test in [Hallin and Paindaveine, 2008].

These various works appear to be a good alternative to the Gaussian-based test for the change detection problem in SAR time series. However, tests on real images proved to be numerically unstable and did not yield satisfactory results compared to the Gaussian-based tests. Indeed, these tests do not have optimality at a finite distance which is the case in SAR change detection application. Moreover, the test assumes the computation of an estimate of the kurtosis which is known to be unstable if the number of samples is not high. The problem is that in this context where we have a sliding windows approach, in order to obtain good estimates that are well conditioned, the size of the windows has to be very large. Thus in practice, there were aberrant results (NaN or infinite values) in the value of the statistic on real images. Finally, the estimates of the covariance matrix is based on the SCM which is not robust to outliers. For those reasons, this approach was not pursued in the present thesis.

We will consider in the chapter 3 of this thesis the GLRT under elliptical distributions and deterministic compound-Gaussian, which to the best of our knowledge has not yet been considered in the literature¹⁵. The advantage of this methodology is that we can introduce robust estimates of the shape matrices in the statistic of decision while the SCM is always the one used in pseudo-Gaussian tests.

1.4.8 Some attempts of equality testing under deterministic compound-Gaussian model

Under this model, the estimation of ξ has been studied in [Tyler, 1987, Pascal et al., 2008a, Soloveychik and Wiesel, 2015, Drašković and Pascal, 2016, Gini and Greco, 2002]. When the textures are considered deterministic, an approximation in the form of a fixed-point estimator also known as Tyler estimator can be used:

$$\hat{\xi}_t^{\text{TE}} = \frac{p}{N} \sum_{k=1}^N \frac{\mathbf{x}_k^t \mathbf{x}_k^{t \text{ H}}}{\mathbf{x}_k^{t \text{ H}} \{\hat{\xi}_t^{\text{TE}}\}^{-1} \mathbf{x}_k^t}. \quad (1.31)$$

¹⁵For the elliptical case, it is mostly due to the impracticality of the approach which requires the knowledge of the density generator function. As a consequence, this single distribution based approach can hardly be seen as a robust method. On the other hand, the deterministic compound-Gaussian model allows adding more degrees of freedom to the problem without additional knowledge about the underlying elliptical distribution, which makes it a good candidate for a robust change detection methodology.

In [Formont et al., 2011], it was proposed for the case $T = 2$, to use this estimator in the statistic of eq. (1.22) in place of the SCM estimator in order to obtain a robust distance between covariance matrices. In fact, when considering this methodology, the statistic loses its CFAR matrix property. Indeed, this is caused by the normalisation that has to be imposed for the shape matrix ($\text{Tr}(\hat{\xi}_t) = p$): the ratio is not invariant when the estimates of covariance matrices are scaled. Since the normalisation is performed by scaling the estimates $\hat{\xi}_t$ by $p/\text{Tr}(\hat{\xi}_t)$, the statistic introduces a ratio of trace terms which is not CFAR.

Moreover, this methodology omits the scale parameters which may be useful for accounting changes. Indeed, since the shape matrices are normalised, the relative power between the images is *contained* in these scale parameters.

Another approach was considered in [Liu et al., 2014b], where estimates of both textures and covariance are plugged into a similarity measure for the case $T = 2$. In [Liu et al., 2011], a 2-step LRT plug-in approach was considered using deterministic compound-Gaussian model to detect a change in the shape matrix. However in these two works, the study was limited to the case of $T = 2$ and there was no theoretical study. Moreover, it is expected that plug-in approaches have poorer performance compared to the GLRT which considers the whole problem from the beginning. Notably in those approaches, the CFAR property is lost due to the normalization as well.

As a consequence, we will consider the GLRT methodology in the chapter 3 which has not yet been applied to the compound-Gaussian case.

1.5 CHANGE-POINT ESTIMATION IN TIME SERIES

In this section, we consider the problem of change-point detection/estimation. While the change detection problem developed in the previous section considers the sole problem of detecting all changes in the series, the change-point detection/estimation problem aims at obtaining at the same time a detection of the changes while also determining the time at which those changes have occurred. It is thus an alternative problem which brings more information about the changes in the time series with regards to basic change detection. This also means that the problem is more complex since the number of changes is also unknown (ranging from 0 to $T - 1$).

We stress here that we consider change-points as impulsive changes in the statistics of the data over time. This definition does not consider slowly changing series for which specific tools relying upon time series modeling are used¹⁶.

While the subject on change-points detection/estimation is well known in the statistical signal processing literature, the choice has been made to consider the methodology of [Conradsen et al., 2016] which relies on iterative hypothesis testing. This choice has been made since it allows to advantageously use the developments of the chapter 3 which focusses on this point. This does not mean however that this methodology yields the best performance from a practical standpoint.

The problem considered presently is the following:

Assume we have a time series of T ordered sets $\mathbf{X}_t = \{\mathbf{x}_i^t \in \mathbb{C}^p : 1 \leq i \leq N\}$. Assuming each set at date t follows a parametric model with parameter $\boldsymbol{\theta}_t$, find all the indexes $\{t_1, \dots, t_K\}$ so that $\forall 1 \leq i \leq K, \boldsymbol{\theta}_{t_{i-1}} \neq \boldsymbol{\theta}_{t_i}$. The number of change-points K is not known.

¹⁶See for example [Scharf, 1991].

1.5.1 Some techniques for estimating the change-point

We describe shortly some methodologies relative to change-point detection and estimation in time series. Far from being comprehensive, the aim here is to situate the methodology we will rely upon with regards to other existing literature¹⁷.

Broadly, two types of methodologies exist within the literature: the parametric and non-parametric approaches. In the first case, the distribution of the data must be known and inference can be done thanks to the estimations of parameters through time. For the other case, Bayesian inference is often used to introduce priors on the parameters seen as random variables and then consider the posterior probability density for inference.

Algorithm 1 CUSUM Algorithm

```

1: Initialize  $Z_0 = 0$ 
2: for  $t = 1, 2, \dots$  do  $Z_t = \max \left\{ 0, Z_{t-1} + \log \frac{p_{\mathbf{X}_t}(\mathbf{X}_t; \boldsymbol{\theta}_1)}{p_{\mathbf{X}_t}(\mathbf{X}_t; \boldsymbol{\theta}_0)} \right\}$ 
3:   if  $Z_t \geq h$  then  $\hat{t}_C = t$ 
4:   end if
5: end for
```

Concerning parametric methodologies, the Cumulative Sum Control Chart (CUSUM) algorithm [Page, 1954] is a sequential algorithm well known in industrial control context where a shift from an expected value is to be detected. The methodology consists in summing a ratio of the instantaneous log-likelihoods of the data being under models $p_{\mathbf{X}_t}(\mathbf{X}_t; \boldsymbol{\theta}_1)$ and $p_{\mathbf{X}_t}(\mathbf{X}_t; \boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_1$ is the parameter in case of a change and $\boldsymbol{\theta}_0$ the expected value of the parameter to control. The algorithm continues until its value exceeds a certain threshold value h . This process is summed-up in Algorithm 1. A major drawback of this approach concerns the fact that both $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ must be known to be applicable. Although methodologies exist when these parameters are unknown, they often suppose additional constraints [Willsky and Jones, 1976]. Moreover, the setting of the threshold h can be arduous depending on the probability model considered, which makes it impractical in the context of SAR change-point estimation.

Conradsen approach is similar in the way that its corresponds to a sequential hypothesis testing but differs from the fact that it does not consider the instantaneous log-likelihood ratio but an alternative GLRT which take into account the whole data before the time considered. This allows to set the threshold as a function of the false alarm rate but increases the complexity.

When it concerns Bayesian methodology, an interesting approach consists in defining binary variables r_t for each time such that:

$$r_t = \begin{cases} 1 & \text{If } \mathbf{X}_t \text{ is a change-point} \\ 0 & \text{otherwise} \end{cases} \quad (1.32)$$

By assigning a *Bernoulli* prior to these random variables, [Tourneret et al., 2003] developed a scheme to estimate the change-points for monovariate SAR image segmentation problem. A more recent approach under the same principle considering multivariate vectors has been introduced in [Harlé et al., 2016] which could be adapted to the context of change-point estimation in SAR images. This, however, moves away from the analysis on

¹⁷See [Basseville and Nikiforov, 1993, Brodsky and Darkhovsky, 2013, Gustafsson, 2000, Aminikhanghani and Cook, 2017] or the ph.d thesis of Flore Harlé [Harlé, 2016] for a more comprehensive overview.

covariance matrices obtained from a local spatial neighborhood which is the main subject of the present work and has thus not been considered.

The approach introduced by [Conradsen et al., 2016], which we will present in more detail hereafter, has the advantage to be directly generalized to non-Gaussian context thanks to the work done in Chapter 3 on covariance homogeneity testing. As a direct application, this methodology has been chosen to tackle the change-point estimation problem.

1.5.2 Joint detection and estimation of change-points for SAR images

Based on the methodology of [Conradsen et al., 2016], the change-point detection/estimation problem can be decomposed in two sub-problems consisting in:

- A detection problem: we first want to know if there is any change-point in the time-series. If the detection process does not yield a detection, the scheme can end here.
- An estimation problem: when a detection has been made, the change-point location (time-stamp) has to be estimated.

Detection problem: Binary Hypothesis testing

The first step consists in detecting the presence of a change in the time series. If the series is stationary, we assume that there is no change-point to be estimated. The so-called omnibus test scheme is intended to choose between the two following hypotheses:

Let $(t_1, t_2) \in \llbracket 1, T \rrbracket^2$, so that $t_2 > t_1$,

$$\begin{cases} H_{0,\text{omni}}^{t_1, t_2} : \boldsymbol{\theta}_{t_1} = \dots = \boldsymbol{\theta}_{t_2} = \boldsymbol{\theta}_{t_1, t_2} \\ H_{1,\text{omni}}^{t_1, t_2} : \exists (t, t') \in \{t_1, \dots, t_2\}^2, \boldsymbol{\theta}_t \neq \boldsymbol{\theta}_{t'} \end{cases} \quad (1.33)$$

An appropriate statistic of the observations must be used to decide between the two hypotheses. This problem has been tackled for various data models in chapter 3.

Estimation Strategy

The scheme (1.33) determines if there is one or more changes. In the case of a positive outcome, the location of the changes in the time series has to be estimated. To this end, successive bi-date detection schemes can be implemented:

$$\forall t \in \llbracket 2, T \rrbracket, \begin{cases} H_{0,\text{bi-date}}^t : \boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1, t} \\ H_{1,\text{bi-date}}^t : \boldsymbol{\theta}_{t-1} \neq \boldsymbol{\theta}_t \end{cases} \quad (1.34)$$

However, this scheme exploits at most the data of two successive dates which is sub-optimal. Moreover, this methodology relies on multiple hypothesis testing in order to derive a decision, for which as we saw in 1.4.5, we must consider corrections on the threshold.

An alternative scheme proposed is to consider successively the following marginal hypotheses:

Consider $(t_1, t_2) \in \llbracket 1, T \rrbracket^2$, so that $t_2 > t_1$,

$$\begin{cases} H_{0,\text{marg}}^{t_1, t_2} : \boldsymbol{\theta}_{t_1} = \dots = \boldsymbol{\theta}_{t_2-1} = \boldsymbol{\theta}_{t_1, t_2-1} \text{ and } \boldsymbol{\theta}_{t_2-1} = \boldsymbol{\theta}_{t_2} \\ H_{1,\text{marg}}^{t_1, t_2} : \boldsymbol{\theta}_{t_1} = \dots = \boldsymbol{\theta}_{t_2-1} = \boldsymbol{\theta}_{t_1, t_2-1} \text{ and } \boldsymbol{\theta}_{t_2-1} \neq \boldsymbol{\theta}_{t_2} \end{cases} \quad (1.35)$$

In this scheme, the data which are not considered as a change are used, leading to a better estimation of the parameters concerning the null hypothesis. This is expected to improve the performance of both detection and estimation given that the number of samples available is an important parameter. Moreover, since each test does not rely on the previous one, there is no need to consider threshold corrections.

The algorithm

Both detection and estimation can be done jointly using problems (1.33) and (1.35). [Conradsen et al., 2016] has proposed the Algorithm 2.

Algorithm 2 Change-Point Detection and Estimation

```

1: Initialize  $t_1 \leftarrow 1$ 
2: while  $H_{1,\text{omni}}^{t_1,T}$  do ▷ Omnibus test
3:   Initialize  $r \leftarrow 1$ 
4:   while  $H_{0,\text{marg}}^{t_1,t_1+r}$  do ▷ Successive marginal tests
5:     Update  $r \leftarrow r + 1$ 
6:   end while
7:   Store  $t_1 + r - 1$  as a change point
8:   Update  $t_1 \leftarrow t_1 + r$ 
9: end while

```

The presented algorithm allows to detect several change points by first detecting a global change in the series and then refining the detection by iterating on the number of processed dates. One key-point in the algorithm is that the statistics of decision for both omnibus and marginal schemes should have the Constant False Alarm (CFAR) property which means that their distribution is independent of the covariance matrix of the input data. Hence, the strength of this method lies in the possibility to select detection thresholds as a function of the probability of false alarm (P_{FA}) and the fact that the number of changes is not required to be known a priori.

Statistics for marginal scheme under Gaussian case

The Gaussian GLRT for the marginal scheme has also been derived in [Conradsen et al., 2016]:

$$\hat{\Lambda}_{\mathbb{C}\mathcal{N},\text{marg}}^{t_1,t_2} = \frac{\left| \frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} \hat{\Sigma}_t \right|^{(t_2-t_1)N}}{\left| \hat{\Sigma}_{t_2} \right|^N \left| \frac{1}{t_2 - t_1 - 1} \sum_{t=t_1}^{t_2-1} \hat{\Sigma}_t \right|^{(t_2-t_1-1)N}} \stackrel{H_1}{\gtrless} \lambda. \quad (1.36)$$

The associated P_{FA} -threshold relationship has also been obtained:

$$\begin{aligned}
P \left\{ 2\rho \log(\hat{\Lambda}_{\mathbb{C}\mathcal{N},\text{marg}}^{t_1,t_2}) \leq z \right\} &\approx P \left\{ \chi_{f^2}^2 \leq z \right\} + \omega_{2,t} [P \left\{ \chi_{f^2+4}^2 \leq z \right\} - P \left\{ \chi_{f^2}^2 \leq z \right\}] \\
f &= p^2, \rho_t = 1 - \frac{(2p^2 - 1)}{6pN} \left(1 + \frac{1}{t(t-1)} \right) \\
\omega_{2,t} &= -\frac{p^2}{4} \left(1 - \frac{1}{\rho_t} \right)^2 + \frac{1}{24N^2} p^2 (p^2 - 1) \left(1 + \frac{2t-1}{t^2(t-1)^2} \right) \frac{1}{\rho_t^2},
\end{aligned} \quad (1.37)$$

where $t = t_2 - t_1$.

This allows to obtain a threshold to decide for each test without requiring expensive Monte Carlo simulations.

With regards to this algorithm, the work presented in the Chapter 4 consists in replacing the Gaussian-derived statistics with those obtained using the broader compound-Gaussian model. This, in theory, allows to bring robustness to the methodology as it is the case for the change detection problem in Chapter 3.

1.6 SUMMARY

We described in this chapter a few issues related to the processing of SAR image time series for change detection purposes. We explained that different types of images can be obtained depending on the acquisition methods. In the context of this thesis, we will consider images corresponding to complex values, usually denominated as single look complex images.

Then, we explained that change detection based on statistical framework relies on the concept of diversity which can be naturally present (polarimetry) but is sometimes lacking. In this case, spectro-angular diversity is an interesting track to characterize the scatterers present in SAR images. However, depending on the methodology used, the wavelet decomposition along sub-apertures and sub-bands can lead to unwanted patterns on the resulting images. For such purposes, we will consider the development of new wavelet packets in the chapter 2.

Using this diversity, it is possible to consider change-detection through the statistical problem of covariance homogeneity testing which has shown promising results in the literature. However, these works have considered a Gaussian hypothesis, which is as we showed inaccurate in high-resolution images. For that purpose, we will consider in the chapter 3 the problem of covariance equality testing in the context of the broader elliptical family of distributions as well as deterministic compound-Gaussian model. Then, exploiting the hypothesis testing schemes, we will consider the problem of change-point detection and estimation in chapter 4.

Finally, we will present in Chapter 5, an opening of this work based on Riemannian geometry which is able to solve problems related to on-line change-point detection as well as clustering problems.

CHAPTER 2

DESIGN OF WAVELET PACKETS FOR HIGH-RESOLUTION SAR IMAGES PROCESSING

Contents:

2.1	Reminding the principle of SAR image reconstruction	56
2.2	Adapting Shannon wavelet packets to SAR geometry	57
2.2.1	Shannon M-band wavelets theory	58
2.2.2	Adaptation of Shannon wavelets to SAR geometry	61
2.3	Design of a new wavelet packets to reduce sidelobes	63
2.4	Application to real SAR images	68
2.4.1	Datasets considered	68
2.4.2	Analysis of high-resolution SAR images decomposition	68
2.4.3	Preliminary results in target detection applications	68
2.5	Conclusions	79
2.A	Appendix	80
2.A.1	Proof of proposition 2.2.1 at p. 59	80
2.A.2	Proof of proposition 2.2.2 at p. 62	80
2.A.3	Proof of proposition 2.2.3 at p. 62	81
2.A.4	Proof of proposition 2.2.4 at p. 63	81
2.A.5	Proof of proposition 2.3.2 at p. 65	81

In this chapter, we consider the design of wavelet packet formalism adapted to SAR image analysis of dispersibility and anisotropy. As described in 1.2.3, Time-Frequency analysis tools have been successfully used in retrieving those behaviors of scatterers. This chapter aims at providing a comprehensive wavelet packet formalism, which as we will see, is an interesting generalization of linear time-frequency tools. To that end, we first describe the geometry of acquisition for a SAR system and give the definition of relevant physical parameters. Then we link standard linear time-frequency tools to wavelet packets formalism. From there the wavelet packet formalism is adapted from cartesian representation to polar one better suited to SAR geometry. Finally, we propose a new wavelet packet aimed at reducing sidelobes related effects.

Notations relevant to his chapter: Given a 2-dimensional (2D) function $g \in L^1(\mathbb{C}^2) \cup L^2(\mathbb{C}^2)$, the 2D Fourier transform (resp. inverse Fourier Transform) is denoted by $\mathcal{F}g(\omega_1, \omega_2) = \int_{\mathbb{R}} g(x_1, x_2) e^{-i\omega_1 x_1} e^{-i\omega_2 x_2} dx_1 dx_2$ (resp. $\mathcal{F}^{-1}g$) and define $\tau_{[p,q]}g(x, y) = g(x - p, y - q)$.

For a function, $\langle \bullet, \bullet \rangle$ is the inner product on $L^1(\mathbb{C}^2) \cup L^2(\mathbb{C}^2)$ and $\|\bullet\|$ is the L^2 norm. $\mathbf{1}_K$ denotes the indicator function of a given set K .

2.1 REMINDING THE PRINCIPLE OF SAR IMAGE RECONSTRUCTION

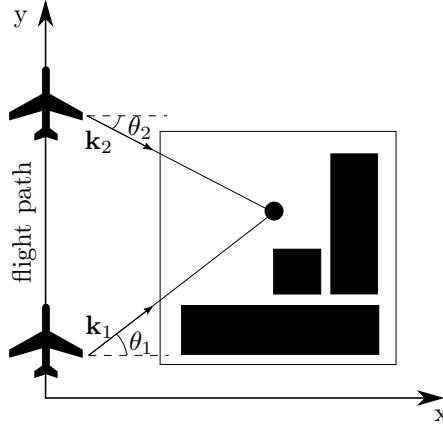


Figure 2.1: SAR acquisition geometry. A reflector is viewed at two different angles of illumination θ_1 and θ_2 in a given frequency. This information is summarized through the wave vectors \mathbf{k}_1 , \mathbf{k}_2 .

Figure 2.1 presents the geometry of acquisition for a SAR system. The moving radar transmits an electromagnetic wave represented by the wave vector $\mathbf{k} = [k_x, k_y]^T$ and recovers the backscattering signal in order to obtain a map of the reflectors of the scene. \mathbf{k} is related to the emitted frequency f by $\|\mathbf{k}\| = \mathfrak{K} = 2f/c$, c being the celerity of the light, and to the angle of illumination θ by $\theta = \arctan(k_y/k_x)$.

The emitted signal is located in a certain range of frequencies defined by: $[f_0 - B/2, f_0 + B/2]$, f_0 being the carrier and B being the bandwidth of the radar. This translates in terms of spatial frequencies \mathfrak{K} to: $[\mathfrak{K}_0 - \mathfrak{K}_B/2, \mathfrak{K}_0 + \mathfrak{K}_B/2]$ with $\mathfrak{K}_0 = 2f_0/c$, $\mathfrak{K}_B = 2B/c$. The angles θ of illumination lies in $[-\theta_B, \theta_B]$ angular domain. The spatial SAR resolutions are given respectively in azimuth by $\delta_y = c/(4f_0\theta_B)$ and in radial range by $\delta_x = c/(2B)$. We define $\mathcal{D} = [\mathfrak{K}_0 - \mathfrak{K}_B/2, \mathfrak{K}_0 + \mathfrak{K}_B/2] \times [-\theta_B, \theta_B]$, and $\mathbf{U}^{S,\angle}$ as the space of functions having spectro-angular features in \mathcal{D} .

In algorithms such as RMA [Carrara et al., 1995], the aim is to collect a backscattering reflection coefficient $\tilde{I}(\mathbf{k})$ and then perform Fourier based spectral estimation in order to build the conventional complex single look (monovariate) SAR image $I(\mathbf{r})$ for each point $\mathbf{r} = [x, y]^T$ on the ground:

$$I(\mathbf{r}) = \int_{\mathcal{D}} \tilde{I}(\mathbf{k}) \exp(2i\pi \mathbf{k}^T \mathbf{r}) d\mathbf{k}, \quad (2.1)$$

where the integration is performed on the whole spectral and angular domains.

As described in section 1.2.3, this methodology relies on the assumption that all scatters have an isotropic and non-dispersive behavior. These assumptions are no longer respected in HR SAR images which lead to artifacts on the resulting image. However, this behavior can, in return, be considered as a source of diversity to characterize the objects present in the scene.

In order to retrieve the behavior of the scatters in the space of spectral and angular features $[\mathfrak{K}, \theta]$, works such as [Tria et al., 2007] or [Ovarlez et al., 2003] have proposed to consider the anisotropic and non-dispersive behavior in terms of non-stationarities of the spectra. Using this model, wavelet analysis is a powerful tool for analyzing the behavior of the colored scatterers. For instance, an hyperimage representing the reflectivity of the scene for any sub-space $\mathcal{E} \subset \mathcal{D}$ is given by:

$$\tilde{I}_{\mathcal{E}}(\mathbf{r}) = \int_{\mathcal{E}} \tilde{I}(\mathbf{k}) \overline{\Psi_{\mathcal{E}}^S(\mathbf{k}, \mathbf{r})} d\mathbf{k}, \quad (2.2)$$

where $\Psi_{\mathcal{E}}^S(\mathbf{k}, \mathbf{r})$ is a wavelet function with spectro-angular support \mathcal{E} . When considering several subsets $\mathcal{E}_{1, \dots, M}$, a wavelet packet $\{\Psi_{\mathcal{E}_i}^S(\mathbf{k}, \mathbf{r}) / i = 1, \dots, M\}$ can be defined. As an example, an illustration of a partitioning of the spectra into 9 subsets is presented in Figure 2.1. The problematic is then to choose the shape of the wavelets and a relevant partition of \mathcal{D} in terms of \mathcal{E}_i as to decompose the image in separate frequency bands and range of angles for a given purpose.

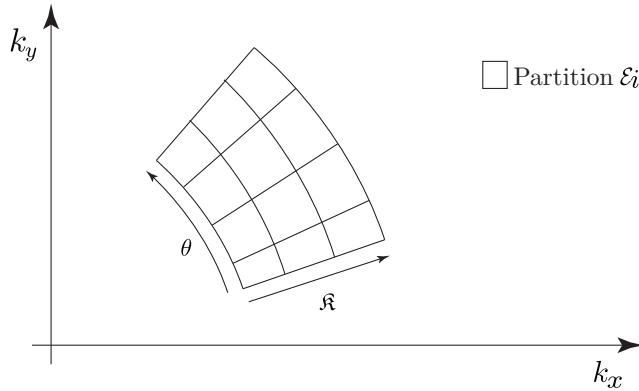


Figure 2.2: Illustration of SAR spectra partitioning

This approach was used in [Ovarlez et al., 2017] where it was adapted to work in target detection problems. However, in this work, the Short Time Fourier Transform (STFT) was considered with no analysis of the algebraic properties of the ensued decomposition. Another issue with the STFT concerns the known relative high side lobes introduced by the rectangular windows while others wavelet alternatives allow limiting this phenomenon with a cost of deforming the input signal. To tackle these issues in a comprehensive manner, we will consider, in the next sections, the problem in terms of wavelet packet decomposition. Indeed, as we will see in the next section, the formalism of wavelet packets is an interesting generalization of the STFT which allows considering the time-frequency analysis in terms of decomposition in algebraic space to analyze problems of signal deformation analysis, reconstruction properties and so on.

2.2 ADAPTING SHANNON WAVELET PACKETS TO SAR GEOMETRY

In this section, new wavelet packets adapted to SAR geometry are developed. The particularity of these new wavelet packets with regards to existing wavelet packets literature is the choice of polar representation which better describes the data with regards to the spectro-angular diversity of interest. To this end, we use classic Shannon M-band filters

that we adapt to take into account the spectral support of SAR images and then we correct the edge effects of those wavelets.

The adaptation of wavelet packets in this geometry can be done using many classic wavelets packets (Gabor, Debaucheries, *etc.*). However, Shannon M-band filters have been chosen as a basis of our design for the following reasons:

- They are separable with regards to the two dimensions (\mathfrak{K}, θ) of the decomposition, which makes them ideal when we want to choose the number of sub-bands and sub-looks (as in looking angle) separately.
- Since we expect to exploit the decomposition in classic statistical detection schemes, there is a need for each coefficient to deliver different information than the others. Otherwise, correlations between sub-bands/sub-looks would be introduced due to the shape of the wavelets and may deteriorate detection performance. This leads to a choice of an orthogonal wavelet packet.
- To better describe the behavior of a possible target as a function of the frequencies and looking angle, we consider wavelets corresponding to a connected subset of the frequency/angular domain.

2.2.1 Shannon M-band wavelets theory

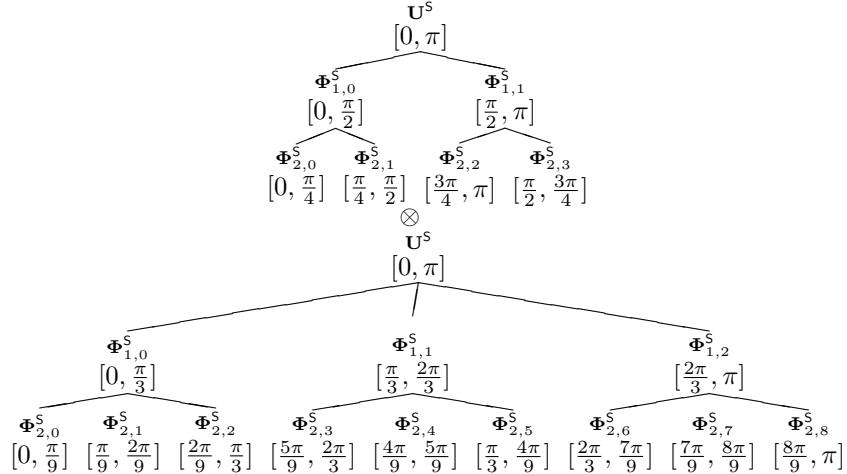


Figure 2.3: Shannon 2×3 multi-band decomposition tree associated with $j = 2$. The positive part of $\Delta_{j,\bullet}^{1,+}$ of $\Delta_{j,\bullet}^1$ are given in the upper tree whereas $\Delta_{j,\bullet}^{2,+}$, positive part of $\Delta_{j,\bullet}^2$ are given at the bottom tree. The frequency tiles associated with the decomposition are the intervals $\Delta_{j,\bullet}^1 \times \Delta_{j,\bullet}^2$ for every fixed j : the whole tree involves all combination of nodes given at a fixed resolution level j .

Let M_1 and M_2 be natural numbers that are both greater than or equal to 2. The Shannon 2D $M_1 \times M_2$ multi-band wavelet filters used in this work follows from a separable 2D extension of 1D filters presented in [Hess-Nielsen, 1994], [Nielsen, 2002]. These filters give a multi-resolution framework for decomposing any image.

We define \mathbf{U}^S as the 2D Paley-Wiener (PW) space composed by elements of $L^2(\mathbb{R}^2)$ whose Fourier transform is supported within $[-\pi, \pi]^2$. Any element of this space satisfies Shannon's sampling theorem. Therefore, when the $M_1 \times M_2$ multi-band decomposition concerns the PW space \mathbf{U}^S , the input data for the decomposition of any element g of this

functional space are the samples $\{g[k, \ell]\}_{k, \ell \in \mathbb{Z}}$ of g (corresponding to the pixels of the image to decompose).

The 2D Shannon wavelet packet function at resolution level j and 2D shift parameters (n_1, n_2) , with $n_\varepsilon \in \{0, \dots, M_\varepsilon^j - 1\}$ for $\varepsilon \in \{1, 2\}$, is given by its Fourier transform as:

$$\Psi_{j, [n_1, n_2]}^S = M_1^{j/2} M_2^{j/2}(k_x, k_y) \mathbf{1}_{\Delta_{j, G^{[1]}(n_1)}^0 \times \Delta_{j, G^{[2]}(n_2)}^1}(k_x, k_y), \quad (2.3)$$

where

$$\Delta_{j, k}^\varepsilon = \left[-\frac{(k+1)\pi}{M_\varepsilon^j}, -\frac{k\pi}{M_\varepsilon^j} \right] \cup \left[\frac{k\pi}{M_\varepsilon^j}, \frac{(k+1)\pi}{M_\varepsilon^j} \right], \quad (2.4)$$

and $(G^{[\varepsilon]})_{\varepsilon \in \{1, 2\}}$ are the permutation maps defined respectively, for $\varepsilon \in \{1, 2\}$, by $G^{[\varepsilon]}(0) = 0$ and by recursively setting, for $k = 0, 1, \dots, M_\varepsilon - 1$ and $\ell = 0, 1, 2, \dots$

$$\begin{aligned} G^{[\varepsilon]}(M\ell + k) \\ = \begin{cases} MG^{[\varepsilon]}(\ell) + k & \text{if } G^{[\varepsilon]}(\ell) \text{ is even,} \\ MG^{[\varepsilon]}(\ell) - k + M - 1 & \text{if } G^{[\varepsilon]}(\ell) \text{ is odd.} \end{cases} \end{aligned} \quad (2.5)$$

Define $\Phi_{j, [n_1, n_2]}^S = \mathcal{F}^{-1}\Psi_{j, [n_1, n_2]}^S$. A 2D wavelet packet subspace $\Phi_{j, [n_1, n_2]}^S$ is generated as the closure of the space spanned by the following translated versions of $\Phi_{j, [n_1, n_2]}^S$:

$$\Phi_{j, [m, n]}^S = \text{Clos} \left\langle \tau_{[M_1^j p, M_2^j q]} \Phi_{j, [m, n]}^S : p \in \mathbb{Z}, q \in \mathbb{Z} \right\rangle. \quad (2.6)$$

These subspaces are such that for any fixed j ,

$$\mathbf{U}^S = \bigoplus_{\substack{m=0, 1, \dots, M_1^j - 1 \\ n=0, 1, \dots, M_2^j - 1}} \Phi_{j, [m, n]}^S$$

where \oplus denotes the direct sum of functional subspaces.

As an illustration, the Shannon 2×3 multi-band wavelet packet tree is given by Figure 2.3 as a tree product resulting from a 2-band and a 3-band 1D trees, where the tree-product involves all combination of nodes given at a fixed resolution level j . In this figure, the positive part $\Delta_{j, k}^{\varepsilon,+}$ of $\Delta_{j, k}^\varepsilon$ is given for each resolution level j under consideration.

The Shannon $M_1 \times M_2$ multi-band coefficients of the projection of g on a 2D wavelet packet subspace $\Phi_{j, [n_1, n_2]}^S$ defines the wavelet coefficients:

$$\mathbf{C}_{j, [n_1, n_2]}^S[p, q] = \iint_{\mathbb{R}^2} g(z, t) \tau_{[M_1^j p, M_2^j q]} \Phi_{j, [n_1, n_2]}^S(z, t) dz dt. \quad (2.7)$$

Proposition 2.2.1. *The coefficients of the projection of g on a wavelet packet subspace $\Phi_{j, [n_1, n_2]}^S$ is a discrete sequence $\mathbf{C}_{j, [n_1, n_2]} = (\mathbf{C}_{j, [n_1, n_2]}^S[p, q])_{p, q \in \mathbb{Z}}$ where*

$$\mathbf{C}_{j, [m, n]}^S[p, q] = M_1^{-j/2} M_2^{-j/2} \mathcal{F}^{-1} U_{j, [m, n]}(M^j p, M^j q), \quad (2.8)$$

with

$$U_{j, [m, n]} = [\mathcal{F}g] \times \mathbf{1}_{\Delta_{j, G(m)}^0 \times \Delta_{j, G(n)}^1}. \quad (2.9)$$

Proof. See in Appendix 2.A.1. \square

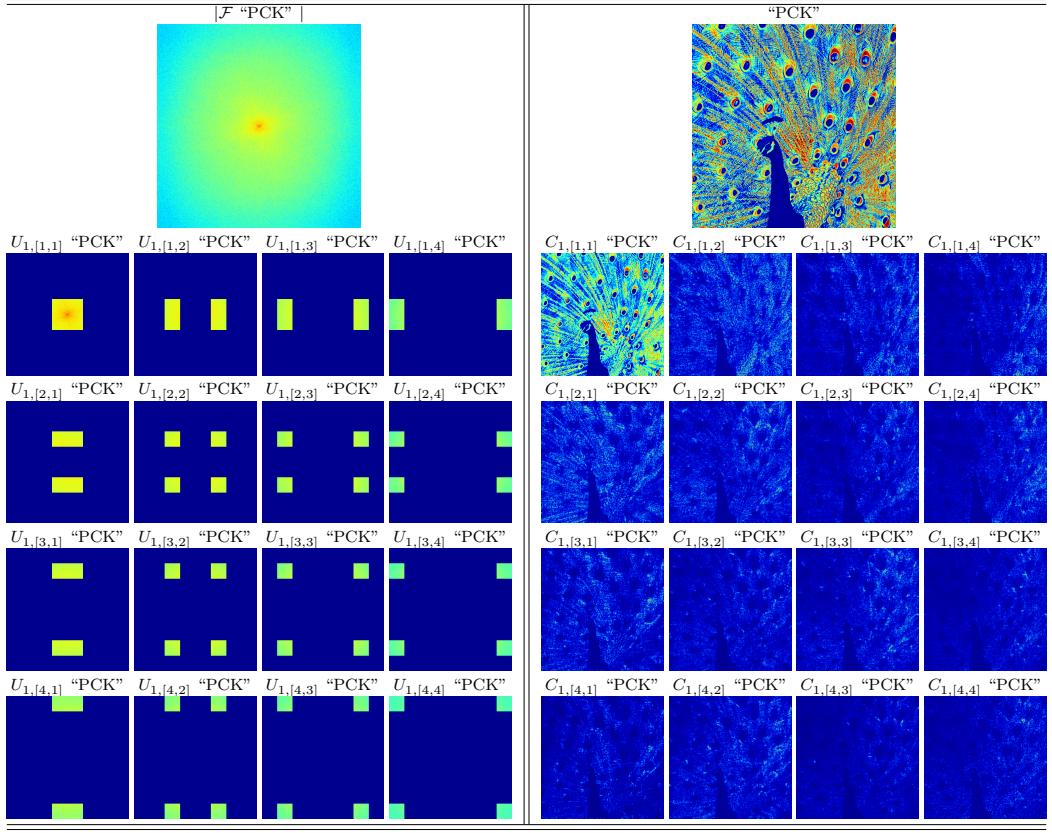


Figure 2.4: Peacock image - Its Cartesian Fourier magnitudes - Shannon $M = 4$ band frequency sub-selections and the corresponding 4-band wavelet subimages.

In practice, g is a discrete image to be decomposed. This proposition shows how wavelet coefficients can be easily obtained in practice through a simple Fast Fourier Transform (FFT). $[p, q]$ are the pixels of the wavelet coefficient for shift parameters $[n_1, n_2]$ at a fixed resolution j . Note that, a decomposition at a given resolution j assumes that the wavelets coefficients correspond to a decimated version of the image (by a factor of M_1^j and M_2^j). This methodology enables to compute efficiently the coefficients with a low-complexity. Indeed, since only a Hadamard product and an FFT is necessary, the complexity is linear with regards to the number of coefficients desired.

The wavelets presented here are designed for images respecting Shannon sampling theorem. The decomposition is done on functions whose frequencies are contained in the space $[-\pi, \pi]^2$. We adapt hereafter Shannon wavelets from the Cartesian space $[-\pi, \pi]^2$ to the polar space \mathcal{D} .

The representation space has indeed a great impact on the ensuing decomposition. To better understand how the representation can impact the decomposition results, an example of M-band decomposition on a classic image of a peacock is presented in Figure 2.4 using a Cartesian representation and polar representation in Figure 2.5 of this response. The comparison between the two Figures highlights that information is represented differently when considering Cartesian and polar features: in the Cartesian 4-band splitting (see Figure 2.4), the main highlighted information is the approximation of the peacock with different

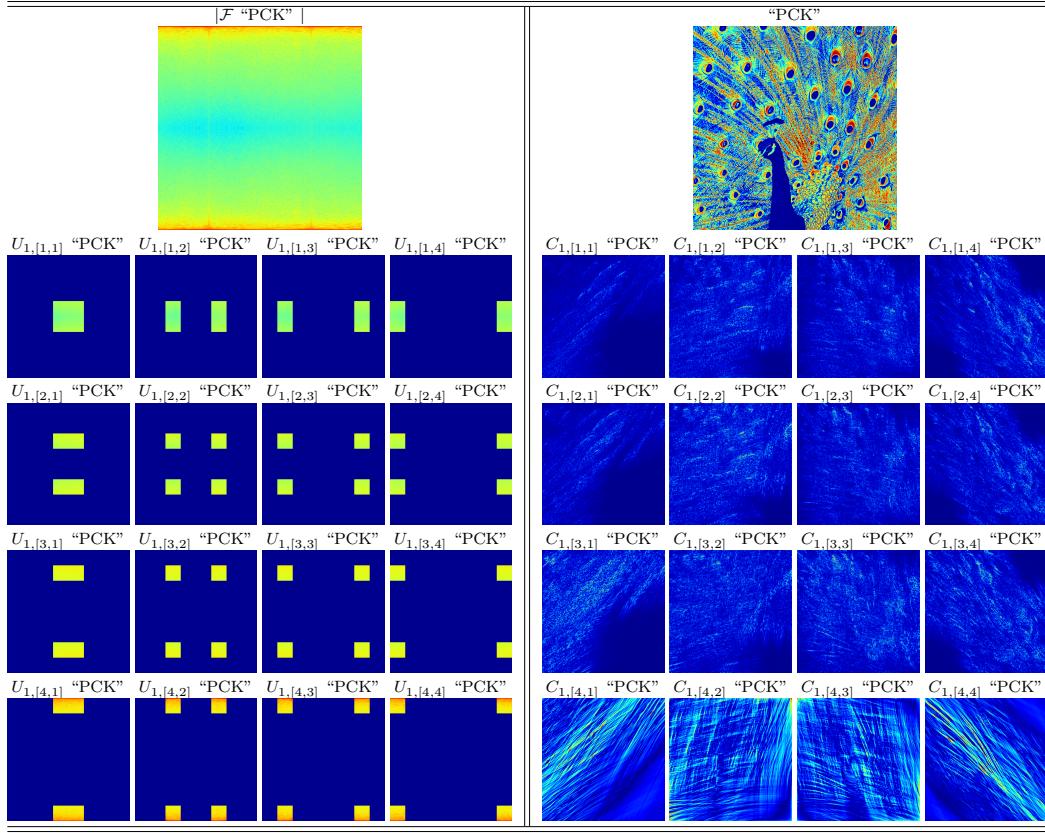


Figure 2.5: Peacock image - Its Polar Fourier magnitudes - Shannon $M = 4$ band frequency sub-selections and the corresponding 4-band wavelet subimages.

edge details. On the contrary, circular peacock's tail range is the main information associated with the polar wavelet domain implementation (see Figure 2.5). When decomposing a SAR image, this difference is primordial since when using a Cartesian representation will not result in a decomposition suited for analyzing the spectro-angular behavior.

2.2.2 Adaptation of Shannon wavelets to SAR geometry

Define

$$\Psi_{j,[m,n]}^{\mathbf{S},\angle}(\mathfrak{K}, \theta) = R^{j/2} L^{j/2} \mathbb{1}_{\Delta_{j,\mathfrak{K}_m} \times \Delta_{j,\theta_n}}(\mathfrak{K}, \theta), \quad (2.10)$$

where

$$\Delta_{j,\mathfrak{K}_m} = \mathfrak{K}_0 - \mathfrak{K}_B + \left[\frac{m\mathfrak{K}_B}{R^j}, \frac{(m+1)\mathfrak{K}_B}{R^j} \right], \quad (2.11)$$

$$\Delta_{j,\theta_n} = \left[\frac{n\theta_B}{L^j}, \frac{(n+1)\theta_B}{L^j} \right]. \quad (2.12)$$

In contrast to eq. (2.3), these functions are defined in polar coordinates in order to span the spectrum of the SAR image as illustrated in Figure 2.1. From this, we define the wavelet functions $\Phi_{j,[m,n]}^{\mathbf{S},\angle}(x, y) = \mathcal{F}^{-1}\Psi_{j,[m,n]}^{\mathbf{S},\angle}(\mathfrak{K}, \theta)$. Here, the variables x and y correspond

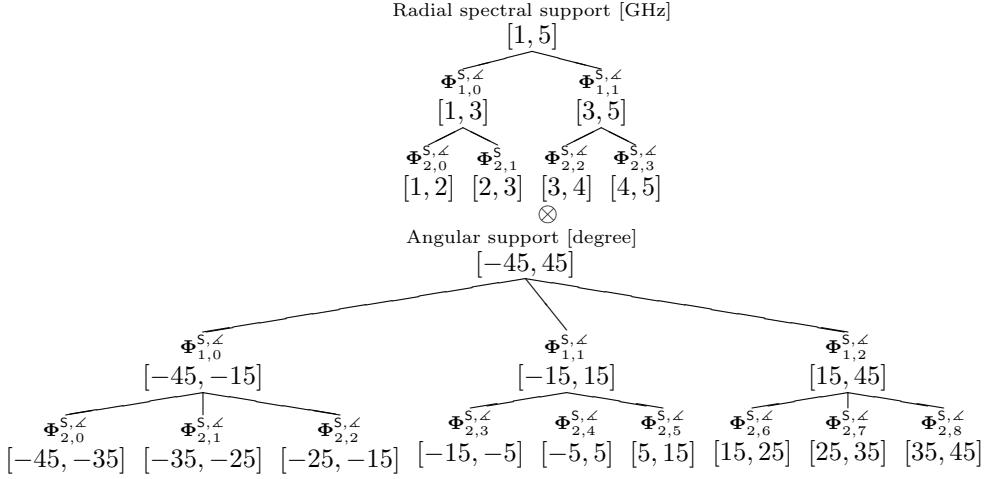


Figure 2.6: Cartesian product \otimes of spectral and angular multi-resolution trees associated with radial ‘2 bands - 2 resolutions’ and angular ‘3 looks - 2 resolutions’ splitting $\Psi_{j,[m,n]}^{S,\angle}(\mathfrak{K}, \theta)$ defined by Eq. (2.16) when considering $[f_0 - B/2, f_0 + B/2] \times [-\theta_B, \theta_B] = [1\text{GHz}, 5\text{GHz}] \times [-45 \text{deg}, 45 \text{deg}]$. The intervals represented are given for illustration.

to the range and cross-range positions as in Figure 2.1. Note that this definition requires computing the Fourier transform on spectral and angular variables. Among the different possible solutions of this problem, we use interpolation from the Fractional Fast Fourier Transform (FFFT or 3FT) in order to fill the polar grid $[\mathfrak{K}_0 - \mathfrak{K}_B/2, \mathfrak{K}_0 + \mathfrak{K}_B/2] \times [-\theta_B, \theta_B]$ from the Cartesian one corresponding to variables k_x, k_y . Among the 3FT implementations, we recommend using that of [Averbuch et al., 2006].

The wavelets thus defined constitute a wavelet packet as per the following proposition:

Proposition 2.2.2 (Vanishing moments). *For any non-negative integers j, m, n, p, q , we have*

$$\iint_{\mathbb{R}^2} x^p y^q \Phi_{j,[m,n]}^{S,\angle}(x, y) dx dy = 0.$$

Proof. See Appendix 2.A.2. □

Functions $\Phi_{j,[m,n]}^{S,\angle}$ defined above have thus an infinite number of vanishing moments. Since they are well localized in space/frequency/angle, they are wavelet functions.

Define the wavelet subspaces $\Phi_{j,[m,n]}^{S,\angle}$ similarly to eq. (2.6). Then we have:

Proposition 2.2.3 (Orthogonality of wavelet packet subspaces). *For any given j and any $(m, n) \neq (m', n')$ we have:*

$$\Phi_{j,[m,n]}^{S,\angle} \perp \Phi_{j,[m',n']}^{S,\angle},$$

where \perp denotes orthogonality symbol.

Proof. See Appendix 2.A.3. □

Proposition 2.2.4 (Completion of wavelet packet subspaces). *For any given j , and any (m, n) , we have*

$$\bigcup_{\substack{m=0,1,\dots,R^j-1 \\ n=0,1,\dots,L^j-1}} \Phi_{j,[m,n]}^{S,\angle} = \mathbf{U}^{S,,\angle}.$$

Proof. See Appendix 2.A.4. □

Propositions 2.2.2, 2.2.3 and 2.2.4 highlight that wavelet subspaces $\Phi_{j,[m,n]}^{S,\angle} : j \geq 1, m = 0, 1, \dots, R^j - 1, n = 0, 1, \dots, L^j - 1$ can thus be used to define several multi-resolution frameworks (specific sub-selection of j, m, n) for analyzing SAR data.

Analyzing SAR data is done by computing the wavelet coefficients as previously stated in (2.2). We have:

Proposition 2.2.5 (Wavelet coefficients). *In practice, at a resolution level j , we obtain the wavelet coefficients by computing the following:*

$$\mathbf{C}_{j,[m,n]}^{S,\angle}[p, q] = R^{-j/2} L^{-j/2} \mathcal{F}^{-1} V_{j,[m,n]}(R^j p, L^j q), \quad (2.13)$$

where $V_{j,[m,n]}$ follows from the back-projection of the SAR image with respect to spectral and illumination features:

$$V_{j,[m,n]}(\mathfrak{K}, \theta) = \tilde{I}(\mathfrak{K}, \theta) \mathbf{1}_{\Delta_{j,\mathfrak{K}_m} \times \Delta_{j,\theta_n}}(\mathfrak{K}, \theta). \quad (2.14)$$

Proof. Analog to the proof of proposition 2.2.5. □

Again, the complexity of this methodology is linear with regards to the number of coefficients ($R \times L$) since coefficients are obtained from a Hadamard product and a 3FT. In practice, any SAR data can efficiently be analyzed using this methodology.

An example of a multi-resolution analysis is given in Figure 2.6 for spectral features and angular illumination in $[f_0 - B/2, f_0 + B/2] \times [-\theta_B, \theta_B] = [1, 5]\text{GHz} \times [-45, 45]\text{deg}$.

2.3 DESIGN OF A NEW WAVELET PACKETS TO REDUCE SIDELOBES

When considering Shannon wavelets, the decomposition is subject to hard transitions in the sense that each filter is an ideal band-pass filter. When considering the wavelet coefficients, this results in a convolution with a cardinal sine function which has high side lobes (see fig. 2.7 for an illustration). This dispersion of energy is problematic in detection schemes when secondary data, corresponding to the surrounding pixels, are needed.

To limit the side lobes on the wavelet coefficients which are due to the sharp edge of the Shannon wavelets, we look for alternatives that are subject to smooth transitions. We derive hereafter, a new family of parametrized R -band / L -look wavelet functions including the Shannon wavelets as a limit case.

We propose the following criteria for the design of the new family of wavelets:

- Well located in frequencies and angles (wavelet function).
- Similar behaviour to Shannon wavelets to preserve the framework presented in 2.2.2.
- Smooth transition with a parameter controlling the decay (for adaptability purposes).

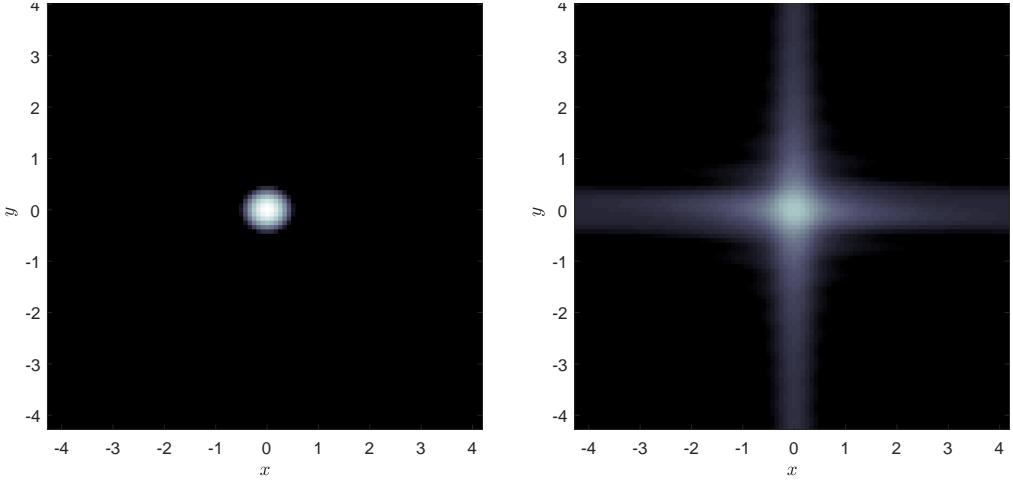


Figure 2.7: Example of bright point decomposition. Left: Image I . Right: $\mathbf{C}_{1,[1,1]}^{\mathbf{S},\angle}$ with $R = L = 2$.

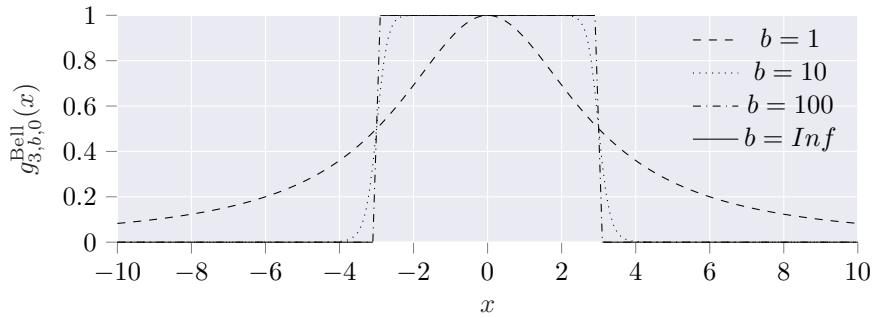


Figure 2.8: Bell-shaped function with width $a = 3$, center $c = 0$ and different slope parameters $b \in \{1, 3, 10, 50\}$.

Many functions respect the two first criteria. However, Bell-shaped membership functions appear to be a good choice as they allow to control both center, extent and slope (and thus smoothness). They are a family of one-dimensional functions defined by:

$$g_{a,b,c}^{\text{Bell}}(x) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}}. \quad (2.15)$$

where the parameter a stands for the width of the function, where the parameter b controls the slope and the parameter c is a location parameter.

Figure 2.8 gives an example of Bell function with different slopes and shows that these functions are good candidates for our problem. Using them as a basis, we define:

$$\Psi_{j,[m,n]}^{[d_1,d_2],\angle}(\mathfrak{K}, \theta) = R^{\frac{j}{2}} L^{\frac{j}{2}} H_{j,[m,n]}^{[d_1,d_2],\angle}(\mathfrak{K}, \theta) \mathbb{1}_{\mathcal{D}}, \quad (2.16)$$

where $H_{j,[m,n]}^{[d_1,d_2],\angle}$ is defined as a product of two Bell functions:

$$H_{j,[m,n]}^{[d_1,d_2],\angle}(\mathfrak{R}, \theta) = H_{j,m}^{d_1,\angle}(\mathfrak{R}) H_{j,n}^{d_2,\angle}(\theta) \quad (2.17)$$

with:

$$\begin{aligned} H_{j,m}^{d_1,\angle}(\mathfrak{R}) &= g_{\frac{\mathfrak{R}_B}{2R^j}, d_1, \mathfrak{R}_0 - \frac{\mathfrak{R}_B}{2} + \frac{(2m+1)\mathfrak{R}_B}{2R^j}}^{\text{Bell}}(\mathfrak{R}), \\ H_{j,n}^{d_2,\angle}(\theta) &= g_{\frac{\theta_B}{L^j}, d_2, -\theta_B + \frac{(2n+1)\theta_B}{2L^j}}^{\text{Bell}}(\theta). \end{aligned}$$

The definition is similar to that of eq. (2.10): the center and width of Bell functions have been adapted to span the SAR geometry domain \mathcal{D} through R^j translations along \mathfrak{R} and L^j translations along θ . The slope parameters d_1 and d_2 are let open as a parametrization of the wavelet family.

Define $\Phi_{j,[m,n]}^{[d_1,d_2],\angle} = \mathcal{F}^{-1}\Psi_{j,[m,n]}^{[d_1,d_2],\angle}(\mathfrak{R}, \theta)$, the wavelet function. We have the following properties:

Proposition 2.3.1 (Vanishing moments). *For any non-negative integers j, m, n, p, q , we have*

$$\iint_{\mathbb{R}^2} x^p y^q \Phi_{j,[m,n]}^{[d_1,d_2],\angle}(x, y) dx dy = 0.$$

Proof. Similar to 2.2.2. The null derivative in $(0, 0)$ is assured by the indicator $\mathbf{1}_{\mathcal{D}}$. \square

Proposition 2.3.1 indicates that the functions presented by eq. (2.16) define wavelets.

Proposition 2.3.2 (Convergence to Shannon wavelets). *The R-band L-look wavelet transform obtained by using Eq. (2.16) is associated with the Shannon wavelet transform when $d_1, d_2 \rightarrow +\infty$:*

$$\lim_{d_1 \rightarrow +\infty} \lim_{d_2 \rightarrow +\infty} \Psi_{j,[m,n]}^{[d_1,d_2],\angle} \stackrel{a.e.}{=} \Psi_{j,[m,n]}^{S,\angle}. \quad (2.18)$$

where equality holds true almost everywhere (a.e.).

Proof. See Appendix 2.A.5. \square

Proposition 2.3.2 highlights that the Bell-shaped wavelets have similar behaviour than Shannon wavelets for high value of d_1 and d_2 and can thus be used for analyzing SAR images. For convenience purposes, we use alternatively the notation $\Psi_{j,[m,n]}^{[\infty,\infty],\angle} = \Psi_{j,[m,n]}^{S,\angle}$.

Figure 2.9 provides the shapes associated with wavelet $\Phi_{1,[\bullet,\bullet]}^{[d_1,d_2],\angle}$ and their corresponding Fourier transforms, when $d_1 = d_2 \in \{0.1, 0.5, 1, 10\}$. As it can be seen in this figure, functions $\Phi_{1,[\bullet,\bullet]}^{[d_1,d_2],\angle}$ are highly selective in frequency and orientation when d_1, d_2 are small. When d_1, d_2 are large, these functions behave as $\Phi_{1,[\bullet,\bullet]}^{S,\angle}$.

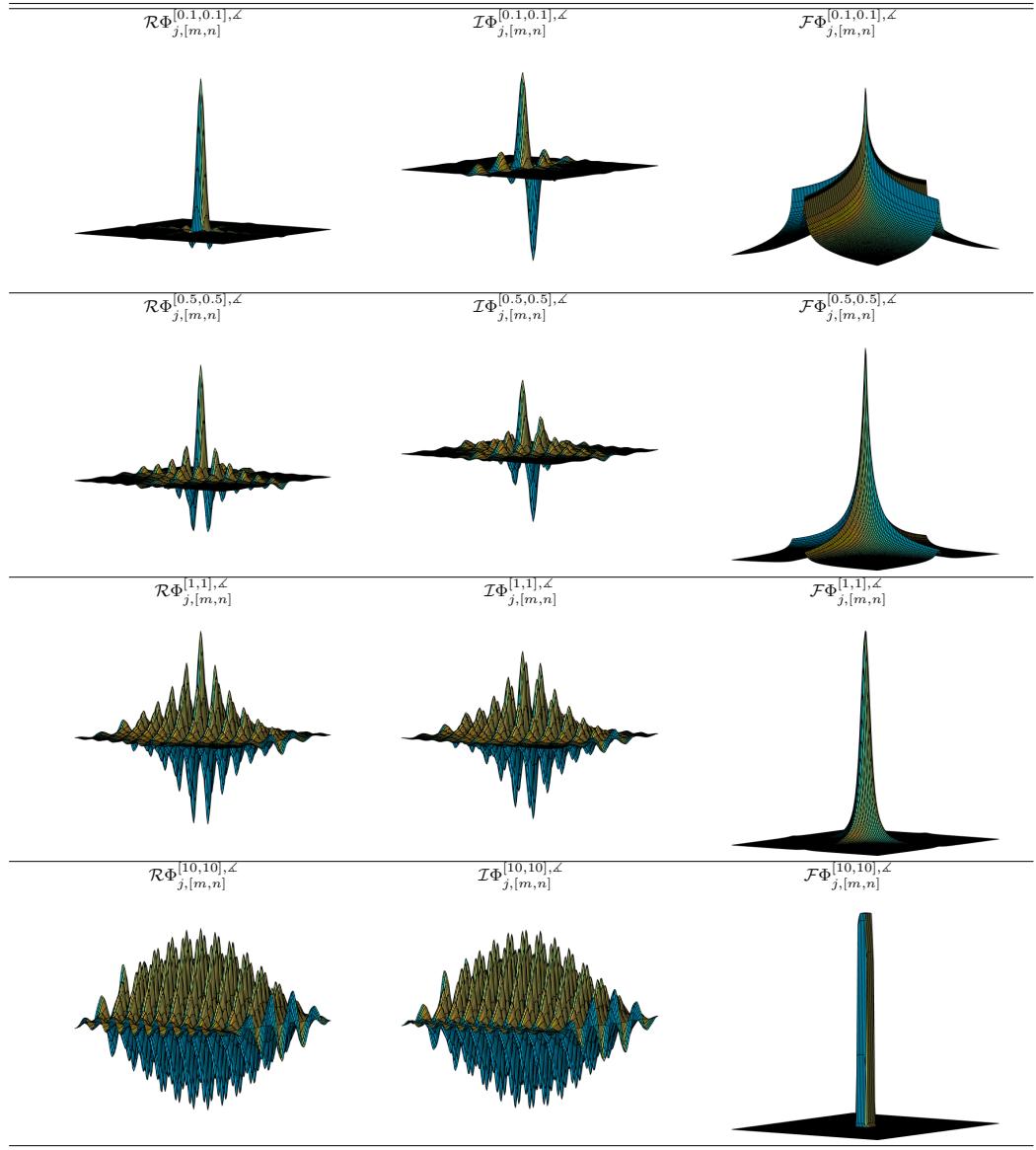


Figure 2.9: Real Part (\mathcal{R}), Imaginary Part (\mathcal{I}) and Fourier transform (\mathcal{F}) shapes of wavelet $\Phi_{1,[\bullet,\bullet]}^{[d_1,d_2],\angle}$ functions, for $d_1 = d_2 \in \{0.1, 0.5, 1, 10\}$.

Selection of slope parameters

A problem arises in the choice of these slope parameters. One can intuit that given their values, the properties of orthogonality and completion of wavelet packet subspaces are not assured. Unfortunately, given the expression of the wavelets, finding an interval of values using orthogonality or completion properties is not possible to our knowledge. As such, we propose to consider the wavelet packet in terms of frames (see [Christensen, 2002] for details) which relaxes the conditions of orthogonality and completion. As suggested in [Cohen et al., 1992], a wavelet packet has good reconstruction property if the energy of the

signal is preserved when doing the decomposition and reconstruction. In practice, this can be ensured if the following condition is respected [Kovacevic et al., 2002]:

$$Q(\mathfrak{K}, \theta) = \sum_{m,n} |H_{j,[m,n]}^{[d_1,d_2],\angle}(\mathfrak{K}, \theta)|^2 \approx 1, \forall \mathfrak{K}, \forall \theta. \quad (2.19)$$

This criterion can be used to grasp qualitatively how the decomposition will treat the frequencies present in the image. If $Q > 1$, the energy increases which means that the packet is redundant. When $Q < 1$, there is a loss of energy and thus information. We propose to use this criterion to select the values of d_1, d_2 which preserves energy the most.

Since the expression in eq. (2.17) is separable in \mathfrak{K} and θ , we can treat both separately and solve the problems: $\forall (\mathfrak{K}, \theta) \in \mathcal{D}$,

- find d_1 subject to $Q_{\mathfrak{K}}(\mathfrak{K}) = \sum_{m,n} |H_{j,[m,n]}^{d_1,\angle}(\mathfrak{K})|^2 \approx 1$ and
- find d_2 subject to $Q_{\theta}(\theta) = \sum_{m,n} |H_{j,[m,n]}^{d_2,\angle}(\theta)|^2 \approx 1$.

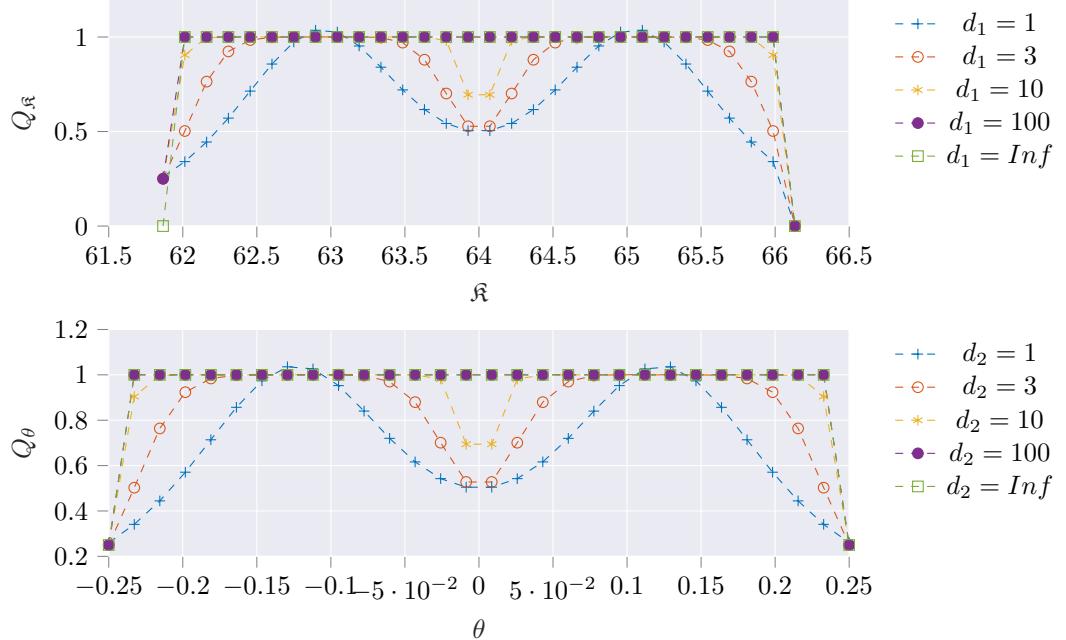


Figure 2.10: Frequencies selectivity of wavelet packet for $f_0=9.6\text{GHz}$, $B=640\text{MHz}$, $\theta_B=0.25$ rad, $j=1$, $R=L=2$.

Figure 2.10 gives the values of Q for several values of d_1 and d_2 for a given set of $(j, R, L, f_0, B, \theta_B)$. We notice that for small values of $d_{\epsilon \in \{1,2\}}$, there is a loss of energy at the transitions between the filters. For values of $d_{\epsilon \in \{1,2\}} \in [10, \infty]$, this loss is acceptable. Indeed, since the wavelet packet is developed for target detection schemes, there is a need to know the spectro-angular behavior in a vector of a fixed size. This means that if the energy of most of each band is preserved in the coefficients, this will not impact much the detection scheme.

Finally we can compute the wavelets coefficients simply by taking expression at eq. (2.13) and using $V_{j,[m,n]}(\mathfrak{K}, \theta) = \tilde{I}(\mathfrak{K}, \theta) H_{j,[m,n]}^{[d_1,d_2],\angle}$.

2.4 APPLICATION TO REAL SAR IMAGES

2.4.1 Datasets considered

Two dataset have been used to test the wavelet decompositions and their impact on target detection schemes:

- SANDIA Dataset, available at http://www.sandia.gov/radar/complex_data/. The image referenced as *MiniSAR20050519p0010image002* is selected.
- SDMS Dataset [Scarborough et al., 2010], available at https://www.sdms.afrl.af.mil/index.php?collection=ccd_challenge. The image referenced as *FP0120* is selected.

Table 2.1 summarizes the information on both datasets.

Table 2.1: Description of Dataset

Dataset	Band	Frequency	Resolution	Scene description
SANDIA	Ku	16.8 GHz	0.10 m	Stationary aircraft, trees
SDMS	X	9.6GHz	0.20 m	Foliage, buildings, vehicles

2.4.2 Analysis of high-resolution SAR images decomposition

Figure 2.11 shows a 2-Band/2-Look decomposition of a portion of SDMS Image. First, the spectro-angular behavior of the data can be analyzed: given the sub-image considered, different patterns emerge. Indeed, for example the object in the bottom-right corner ($0 > x > 100$ and $-50 < y < 0$), is not present in the coefficients $\mathbf{C}_{1,[1,2]}$ and $\mathbf{C}_{1,[2,2]}$. Next, the wavelet decomposition is compared with two parameters d_1 and d_2 . When comparing both Shannon and Bell decomposition in Figure 2.11, we observe that for Shannon wavelets, linear patterns (sidelobes for the strong scatterers present in the scene). When considering $d_1 = d_2 = 3$, the undesired linear patterns are less prominent. This result was expected as Bell-shaped wavelets make a more concise decomposition in the spatial domain.

Figure 2.12 shows a 3-Band 3-Look decomposition of a portion of SANDIA Image. The decomposition shows here as well that the scatterers do not behave as isotropic and non-dispersive since the coefficients share dissimilarities. The choice of Bell wavelets has also reduced sidelobes of the strong scatterers considerably compared to the Shannon wavelets.

2.4.3 Preliminary results in target detection applications

In this subsection, we consider a target detection problem, as an illustration, to show that the wavelet decomposition presented here fits well with statistical adaptive detection schemes. Hence, we propose to use the wavelet decomposition of the previous section in order to detect a target in a noisy SAR image. First, we give a statistical model for noise disturbances. Then we present the target detection problem.

Data Model

In the following, each pixel location (p, q) of the SAR image will be represented, *at a resolution level j* , by a set of R -radius and L -look wavelets features encapsulated in the random

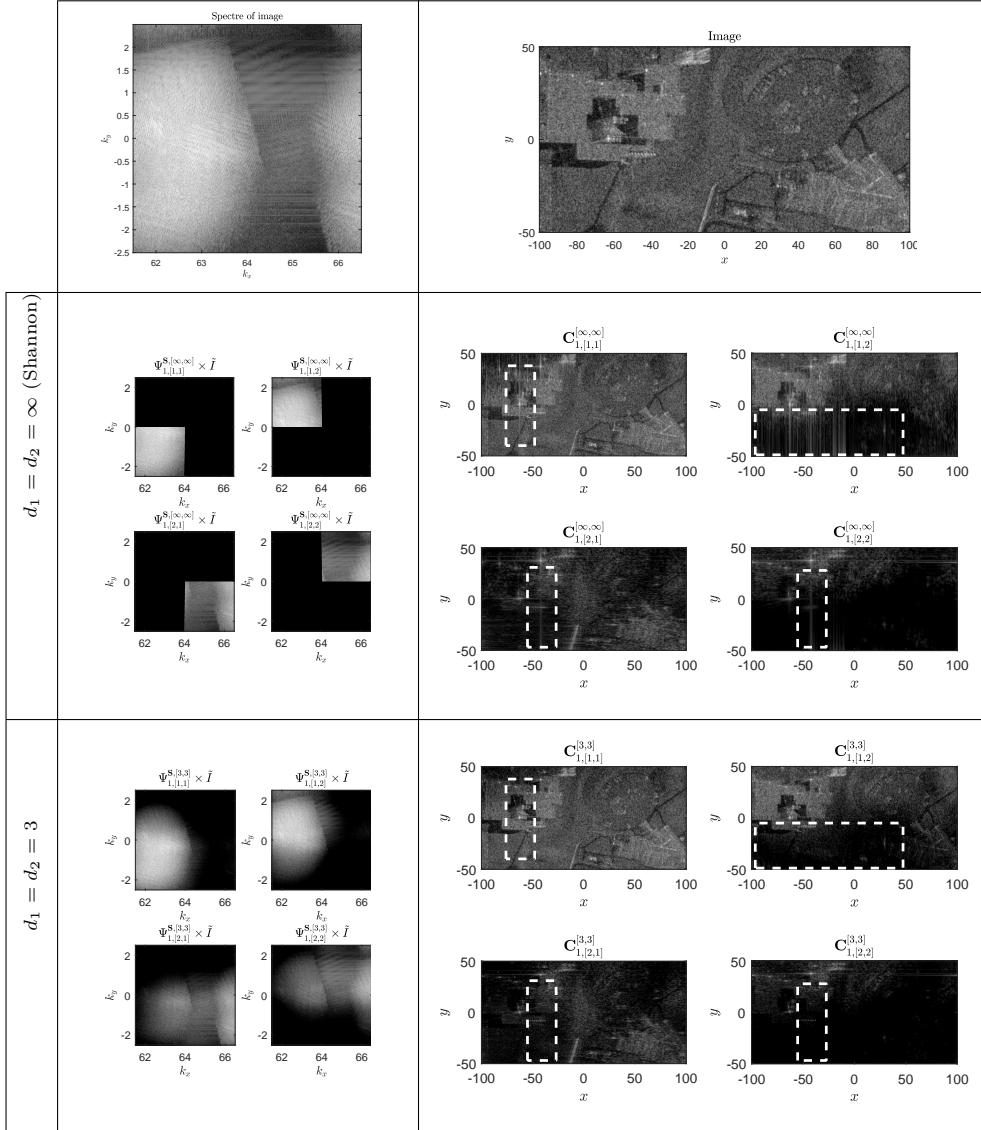


Figure 2.11: Coefficients for Shannon and Bell-Shaped wavelets on SDMS Image with $R = 2$, $L = 2$. The improvement in terms of linear patterns (side lobes of bright points) for $d_1 = d_2 = 3$ are highlighted using dashed boxes.

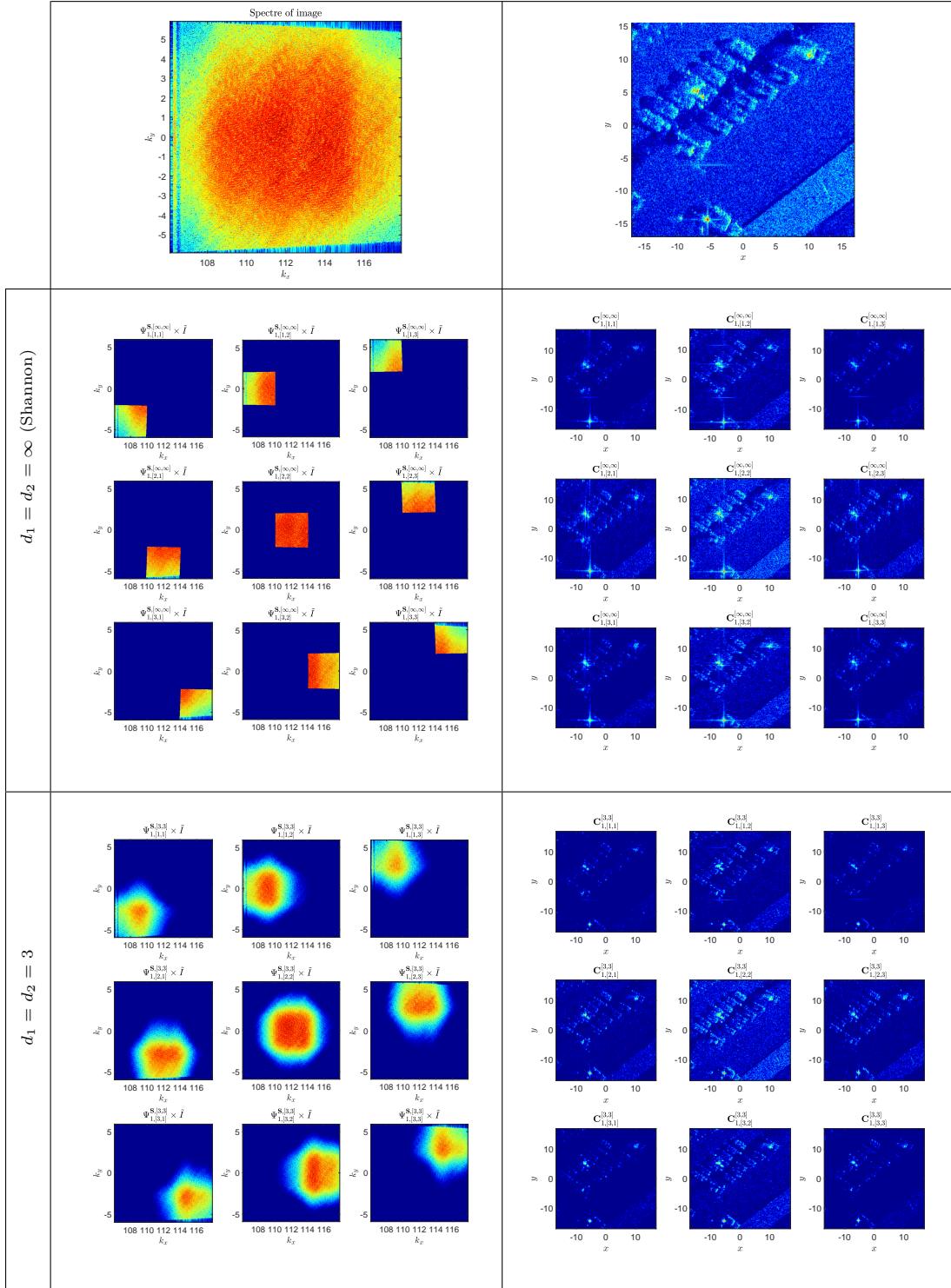


Figure 2.12: Coefficients for Shannon and Bell-Shaped wavelets on SANDIA image with $R = 3$, $L = 3$.

complex vector

$$\mathbf{c}_j[p, q] \triangleq \left\{ \mathbf{C}_{j,[m,n]}^{[d_1, d_2], \angle} [p, q] \right\}_{\substack{m=0, 1, \dots, R^j - 1 \\ n=0, 1, \dots, L^j - 1}} \in \mathbb{C}^N,$$

where $N = R^j \times L^j$.

In standard applications, the vector $\mathbf{c}_j[p, q]$ is modeled as a multivariate Gaussian vector: $\mathbf{c}_j[p, q]$ follows a Gaussian distribution $\mathcal{CN}(\mathbf{0}, \Sigma)$ where Σ is the unknown covariance matrix of the data. This model is accurate for SAR images where each pixel is often assumed to be the sum of the contributions of all the scatterers inside its range.

However, as described in section 1.3, when considering HR SAR images, the number of scatterers present in any pixel of the image is small, meaning that the Central Limit Theorem may no longer be applicable. Moreover, there are many non-stationarities inherent to this kind of images where the backscattered power can vary greatly spatially inside the analysis windows. Thus, the Gaussian hypothesis may no longer be applicable. To generalize the Gaussian statistic, we assume that $\mathbf{c}_j[p, q]$ follows a CES distribution $\mathcal{CE}(\mathbf{0}, g, \Sigma)$ as defined in 1.3.4 where the scatter matrix Σ is unknown and where g stands for any characteristic function generator.

In both models, the matrix Σ characterizes the angular and the spectral behaviour of each scatterer. To estimate this matrix, the following K secondary vectors surrounding the pixel (p, q) under test (supposed homogeneous in terms of angular and spectral behaviour) are used:

$$\{\mathbf{c}_j[p - \ell_1, q - \ell_2]\}_{\substack{\ell_1 = -K_1, \dots, K_1 \\ \ell_2 = -K_2, \dots, K_2 \\ (\ell_1, \ell_2) \neq (0, 0)}}$$

with $K = (2K_1 + 1)(2K_2 + 1) - 1$.

We consider two covariance matrix estimators on wavelet feature vectors: the standard Sample Covariance Matrix (SCM) which can be written, under zero-mean wavelet coefficient assumption, in the form

$$\begin{aligned} \hat{\Sigma}_{\text{SCM}, j}[p, q] &= \frac{1}{K} \times \\ &\sum_{\substack{\ell_1 = -K_1, \dots, K_1 \\ \ell_2 = -K_2, \dots, K_2 \\ (\ell_1, \ell_2) \neq (0, 0)}} \mathbf{c}_j[p - \ell_1, q - \ell_2] \mathbf{c}_j^H[p - \ell_1, q - \ell_2], \end{aligned} \quad (2.20)$$

and, as alternative to SCM (which can have poorer performance under generalized CES model assumption), the Tyler's Estimator (TE) which has proven some robustness in both Gaussian and non-Gaussian cases and which is defined as the solution of the fixed point equation [Pascal et al., 2008b]:

$$\begin{aligned} \hat{\Sigma}_{\text{TE}, j}[p, q] &= \frac{N}{K} \times \\ &\sum_{\substack{\ell_1 = -K_1, \dots, K_1 \\ \ell_2 = -K_2, \dots, K_2 \\ (\ell_1, \ell_2) \neq (0, 0)}} \frac{\mathbf{c}_j[p - \ell_1, q - \ell_2] \mathbf{c}_j^H[p - \ell_1, q - \ell_2]}{\mathbf{c}_j^H[p - \ell_1, q - \ell_2] \hat{\Sigma}_{\text{TE}, j}^{-1}[p, q] \mathbf{c}_j[p - \ell_1, q - \ell_2]}. \end{aligned} \quad (2.21)$$

The TE estimator is robust to non-stationarities that are naturally present in HR SAR images.

For both SCM and TE estimators, the number K has to be around $K \approx 2N$ for a good estimation [Robey et al., 1992]. For high values of R or L , the vectors become very large. The target detection problem necessitates the estimation of the covariance matrix of the noise in order to compute the detectors. This covariance is estimated using a window with some guard pixels around the test pixel. As the dimension increases, more samples are needed for the estimation which means that we have to increase the size of the window. This leads to a very poor spatial resolution of detection. That's why we usually prefer to keep the size of the window to the lowest possible value. In those cases, regularized versions of SCM and TE exist in the literature and have shown good results for many applications [Chen et al., 2011, Pascal et al., 2014].

Note that both estimators are used to estimate the covariance matrix of the clutter around a target. Hence, the test pixel (namely $(\ell_1, \ell_2) \neq (0, 0)$) is excluded in the process.

Detection Schemes

Target detection schemes usually consist in estimating the statistics of the background around the possible location of a target and computing a ratio measuring the likelihood of a target being present. Most of the time, an appropriate scaling is chosen to guarantee a certain probability of false alarm. This is possible if the statistic of the ratio is independent of the data being tested and is known as the Constant False Alarm Rate (CFAR) property (see [Li and Zelnio, 1996] or [Tao et al., 2016] for an accurate description of such detection schemes).

We assume that a target with a known steering vector $\mathbf{p} \in \mathbb{C}^N$ could be present in some pixels in the SAR image. It is an a priori information about the spectro-angular behaviour of the target we want to detect. In practice, it can be obtained by experimental measurements or by means of wavelet transform on an image where we have isolated the target¹. We have for each pixel $I[p, q]$ to solve the standard binary hypothesis test:

$$\begin{cases} H_0 : \mathbf{c}_j[p, q] = \mathbf{n}, & \mathbf{c}_k = \mathbf{n}_k \quad \forall k \in [1, K] \\ H_1 : \mathbf{c}_j[p, q] = a\mathbf{p} + \mathbf{n}, & \mathbf{c}_k = \mathbf{n}_k \quad \forall k \in [1, K], \end{cases} \quad (2.22)$$

where $(\mathbf{n}, \mathbf{n}_k)$ both represent a noise with the same distribution, a is an unknown complex amplitude of the potential target with spectro-angular steering vector \mathbf{p} to be detected and $\{\mathbf{c}_k\}_{k \in [1, K]}$ being the K secondary data.

In this detection issue, we decide to test different adaptive detectors like the well-known Adaptive Matched Filter which corresponds to a two-step Generalized Likelihood Ratio Test (GLRT) in homogeneous Gaussian noise [Robey et al., 1992]):

$$\Lambda_j^{AMF}[p, q] = \frac{\left| \mathbf{p}^H \hat{\Sigma}_{SCM,j}^{-1}[p, q] \mathbf{c}_j[p, q] \right|^2}{\mathbf{p}^H \hat{\Sigma}_{SCM,j}^{-1}[p, q] \mathbf{p}} \underset{H_0}{\gtrless} \underset{H_1}{\lambda}, \quad (2.23)$$

where λ is the detection threshold.

For partially homogeneous Gaussian noise or for CES distributed noises [Ollila et al., 2012a], the derivation of the detection problem leads to the Adaptive Normalized Matched

¹ Note that when the steering vector is not known, it is possible to develop Bayesian target detection schemes using works such as [Besson et al., 2017, Coluccia and Ricci, 2017].

Filter [Kraut et al., 2001, Greco and De Maio, 2016]:

$$\begin{aligned} \Lambda_j^{ANMF}[p, q] = & \\ & \frac{\left| \mathbf{p}^H \hat{\Sigma}_{TE,j}^{-1}[p, q] \mathbf{c}_j[p, q] \right|^2}{\left(\mathbf{p}^H \hat{\Sigma}_{TE,j}^{-1}[p, q] \mathbf{p} \right) \left(\mathbf{c}_j^H[p, q] \hat{\Sigma}_{TE,j}^{-1}[p, q] \mathbf{c}_j[p, q] \right)} \stackrel{H_1}{\geqslant} \lambda. \end{aligned} \quad (2.24)$$

The AMF detector has the Constant False Alarm (CFAR) property relative to the Gaussian distribution, while the ANMF is CFAR for both Gaussian and CES distributions. This is an important property since it allows to select a unique detection threshold to ensure a probability of false alarm (P_{Fa}) independently of the data being tested.

Concerning the complexity of these methods, the limiting factor is the need to compute the inverse of the covariance matrix for both AMF and ANMF schemes. Then, if the number of coefficients is high, this operation becomes time-consuming (typically $O((R \times L)^3)$).

Methodology for simulating targets

For a given SAR image, an artificial target with a given steering vector (representing its spectro-angular behavior) is embedded. In this way, we can control both position and Signal to Noise Ratio (SNR) of the target to be detected. For a given image I , steering vector $\mathbf{p} \in \mathbb{C}^{R \times L}$, a pixel $[i_t, j_t]^T$ corresponding to position $\mathbf{r}_t = [\mathbf{x}(i_t), \mathbf{y}(j_t)]^T$ and a given SNR in dB, an image with the target is obtained through:

$$I_t = I + \frac{T}{\|T\|_{l_2}} \sigma 10^{\frac{\text{SNR}_{\text{dB}}}{20}} \quad (2.25)$$

with:

$$T = \sum_{\substack{m=0, \dots, R^j-1 \\ n=0, \dots, L^j-1}} \mathcal{F}^{-1} \left\{ p[m, n] \mathbb{1}_{\Delta_j, \mathfrak{R}_m \times \Delta_j, \theta_n} e^{i 2 \pi \mathbf{k}^T \mathbf{r}_t} \right\} \text{ and}$$

$$\sigma^2 = \sum_{i=-10, \dots, 10} I(i_t + i, j_t + i)^2, \text{ the variance of the noise on a window around the target.}$$

T is an image generated to only possess the target at the position wanted. It is constructed in the spectral domain by generating a piecewise constant spectra where each constant part corresponds to the value of the steering vector in this band of frequency and interval of angles. Then it is multiplied by a phase term to input the location of the target. Finally, the image is obtained by inverse Fourier transform as usually done in SAR image reconstruction.

This process is done as follows:

- Choose a steering vector \mathbf{p} .
- Build spectrum according to the steering vector and create image of the targets using eq. (2.25)
- Perform the wavelet decomposition and create the hyperimage using eq. (2.13).
- Apply the detectors (2.23), (2.24) with the given steering vector.

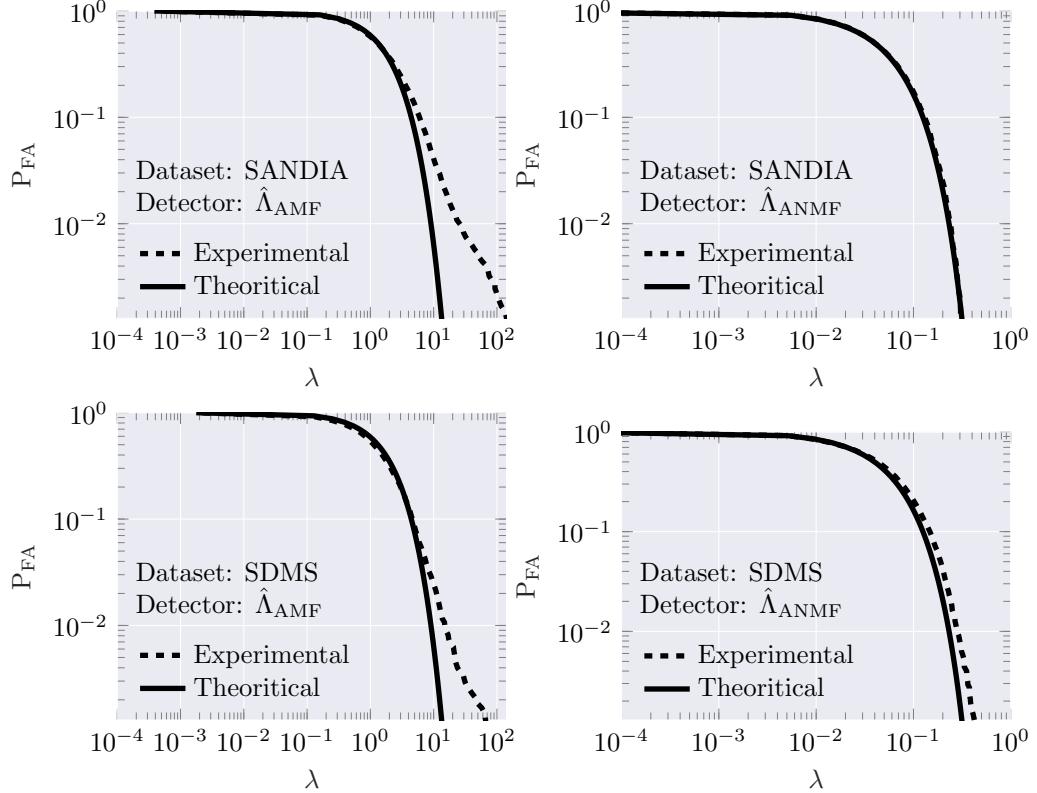


Figure 2.13: P_{FA} - λ curves ($j = 1, R = 5, L = 5, K = 88$). $d_1 = d_2 = \infty$. Left: AMF, Right: ANMF. Top: SANDIA dataset, Bottom: SDMS dataset.

Analysis of Pfa regulation over datasets

Next, we plot in Figure 2.13, the P_{FA} - λ (probability of false alarm versus threshold of detection) relationships for both AMF and ANMF detectors to study the CFAR behavior of the detectors on the datasets. We choose a random steering vector and apply detectors on the image without any target. It can be observed that the ANMF detector fares a lot better in terms of regulation of false alarm than the AMF. When compared to the theoretical relationship, the AMF detector has an experimental threshold higher at low P_{FA} whereas the ANMF detector stays close to its theoretical performance. This can be interpreted by the heterogeneous nature of the datasets which are not well modeled by Gaussian assumption.

Analysis of a scenario of detection near a bright point

We choose $R = L = 5$ and $j = 1$ and we place the target to be detected, with an SNR of 20 dB, near a synthetic bright point with Gaussian spectro-angular behavior. Figure 2.14 gives the steering vector of the target to detect, the spectrum of both targets and the image obtained by the procedure presented previously. The dataset used here is the SANDIA one.

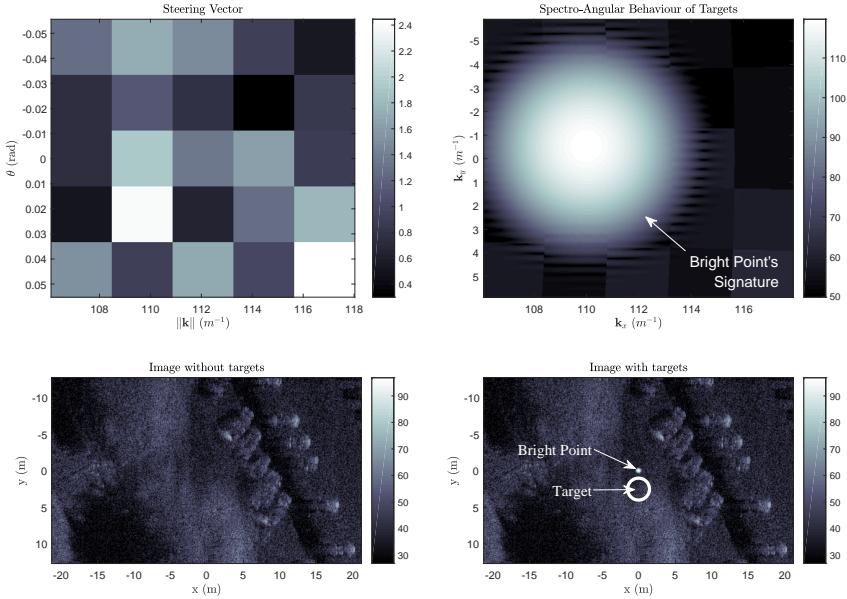


Figure 2.14: A target near a bright point. Dataset is SANDIA, $R = L = 5$. The target has an amplitude of -60 dB when compared to the bright point

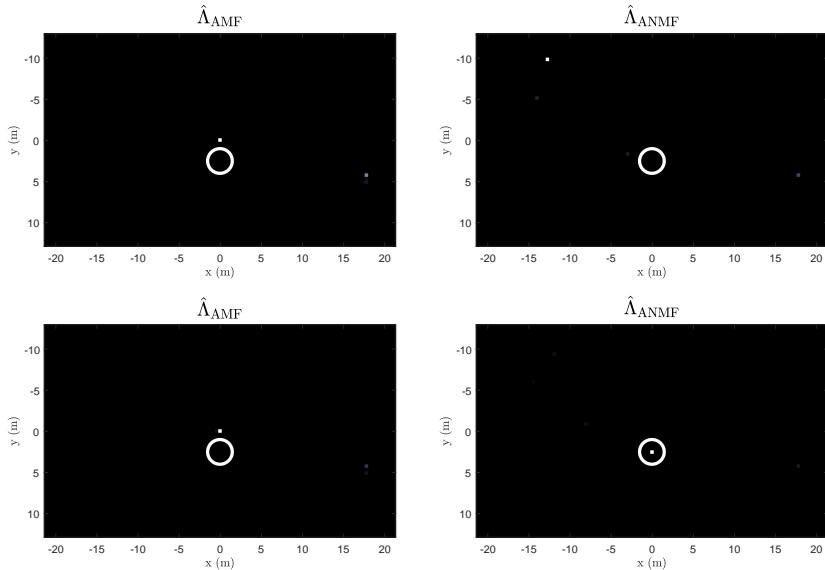


Figure 2.15: Results at $P_{FA} = 10^{-3}$. Top: $d_1 = d_2 = \infty$. Bottom: $d_1 = d_2 = 10$. Dataset is SANDIA, $R = L = 5$. The target has an amplitude of -60 dB when compared to the bright point.

Then we apply both detectors on the wavelet coefficients characterized by $d_1 = d_2 = \infty$ and $d_1 = d_2 = 10$. Figure 2.15 shows the result of the detection at $P_{FA} = 10^{-3}$. The

threshold guaranteeing the P_{FA} was picked-up from the experimental curves of figure 2.13.

Discussion: We focus first on the test of detection with $d_1 = d_2 = \infty$ (top of the figure). It can be observed that for both detectors, the target is not detected. The AMF detector outputs a false detection at the position of the bright point, which is expected given that this detector is mostly a power-based detector, whereas, the ANMF detector does not detect the bright point as it does not have similar spectro-angular behavior as the steering vector. However, a false alarm is still present which can be explained by a similar of the scene signature than the steering vector. If we take a look at detection tests for $d_1 = d_2 = 10$ (bottom of the figure), we observe that the target is detected with the ANMF detector but not the AMF one. This can be explained by the fact that with $d_1 = d_2 = 10$, we have reduced the sidelobes of the bright point which does not pollute the pixel of the target anymore resulting in better detection. The AMF detector does not yield better results for the same reason as previously.

These results confirm that the parametrization of the wavelet decomposition impacts the performance of detection.

P_D -SNR Curves

By randomizing the spatial location of the target for Monte Carlo trials, we obtain P_D -SNR relationships for both detectors presented in Figure 2.16. The steering vector is set to fixed values for all the trials.

Discussion: We first observe that the ANMF detector performs better in terms of detection than the AMF one for both datasets: if we look at $P_D = 0.7$, a gain of almost 7 dB is observed for the ANMF for SDMS dataset and 10 dB for SANDIA Dataset. This can be interpreted by the non-Gaussian nature of the data which makes regulation of false alarm difficult for the AMF detector and by the fact that ANMF is better suited for heterogeneous data.

The plots for the SANDIA dataset show overall lower performance than the SDMS dataset. This can be explained by the different nature of the datasets: the SANDIA image is more heterogeneous than the SDMS one and the speckle noise is more important.

Next, the different plots for each $d_1 = d_2$, lead to a significant gain when considering $d_1 = d_2 = 3$ or 10 compared to $d_1 = d_2 = \infty$ for the SDMS dataset (about 8 dB at $P_D = 0.6$). This result is coherent with the observations done previously when considering near bright-point detection. Indeed, the sidelobes are contained in the secondary data that is used for the estimation of the covariance matrix. These outliers lead to a loss of accuracy in the estimation which in turn decrease the performances of detection. It can be also observed a loss in detection for $d_1 = d_2 = 1$. This is coherent with the analysis of the previous section.

The gain using the new wavelet is lower on the SANDIA image (about 1 dB at $P_D = 0.6$). It is to be expected since the SANDIA dataset contains few bright points spread over the scene.

Impact of the steering vector

In order to assess the impact of the steering vector, another Monte-Carlo simulation has been done by setting the SNR to 0 dB and randomizing the target signature at each trial. For each trial, the target has been set to 100 different location to compute a probability of detection. Table 2.2 gives the performance of detection for two values of $d_1 = d_2$ and for both datasets. The same conclusions as previously can be drawn: the ANMF detector performs better than the AMF one on both datasets and using a bell-shaped wavelet with a parameter $d_1 = d_2 = 10$ allows to improve the detection rate.

SDMS				SANDIA				
		mean	min	max		mean	min	max
AMF	$d_1 = 10$	0.81	0.67	0.91	$d_1 = 10$	0.04	0	0.11
	$d_1 = \infty$	0.50	0.42	0.70	$d_1 = \infty$	0.02	0	0.10
ANMF	$d_1 = 10$	0.95	0.89	1	$d_1 = 10$	0.50	0.34	0.71
	$d_1 = \infty$	0.79	0.68	0.92	$d_1 = \infty$	0.46	0.22	0.66

Table 2.2: Results of detection when randomising steering vector. SNR=0 dB, 100 different signatures have been tested for 100 different target positions on each image.

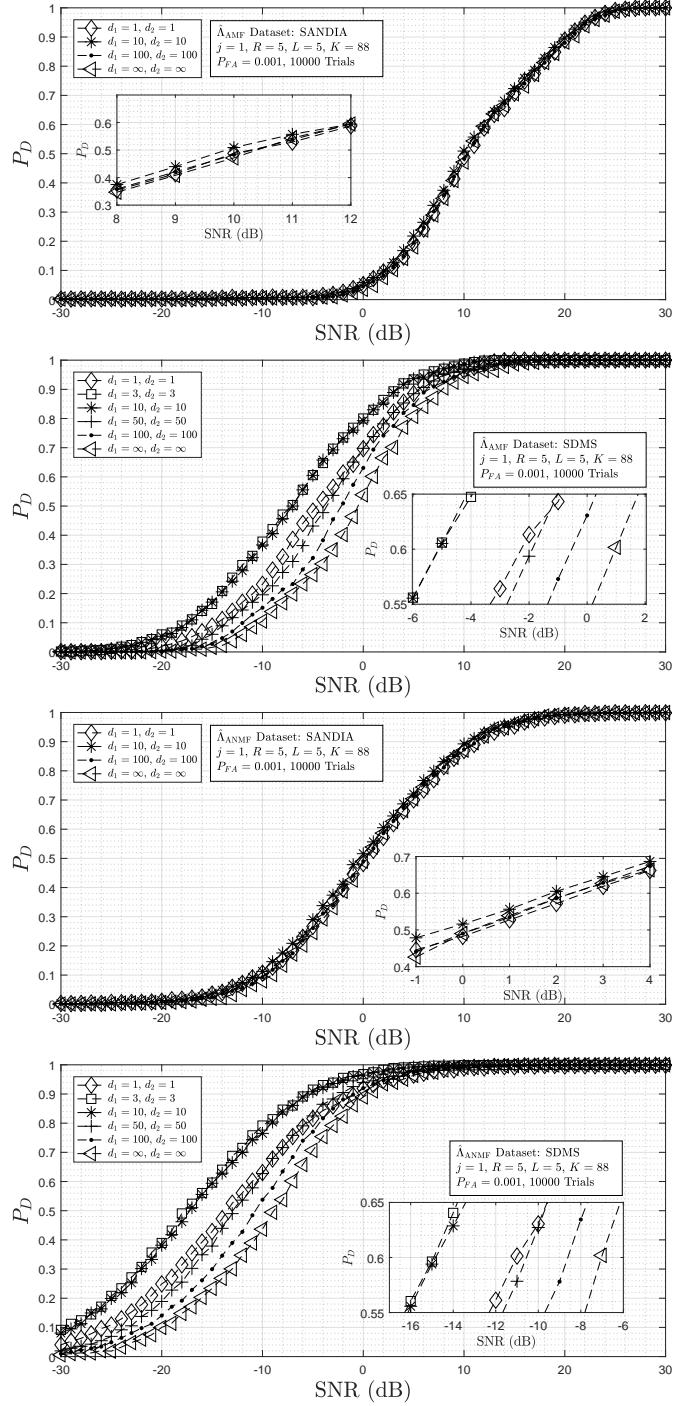


Figure 2.16: P_d -SNR plots for several values of $d_1 = d_2$. Top: AMF Detector. Bottom: ANMF Detector.

2.5 CONCLUSIONS

This chapter presented an adaptation of Shannon wavelet packets to SAR geometry in order to retrieve a physical diversity of interest. To reduce the side lobes, which are inherent to time frequency decomposition, a new family of parametrized wavelets has been proposed. These wavelets have the Shannon wavelets as a limit case and are tuned using a redundancy criterion.

This wavelet decomposition has been used in target detection schemes. It has been shown that the spectro-angular diversity, inherent to HR SAR Images, can be used in classic adaptive detection framework. First, the robust framework has proven to be more effective over the Gaussian one in both false alarms regulation and performance of detection. Then, the reduction of side lobes with the new family of wavelets, yields significantly gain in the performance of detection when the image contains numerous bright points.

2.A APPENDIX

2.A.1 Proof of proposition 2.2.1 at p. 59

Proposition. *The coefficients of the projection of g on a wavelet packet subspace $\Phi_{j,[n_1,n_2]}^S$ is a discrete sequence $\mathbf{C}_{j,[n_1,n_2]} = (\mathbf{C}_{j,[n_1,n_2]}^S[p, q])_{p,q \in \mathbb{Z}}$ where*

$$\mathbf{C}_{j,[m,n]}^S[p, q] = M_1^{-j/2} M_2^{-j/2} \mathcal{F}^{-1} U_{j,[m,n]}(M^j p, M^j q), \quad (2.26)$$

with

$$U_{j,[m,n]} = [\mathcal{F}g] \times \mathbf{1}_{\Delta_{j,G(m)}^0 \times \Delta_{j,G(n)}^1}. \quad (2.27)$$

Proof. Defining the wavelet coefficients from the following integral

$$\mathbf{C}_{j,[n_1,n_2]}^S[p, q] = \iint_{\mathbb{R}^2} g(z, t) \tau_{[M_1^j p, M_2^j q]} \Phi_{j,[n_1,n_2]}^S(z, t) dz dt.$$

we have (through Parseval):

$$\begin{aligned} \mathbf{C}_{j,[m,n]}^S[p, q] &= \frac{1}{4\pi} \times \\ &\iint \mathcal{F}g(\omega_1, \omega_2) \overline{\mathcal{F}\tau_{[M_1^j p, M_2^j q]} \Phi_{j,[m,n]}^S(\omega_1, \omega_2)} d\omega_1 d\omega_2. \end{aligned}$$

As (Fourier transform property on the translation)

$$\begin{aligned} \overline{\mathcal{F}\tau_{[M_1^j p, M_2^j q]} \Phi_{j,[m,n]}^S}(\omega_1, \omega_2) &= e^{i(M_1^j p \omega_1)} e^{i(M_2^j q \omega_2)} \times \\ &\mathcal{F}\Phi_{j,[m,n]}^S(\omega_1, \omega_2), \end{aligned}$$

we obtain:

$$\begin{aligned} \mathbf{C}_{j,[m,n]}^S[p, q] &= \frac{1}{4\pi} \times \\ &\iint e^{i(M_1^j p \omega_1)} e^{i(M_2^j q \omega_2)} \mathcal{F}g \mathcal{F}\Phi_{j,[m,n]}^S(\omega_1, \omega_2) d\omega_1 d\omega_2. \end{aligned}$$

The integral corresponds to the 2D inverse Fourier transform of $[\mathcal{F}g] \Psi_{j,[m,n]}^S$. \square

2.A.2 Proof of proposition 2.2.2 at p. 62

Proposition (Vanishing moments). *For any non-negative integers j, m, n, p, q , we have*

$$\iint_{\mathbb{R}^2} x^p y^q \Phi_{j,[m,n]}^{S,\angle}(x, y) dx dy = 0.$$

Proof. Function $\Phi_{j,[m,n]}^{S,\angle}(x, y)$ being separable with respect to x and y , we have:

$$\begin{aligned} \iint_{\mathbb{R}^2} x^p y^q \Phi_{j,[m,n]}^S(x, y) dx dy &= \\ &-i^{p+q} \frac{d}{d\mathfrak{K}^p} \frac{d}{d\theta^q} \Psi_{j,[m,n]}^{S,\angle}(\mathfrak{K}, \theta) \Big|_{\mathfrak{K}=0, \theta=0} \end{aligned}$$

Proposition 2.2.2 follows by noting that $\Psi_{j,[m,n]}^S$ has null derivatives. \square

2.A.3 Proof of proposition 2.2.3 at p. 62

Proposition (Orthogonality of wavelet packet subspaces). *For any given j and any $(m, n) \neq (m', n')$ we have:*

$$\Phi_{j,[m,n]}^{\mathbf{S},\angle} \perp \Phi_{j,[m',n']}^{\mathbf{S},\angle},$$

where \perp denotes orthogonality symbol.

Proof.

$$\begin{aligned} \left\langle \Phi_{j,[m,n]}^{\mathbf{S},\angle}, \Phi_{j,[m',n']}^{\mathbf{S},\angle} \right\rangle &= \\ \iint_{\mathbb{R}^2} \Phi_{j,[m,n]}^{\mathbf{S},\angle}(x,y) \overline{\Phi_{j,[m',n']}^{\mathbf{S},\angle}(x,y)} dx dy \end{aligned}$$

By using Parseval formula, we derive

$$\begin{aligned} \left\langle \Phi_{j,[m,n]}^{\mathbf{S},\angle}, \Phi_{j,[m',n']}^{\mathbf{S},\angle} \right\rangle &= \\ \frac{1}{4\pi^2} \iint_{\mathbb{R}^2} \mathcal{F}\Phi_{j,[m,n]}^{\mathbf{S},\angle}(\tilde{\kappa},\theta) \mathcal{F}\Phi_{j,[m',n']}^{\mathbf{S},\angle}(\tilde{\kappa},\theta) d\tilde{\kappa} d\theta. \end{aligned}$$

which reduces to

$$\begin{aligned} \left\langle \Phi_{j,[m,n]}^{\mathbf{S},\angle}, \Phi_{j,[m',n']}^{\mathbf{S},\angle} \right\rangle &= \\ \frac{R^j L^j}{4\pi^2} \iint_{\mathbb{R}^2} \mathbb{1}_{\Delta_{j,\tilde{\kappa}_m} \times \Delta_{j,\theta_n}}(\tilde{\kappa},\theta) \mathbb{1}_{\Delta_{j,\tilde{\kappa}_{m'}} \times \Delta_{j,\theta_{n'}}}(\tilde{\kappa},\theta) d\tilde{\kappa} d\theta. \end{aligned}$$

For $m \neq m'$ or $n \neq n'$, intersection $\Delta_{j,\tilde{\kappa}_m} \times \Delta_{j,\theta_n} \cap \Delta_{j,\tilde{\kappa}_{m'}} \times \Delta_{j,\theta_{n'}}$ of the supports of $\mathcal{F}\Phi_{j,[m,n]}^{\mathbf{S}}$ and $\mathcal{F}\Phi_{j,[m',n']}^{\mathbf{S}}(\tilde{\kappa},\theta)$ are either disjoint, or reduce to a null set. This ends the proof. \square

2.A.4 Proof of proposition 2.2.4 at p. 63

Proposition (Completion of wavelet packet subspaces). *For any given j , and any (m, n) , we have*

$$\bigcup_{\substack{m=0,1,\dots,R^j-1 \\ n=0,1,\dots,L^j-1}} \Phi_{j,[m,n]}^{\mathbf{S},\angle} = \mathbf{U}^{\mathbf{S},\angle}.$$

Proof. The proof is a consequence of Shannon band-limited function representation and the fact that for any $j \geq 1$, the sets $\Delta_{j,\tilde{\kappa}_m} \times \Delta_{j,\theta_n}$, for $m = 0, 1, \dots, R^j - 1$ and $n = 0, 1, \dots, L^j - 1$, are constructed so as to form a partition of \mathcal{D} . \square

2.A.5 Proof of proposition 2.3.2 at p. 65

Proposition (Convergence to Shannon wavelets). *The R -band L -look wavelet transform obtained by using Eq. (2.16) is associated with the Shannon wavelet transform when $d_1, d_2 \rightarrow +\infty$:*

$$\lim_{d_1 \rightarrow +\infty} \lim_{d_2 \rightarrow +\infty} \Psi_{j,[m,n]}^{[d_1,d_2],\angle} \stackrel{a.e.}{=} \Psi_{j,[m,n]}^{\mathbf{S},\angle}. \quad (2.28)$$

where equality holds true almost everywhere (a.e.).

Proof.

It suffices to show that $\lim_{d_1, d_2 \rightarrow +\infty} \Psi_{0,[0,0]}^{[d_1, d_2], \mathcal{L}} = \Psi_{0,[0,0]}^{\mathbf{S}, \mathcal{L}}$. From Eq. (2.17), we have

$$H_{0,[0,0]}^{[d_1, d_2]}(\mathfrak{K}, \theta) = \frac{1}{1 + \left| \frac{2}{\mathfrak{K}_B} (\mathfrak{K} - \mathfrak{K}_0) \right|^{2d_1}} \frac{1}{1 + \left| \frac{\theta}{\theta_B} \right|^{2d_2}}, \quad (2.29)$$

As a consequence, if $|\mathfrak{K} - \mathfrak{K}_0| < \frac{\mathfrak{K}_B}{2}$ and $|\theta| < \theta_B$, then:

$$\lim_{d_1, d_2 \rightarrow +\infty} H_{0,[0,0]}^{[d_1, d_2]}(\mathfrak{K}, \theta) = 1.$$

In contrast, if $|\mathfrak{K} - \mathfrak{K}_0| > \frac{\mathfrak{K}_B}{2}$ or $|\theta| > \theta_B$, then:

$$\lim_{d_1, d_2 \rightarrow +\infty} H_{0,[0,0]}^{[d_1, d_2]}(\mathfrak{K}, \theta) = 0,$$

which ends the proof. □

CHAPTER 3

MUTIVARIATE STATISTICAL ANALYSIS FOR ROBUST CHANGE DETECTION

Contents:

3.1	Description of the problem	84
3.1.1	Definitions	84
3.1.2	Defining the problem of covariance equality testing	84
3.2	Adapting robust M-estimation and Tyler's theory to covariance testing	85
3.2.1	GLRT under Elliptical model	86
3.2.2	GLRT under CAE model	86
3.2.3	GLRT under deterministic compound-Gaussian model	88
3.2.4	Convergence considerations	90
3.2.5	Statistical properties	93
3.2.6	Performance study of new statistics	97
3.3	Application of robust statistics to SAR change detection	101
3.3.1	Data description	101
3.3.2	First analysis	103
3.4	Conclusions	108
3.A	Appendix	109
3.A.1	Proof of proposition 3.2.1 at p. 86	109
3.A.2	Proof of proposition 3.2.2 at p. 88	111
3.A.3	Proof of proposition 3.2.3 at p. 89	111
3.A.4	Proof of proposition 3.2.4 at p. 89	115
3.A.5	Proof of proposition 3.2.5 at p. 90	117
3.A.6	Proof of proposition 3.2.6 at p. 91	120
3.A.7	Proof of proposition 3.2.7 at p. 91	121
3.A.8	Proof of proposition 3.2.8 at p. 92	122
3.A.9	Proof of proposition 3.2.10 at p. 95	124
3.A.10	Proof of proposition 3.2.12 at p. 95	125
3.A.11	Proof of proposition 3.2.13 at p. 96	125
3.A.12	Proof of proposition 3.2.14 at p. 96	127
3.B	Detection Methods Based on Structured Covariance Matrices for Multivariate SAR Images Processing	127
3.C	Robust Low-Rank Change Detection for SAR Image Time Series	133

In this chapter, we consider the problem of testing the equality of covariances matrices in a robust framework and its application to SAR change detection. We mentioned in the introductory chapter 1 that although the testing of covariances is a promising methodology for change detection in SAR images, the results could be limited by the Gaussian model assumption in high-resolution images. Thus, we consider here the development of new statistics for testing the equality of covariances using broader and more robust data models: the elliptically distributed random vectors and the deterministic compound-Gaussian model.

We consider the derivation of statistics using the generalized likelihood technique for both elliptical and deterministic compound-Gaussian models. The new derived statistics require the computation of fixed-point estimates for which we will consider the convergence properties. Next, we will consider the statistical properties (constant false alarm rate) of the new statistics. Finally, simulations on synthetic and real data are undertaken to study the effectiveness of the proposed approaches.

An extension of this work when the covariance matrices are assumed to be low rank has also been investigated for both Gaussian and deterministic compound-Gaussian models. This work has been done in collaboration with Arnaud Breloy of Paris Nanterre University. In order to keep a synthetic chapter centered about the extension of standard approaches to non-Gaussian model, we will omit this study in the main body of this chapter and provide alternatively the two papers describing these approaches in appendices 3.B and 3.C.

Notations relevant to his chapter: Given $(a, b) \in \mathbb{N}^2$, $b \geq a$, $\llbracket a, b \rrbracket$ denotes the set $\{a, \dots, b\}$. δ_{ik} is the Kronecker symbol. Θ is an arbitrary parameter space. Given a scalar valued function f , $\frac{\partial f}{\partial \bullet}$ denotes the gradient of f w.r.t \bullet arranged in a column. \mathbf{x} will always represent a random vector of size p . Any subscript or superscript serves to indicate a specific observation. Σ and ξ will always be Hermitian matrices of size $p \times p$ and ξ will always be normalized by the trace. The symbol \sim means "distributed as". H_0 and H_1 denote both possible hypothesis in a binary hypotheses test scheme. The abbreviation i.i.d means independent identically distributed. To simplify the equations, we define the following quantities:

$$\begin{aligned} q(\xi, \mathbf{x}) &= \mathbf{x}^H \xi^{-1} \mathbf{x}, \\ \forall k, \forall t, \mathbf{S}_k^t &= \mathbf{x}_k^t \mathbf{x}_k^{t^H}. \end{aligned} \tag{3.1}$$

3.1 DESCRIPTION OF THE PROBLEM

In this section, we remind the problem of covariance equality testing in the general context.

3.1.1 Definitions

Denote by $\mathbb{W}_{1,T} = \{\mathbf{X}_1, \dots, \mathbf{X}_T\}$ a collection of T mutually independent groups of p -dimensional i.i.d complex vectors: $\mathbf{X}_t = [\mathbf{x}_1^t, \dots, \mathbf{x}_N^t] \in \mathbb{C}^{p \times N}$. With regards to SAR images these sets correspond to the local observations on a spatially sliding windows of size $N_1 \times N_2$ as illustrated in Figure 1.10. The subscript k corresponds to a spatial index while the superscript t corresponds to a time index.

3.1.2 Defining the problem of covariance equality testing

The testing of equality of covariances, sometimes referred as homogeneity of covariances, can be written as a binary hypothesis testing problem of the following form:

$$\left\{ \begin{array}{l} H_0 : \theta_1 = \dots = \theta_T = \theta_0 \quad \& \quad \Phi_1 \neq \dots \neq \Phi_T, \\ H_1 : \exists (t, t') \in \llbracket 1, T \rrbracket^2, \theta_t \neq \theta_{t'} \quad \& \quad \Phi_1 \neq \dots \neq \Phi_T \end{array} \right., \tag{3.2}$$

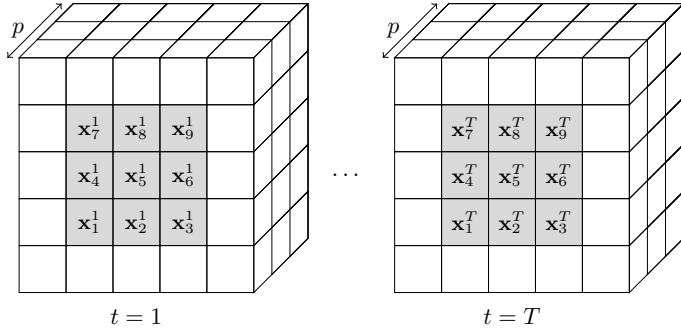


Figure 3.1: Illustration of local data selection ($N_1 = N_2 = p = 3$) for detection test. The gray area corresponds to $\mathbb{W}_{1,T}$ and the central pixel (\mathbf{x}_5^t) is the test pixel.

where $\boldsymbol{\theta}$ is either covariance, scale parameter only or shape matrix only. The nuisances parameters which we do not want to test are denoted Φ . These nuisances parameters are either scale, shape matrix or the empty space, depending on the definition of $\boldsymbol{\theta}$.

The problem is formulated as testing that all the parameters $\boldsymbol{\theta}_t$ are equal to some value $\boldsymbol{\theta}_0$ against all possible alternatives. In this case, the derivation of the GLRT must take into account the worst possible case which is that all parameters are different.

It would be possible to test others alternatives that some subgroup of $\boldsymbol{\theta}_t$ are equal while other are different but there are many possibilities which are incompatible between themselves. This is also why there is no Uniformly Most Powerful test for this problem. So the only possibility way to compute a GLRT is to assume that every parameter is different which was the methodology used in [Anderson, 2003].

3.2 ADAPTING ROBUST M-ESTIMATION AND TYLER'S THEORY TO COVARIANCE TESTING

We saw in the introductory chapter 1 that the covariance equality test is a promising approach to local change-detection in SAR images. We explained that this approach was considered using the Gaussian model solely, which could be inappropriate for high-resolution SAR images. Thus we introduced the elliptical families of distributions better suited for modeling the data distribution, presented some results of the statistical litterature under this assumption and explained how these approaches did not yield satisfactory results in SAR data analysis. Then we saw in chapter 2 that the robust Tyler estimator associated with the ANMF detection scheme had significantly improved the detection performance in target detection problems for SAR images. Thus, the non-Gaussian framework is valid when it concerns the processing of multivariate SAR data. In those regards, it would be interesting to consider these robust approaches in the context of change detection using covariance equality method. In this section, we will consider the derivation of new detection using GLRT methodology for both elliptical and deterministic compound-Gaussian model. The analysis of the obtained statistics will rely on the well-known M-estimation and Tyler theory of covariance matrix estimation.

3.2.1 GLRT under Elliptical model

We consider here the elliptical model defined in 1.3 associated with the problem of detecting a change in both shape matrix and scale parameter:

$$\begin{aligned}\boldsymbol{\theta}_t &= \{\boldsymbol{\Sigma}_t\} = \{\tau_t, \boldsymbol{\xi}_t\}, \\ \boldsymbol{\Phi}_t &= \emptyset.\end{aligned}\tag{3.3}$$

Here the density generator function is assumed to be the same for every single date and is known. We have the following result:

Proposition 3.2.1. *The GLRT under problem (3.3) is the following quantity:*

$$\hat{\Lambda}_{\mathbb{C}\mathcal{E}}^g = \frac{\left| \hat{\boldsymbol{\Sigma}}_0^M \right|^{TN}}{\prod_{t=1}^T \left| \hat{\boldsymbol{\Sigma}}_t^M \right|^N} \prod_{k=1}^{N-T} \frac{g\left(q(\hat{\boldsymbol{\Sigma}}_t^M, \mathbf{x}_k^t)\right)}{g\left(q(\hat{\boldsymbol{\Sigma}}_0^M, \mathbf{x}_k^t)\right)} \stackrel{H_1}{\gtrless} \lambda, \tag{3.4}$$

where:

$$\hat{\boldsymbol{\Sigma}}_t^M = f_t(\hat{\boldsymbol{\Sigma}}_t^M), \quad \hat{\boldsymbol{\Sigma}}_0^M = \frac{1}{T} \sum_{t=1}^T f_t(\hat{\boldsymbol{\Sigma}}_0^M) \text{ and}$$

$$f_t(\boldsymbol{\Sigma}) = \frac{1}{N} \sum_{k=1}^N \frac{-g'(q(\boldsymbol{\Sigma}, \mathbf{x}_k^t))}{g(q(\boldsymbol{\Sigma}, \mathbf{x}_k^t))} \mathbf{x}_k^t \mathbf{x}_k^{t^H} \tag{3.5}$$

Proof. See Appendix 3.A.1 □

Discussion: The statistic obtained shares similarities to the Gaussian GLRT one. We recognize indeed the ratio of determinants which is equal to one if all matrices are equals. The estimates, however, are no longer the SCM of the samples but can be seen as the M-estimators described in [Ollila et al., 2012a] if we consider the ratio $u = -g'/g$ as an arbitrary function not dependent of g with some regularity conditions. However, in this case, the choice of a function u different from the ratio $-g'/g$ is of no interest since the statistic requires the knowledge of the density generator function g . This in practice is a very restrictive hypothesis that cannot be guaranteed since SAR images are diverse. Indeed, we could estimate the g function and we wouldn't even need to rederive the GLRT since we just have to replace the g by its expression in the expression of the GLRT at eq. (3.4). The problem is that it requires to selection the model by fitting a large family of distribution which is a heavy approach. Moreover, we can't guarantee that the function will not change over time. In which case, the computation of the GLRT is more difficult.

Nonetheless, the study of convergence of fixed-point estimates still stands. This means that for any valid density generator g , the estimates will converge to a unique value allowing to compute the detection test.

3.2.2 GLRT under CAE model

An interesting alternative to tackle the problem of not knowing the density generator g is to consider the normalized observations: $\{\mathbf{z}_k^t = \mathbf{x}_k^t / \|\mathbf{x}_k^t\|_2 | 1 \leq k \leq N, 1 \leq t \leq T\}$. The intuitive idea behind this approach is normalizing two groups of observations with the same shape matrix will result in the same distribution over the unit sphere. As an example,

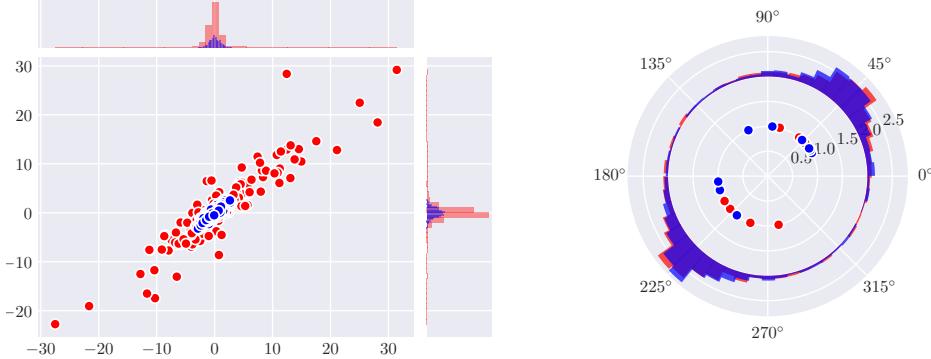


Figure 3.2: Example of normalization effect when the shape matrix is the same for two K-distributed group of samples (blue and red dots). Left: The original bivariate data. Right: Normalized data in polar representation. In polar representation both histograms coincide.

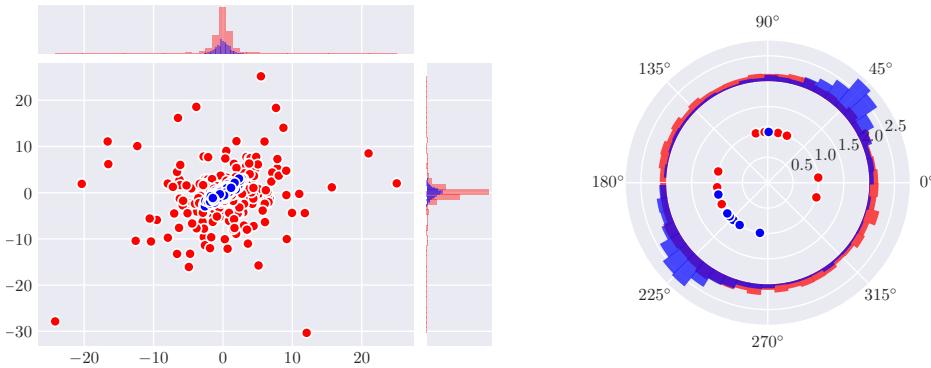


Figure 3.3: Example of normalization effect when the shape matrix is the different for two K-distributed group of samples (blue and red dots). Left: The original bivariate data. Right: Normalized data in polar representation. In polar representation the two histograms do not coincide.

Figures 3.2 and 3.3 present histograms of bivariate K-distributed groups of data for which the shape matrix is either the same or different. As described in the chapter 1, the p.d.f of those vectors is known to be \mathbb{CAE} distributed which is not dependent of the density generator function. However, since we normalized the vectors, the model only considers the shape matrix for which we can test equality through the following scheme:

$$\begin{aligned} \boldsymbol{\theta}_t &= \{\boldsymbol{\xi}_t\}, \\ \boldsymbol{\Phi}_t &= \emptyset. \end{aligned} \tag{3.6}$$

The derivation of the GLRT leads to the following result:

Proposition 3.2.2. *The GLRT under problem (3.6) is the following quantity:*

$$\hat{\Lambda}_{\mathcal{CAE}} = \frac{\left| \hat{\xi}_0^{\text{TE}} \right|^{TN} \prod_{t=1}^N \frac{\left(q(\hat{\xi}_0^{\text{TE}}, \mathbf{x}_k^t) \right)^p}{\left(q(\hat{\xi}_t^{\text{TE}}, \mathbf{x}_k^t) \right)^p}}{\prod_{t=1}^T \left| \hat{\xi}_t^{\text{TE}} \right|^N} \gtrless_{H_0}^{\text{H}_1} \lambda, \quad (3.7)$$

where:

$$\hat{\xi}_t^{\text{TE}} = f_t^{\text{TE}}(\hat{\xi}_t^{\text{TE}}), \quad \hat{\xi}_0^{\text{TE}} = \frac{1}{T} \sum_{t=1}^T f_t^{\text{TE}}(\hat{\xi}_0^{\text{TE}}) \text{ and}$$

$$f_t^{\text{TE}}(\boldsymbol{\xi}) = \frac{p}{N} \sum_{k=1}^N \frac{\mathbf{x}_k^t \mathbf{x}_k^t \mathbf{H}}{q(\boldsymbol{\xi}, \mathbf{x}_k^t)} \quad (3.8)$$

Proof. See Appendix 3.A.2 □

Discussion: The statistic obtained here can be seen as a special case of the one at eq. (3.4) where $g = 1/x^p$. This result is similar to the link between Maronna M-estimators [Maronna, 1976b] and Tyler's estimator [Tyler, 1987] where the same choice of $u = g'/g = 1/x$ leads from the M-estimator to the Tyler one. This is precisely what is observed in the fixed-point estimates where the MLE estimates are replaced by the Tyler estimate. The convergence of the estimate is a direct consequence of Tyler's analysis.

The fixed-point equation is computed as follows: We initialise to the identity matrix $(\hat{\xi})_0$ then we compute $(\hat{\xi})_{n+1} = f_t^{\text{TE}}((\hat{\xi})_n)$ until that the criterion:

$$\frac{\left\| (\hat{\xi})_{n+1} - (\hat{\xi})_n \right\|_{\text{F}}}{\left\| (\hat{\xi})_n \right\|_{\text{F}}} \geq \delta,$$

where δ is a tolerance value, is no longer respected.

The resulting statistic allows testing a change in the shape without regards to the density generator g which makes it a robust statistic. In fact, the assumption that the density function is the same for every date can be dropped here since the GLRT would yield the same result.

Although we have obtained a distribution-free statistic, it only allows to test a change in the shape matrix and discard any variation of the scale parameter over time. This can limit the effectiveness in SAR change detection applications where the scale accounts for the power of scatters and would thus be a piece of great information in order to account for changes. From this work, there appears to be no solution for testing jointly the scale and shape while being distribution free while considering the robust M-estimation/Tyler approach. We will thus restrict our model to the deterministic compound-Gaussian one for which derivations are possible.

3.2.3 GLRT under deterministic compound-Gaussian model

Since in this model, the scale parameter corresponds to a vector of deterministic values, we consider the following problems:

- Scale and shape matrix testing:

$$\begin{aligned}\boldsymbol{\theta}_t &= \{\tau_1^t, \dots, \tau_N^t, \boldsymbol{\xi}_t\}, \\ \boldsymbol{\Phi}_t &= \emptyset.\end{aligned}\tag{3.9}$$

- Shape matrix testing:

$$\begin{aligned}\boldsymbol{\theta}_t &= \{\boldsymbol{\xi}_t\}, \\ \boldsymbol{\Phi}_t &= \{\tau_1^t, \dots, \tau_N^t\}.\end{aligned}\tag{3.10}$$

- Scale testing:

$$\begin{aligned}\boldsymbol{\theta}_t &= \{\tau_1^t, \dots, \tau_N^t\}, \\ \boldsymbol{\Phi}_t &= \{\boldsymbol{\xi}_t\}.\end{aligned}\tag{3.11}$$

We derive hereafter the GLRT for each problem of detection:

Proposition 3.2.3. *The GLRT ratio under hypotheses of Problem (3.9) is the following:*

$$\hat{\Lambda}_{\text{MT}} = \frac{\left| \hat{\boldsymbol{\xi}}_0^{\text{MT}} \right|^{TN}}{\prod_{t=1}^T \left| \hat{\boldsymbol{\xi}}_t^{\text{TE}} \right|^N} \prod_{k=1}^N \frac{\left(\sum_{t=1}^T q(\hat{\boldsymbol{\xi}}_0^{\text{MT}}, \mathbf{x}_k^t) \right)^{Tp}}{T^{Tp} \prod_{t=1}^T \left(q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t) \right)^p} \stackrel{H_1}{\gtrless} \lambda, \tag{3.12}$$

where

$$\hat{\boldsymbol{\xi}}_0^{\text{MT}} = f_{N,T}^{\text{MT}} \left(\hat{\boldsymbol{\xi}}_0^{\text{MT}} \right) = \frac{p}{N} \sum_{k=1}^N \frac{\sum_{t=1}^T \mathbf{s}_k^t}{\sum_{t=1}^T q(\hat{\boldsymbol{\xi}}_0^{\text{MT}}, \mathbf{x}_k^t)}. \tag{3.13}$$

Proof. See Appendix 3.A.3. □

Discussion: The statistic obtained here is similar to the one obtained using the Gaussian assumption. The term involving determinant is the same except that now the estimates are the solution of a fixed-point equation. $\hat{\boldsymbol{\xi}}_t^{\text{TE}}$ is the Tyler estimator described previously. $\hat{\boldsymbol{\xi}}_0^{\text{MT}}$ is similar but corresponds to a different fixed-point equation involving the observations for all the dates. The properties of this new estimate will be studied in the next section.

Due to the normalization of covariance matrices, the term involving determinants is a test involving solely the structure of the covariance matrices and do not consider the relative power of the pixels between the dates. The ratio of the quadratic forms allows testing the change in power in the same way it is done for the correlations in the determinants term: a ratio of arithmetic and geometric means.

Proposition 3.2.4. *The GLRT ratio under hypotheses of problem (3.10) is the following:*

$$\hat{\Lambda}_{\text{Mat}} = \frac{\left| \hat{\boldsymbol{\xi}}_0^{\text{Mat}} \right|^{TN}}{\prod_{t=1}^T \left| \hat{\boldsymbol{\xi}}_t^{\text{TE}} \right|^N} \prod_{k=1}^N \frac{\left(q(\hat{\boldsymbol{\xi}}_0^{\text{Mat}}, \mathbf{x}_k^t) \right)^p}{\left(q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t) \right)^p} \stackrel{H_1}{\gtrless} \lambda, \tag{3.14}$$

where

$$\hat{\boldsymbol{\xi}}_0^{\text{Mat}} = f_{N,T}^{\text{Mat}}\left(\hat{\boldsymbol{\xi}}_0^{\text{Mat}}\right) = \frac{p}{TN} \sum_{k=1}^{k=N} \frac{\mathbf{S}_k^t}{q\left(\hat{\boldsymbol{\xi}}_0^{\text{Mat}}, \mathbf{x}_k^t\right)}.$$

Proof. See Appendix 3.A.4. \square

Discussion: Without surprise, the statistic obtained is exactly the same as the one obtained through the \mathcal{CAE} model. Indeed, it is known that Tyler estimator can be obtained doing an MLE estimation under either \mathcal{CAE} [Ollila et al., 2012a] or deterministic compound-Gaussian model [Pascal et al., 2008a]. Thus, the same result is expected when it comes to the GLRT which relies on an MLE estimation.

Proposition 3.2.5. *The GLRT ratio under hypotheses of problem (3.11) is the following:*

$$\hat{\Lambda}_{\text{Tex}} = \prod_{t=1}^T \frac{\left|\hat{\boldsymbol{\xi}}_t^{\text{Tex}}\right|^N}{\left|\hat{\boldsymbol{\xi}}_t^{\text{TE}}\right|^N} \prod_{k=1}^N \frac{\left(\sum_{t=1}^T q\left(\hat{\boldsymbol{\xi}}_t^{\text{Tex}}, \mathbf{x}_k^t\right)\right)^{Tp}}{T^{Tp} \prod_{t=1}^T \left(q\left(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t\right)\right)^p} \stackrel{H_1}{\gtrless} \lambda, \quad (3.15)$$

where

$$\hat{\boldsymbol{\xi}}_t^{\text{Tex}} = f_{N,T,t}^{\text{Tex}}\left(\hat{\boldsymbol{\xi}}_1^{\text{Tex}}, \dots, \hat{\boldsymbol{\xi}}_T^{\text{Tex}}\right), \quad (3.16)$$

$$= \frac{Tp}{N} \sum_{k=1}^N \frac{\mathbf{S}_k^t}{\sum_{t'=1}^T q\left(\hat{\boldsymbol{\xi}}_{t'}^{\text{Tex}}, \mathbf{x}_k^t\right)}. \quad (3.17)$$

Proof. See Appendix 3.A.5 \square

Discussion: In this last statistic, the detection is done solely on the texture parameters. This leads to an interesting estimation: each $\hat{\boldsymbol{\xi}}_t^{\text{Tex}}$ is solution of a fixed-point equation which involves all the estimates $\hat{\boldsymbol{\xi}}_{t'}^{\text{Tex}}$. In practice, this can lead to convergence issues when considering the computation. This problem will be treated in the next section and it can be shown that the estimates can be implemented simply.

3.2.4 Convergence considerations

Theoretical study of convergence

We consider here the validity of the alternate maximization done when deriving the new statistics and the convergence problems that arise. We will, however, limit ourselves to the case of compound-Gaussian derived statistics, since the optimizations in the \mathbb{CE} and \mathbb{CAE} cases are well-known problems for which extensive analysis exist¹. To this end, we consider geodesic convexity (g-convexity) on the manifold $\mathbb{S}_{\mathbb{H}}^p$ as presented in [Wiesel, 2012] which is defined as follows:

¹See for example [Ollila et al., 2012a, Wiesel, 2012] among many works.

Definition 1. (*Geodesic convexity*) Let \mathcal{M} be an arbitrary manifold. For each pair $q_0, q_1 \in \mathcal{M}$, we define a geodesic $q_t^{q_0, q_1} \in \mathcal{M}$ for $t \in [0, 1]$. A real valued function f with domain \mathcal{M} is g -convex if $f(q_t^{q_0, q_1}) \leq tf(q_1) + (1 - t)f(q_0)$ for any $q_0, q_1 \in \mathcal{M}$ and $t \in [0, 1]$.

The g -convexity, which is illustrated in Figure 3.4, extends the definition of the traditional Euclidean convexity to curved spaces. This concept is useful for optimizations done on covariances matrices which are known to lie on the $\mathbb{S}_{\mathbb{H}}^p$ manifold.

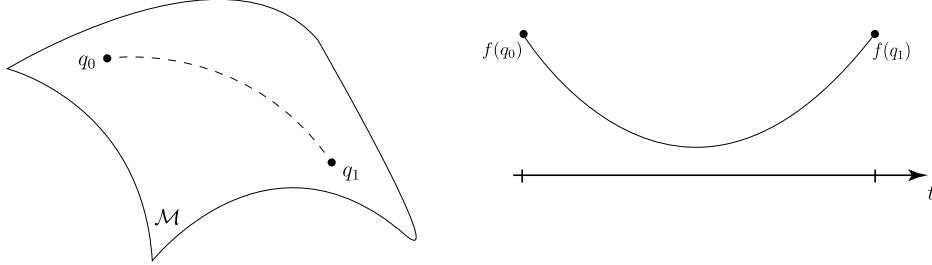


Figure 3.4: Illustration of g -convexity concept: For any set $(q_0, q_1) \in \mathcal{M}^2$, the function is convex when following the geodesic path between these two points.

Notably, we can use this property of the log-likelihood to show the following propositions:

Proposition 3.2.6. $\hat{\xi}_0^{\text{MT}}, \hat{\xi}_0^{\text{Mat}}$ and $\hat{\xi}_t^{\text{Tex}}$ are the unique arguments of the global maxima of their respective log-likelihood cost functions over the observations.

Proof. See Appendix 3.A.6. □

This proposition is necessary to justify the alternate maximization done when deriving the expression of the statistics. However, when considering optimization on manifolds, this in itself does not guarantee that the solution corresponding to the global maxima is part of the manifold. This point is important since we want a solution that is both computable and in the set $\mathbb{S}_{\mathbb{H}}^p$. The following proposition can be effectively shown:

Proposition 3.2.7. $\hat{\xi}_0^{\text{MT}}, \hat{\xi}_0^{\text{Mat}}$ and $\hat{\xi}_t^{\text{Tex}}$ are the arguments to the global minima obtained inside $\mathbb{S}_{\mathbb{H}}^p$.

Proof. See Appendix 3.A.7. □

Now that we know that the solution of the fixed-point equations is the unique arguments to the global maxima of their log-likelihood and that they are obtained inside the manifold $\mathbb{S}_{\mathbb{H}}^p$, the convergence of the fixed-point algorithms can be considered.

We have:

Proposition 3.2.8. Let $\{\mathbf{x}_k^t | k \in \llbracket 1, N \rrbracket, t \in \llbracket 1, T \rrbracket\}$ be a set of observations. Let us define vectors $\mathbf{v}_i \in \mathbb{R}^p$ such that $\forall k, \forall t, \mathbf{v}_{(T-1)*N+k} = (\Re(\mathbf{x}_k^t)^T, \Im(\mathbf{x}_k^t)^T)^T$ and $\mathbf{v}_{(2T-1)*N+k} = (-\Im(\mathbf{x}_k^t)^T, \Re(\mathbf{x}_k^t)^T)^T$. Let $\mathbb{P}_{2TN}(\bullet)$ be the empirical distribution of samples $\{\mathbf{v}_i | i \in \llbracket 1, 2TN \rrbracket\}$. Then the fixed-point algorithms $(\hat{\xi}_0^{\text{MT}})_{k+1} = f_{N,T}^{\text{MT}}((\hat{\xi}_0^{\text{MT}})_k)$ and $(\hat{\xi}_0^{\text{Mat}})_{k+1} = f_{N,T}^{\text{Mat}}((\hat{\xi}_0^{\text{Mat}})_k)$ converge to unique solutions up to a scale factor if and only if the following condition is respected:

Algorithm 3 Computation of $\hat{\xi}_t^{\text{Tex}}$

```

1: Initialize  $\forall t \in \llbracket 1, T \rrbracket$ ,  $(\hat{\xi}_t^{\text{Tex}})_0 = \mathbf{I}_p$ 
2: while  $\delta > \epsilon$  do
3:   for  $t \in \llbracket 1, T \rrbracket$  do
4:     Compute:

$$(\hat{\xi}_t^{\text{Tex}})_{n+1} = f_{N,T,t}^{\text{Tex}} ((\hat{\xi}_1^{\text{Tex}})_n, \dots, (\hat{\xi}_T^{\text{Tex}})_n).$$

5:   Impose Trace normalization by:

$$(\hat{\xi}_t^{\text{Tex}})_{n+1} = \frac{p (\hat{\xi}_t^{\text{Tex}})_{n+1}}{\text{Tr}((\hat{\xi}_t^{\text{Tex}})_{n+1})}.$$

6:   end for
7:   Compute criterion
8:    $\delta = \max \left\{ \frac{\left\| (\hat{\xi}_t^{\text{Tex}})_{n+1} - (\hat{\xi}_t^{\text{Tex}})_n \right\|}{\left\| (\hat{\xi}_t^{\text{Tex}})_n \right\|} / t \in \llbracket 1, T \rrbracket \right\}.$ 
9: end while

```

(C1) $\mathbb{P}_{2TN}(\{\mathbf{0}\}) = 0$ and for all linear subspaces $V \subset \mathbb{R}^{2p}$, we have $\mathbb{P}_{2TN}(V) < \dim(V)/2p$.

Proof. See Appendix 3.A.8. □

For practical purposes the condition (C1) can be achieved when there are at least $p + 1$ linearly independent observations \mathbf{x}_k^t , that is ensured if the size of the window is sufficient. Again, the uniqueness is guaranteed by the trace normalization which has to be imposed at each step of the algorithm. It is important to notice that the convergence of the algorithms is ensured for any set of observations $\{\mathbf{x}_k^t | t \in \llbracket 1, T \rrbracket, k \in \llbracket 1, N \rrbracket\}$ that respects condition (C1), even if the observations do not follow the same distributions (typically if the hypothesis H₁ is correct).

The case of $\hat{\xi}_t^{\text{Tex}}$ is in this regard trickier. Indeed, since each step requires the knowledge of the other estimates, we propose the cyclic Algorithm 3 which will iterate each matrix alternatively. While it is easy to show that if only one of the matrices is unknown², the fixed-point algorithm will converge, it is difficult to conclude on a theoretical standpoint about the convergence of the alternate estimation algorithm. Nonetheless, when doing extensive simulations, as will be shown shortly afterward, on both theoretical and real-data, there has been no case when the algorithm does not converge, except when the condition (C1) is not respected.

²using the same considerations as in the previous theorem.

Experimental study of convergence

In order to test the convergence property of matrix estimates, we consider realisations of random variables $\mathbf{x} = \sqrt{\tau} \mathbf{z}$ where τ follows a Γ -distribution with shape parameter α and scale parameter β . \mathbf{z} is generated through a Gaussian realization with covariance matrix chosen to be Toeplitz structured of the form:

$$\boldsymbol{\xi} = (\sigma_{m,n})_{\substack{1 \leq m \leq p, \\ 1 \leq n \leq p}} \text{ where : } \sigma_{m,n} = \rho^{|m-n|}.$$

We consider two settings: first, we generate a time series $\{\mathbf{x}_k^t | k \in \llbracket 1, N \rrbracket, t \in \llbracket 1, T \rrbracket\}$ where each \mathbf{x}_k^t is distributed with the same covariance matrix $\boldsymbol{\xi}_0$. Then, we generate a time series where each \mathbf{x}_k^t is distributed with a covariance matrix $\boldsymbol{\xi}_t$ different for each date. Figure 3.5 presents a Monte-Carlo (MC) simulation where the criterion $\delta = \frac{\|\boldsymbol{\xi}_{n+1} - \boldsymbol{\xi}_n\|_2}{\|\boldsymbol{\xi}_n\|_2}$ of convergence is plotted against the number of iterations n of the fixed-point algorithm. The plot shows that for whatever the setting, all estimates converge since the criterion reaches the working precision of the machine. We observe that $\hat{\boldsymbol{\xi}}_t^{\text{Tex}}$ needs more iterations to converge. This was expected, since in this case three different matrices were estimated while for the others a single matrix was computed. These results comfort the theoretical considerations of 3.2.4.

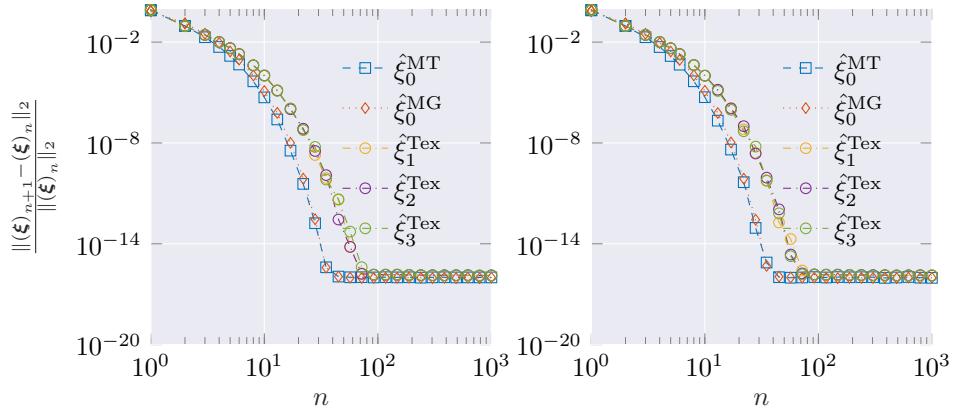


Figure 3.5: Convergence property of estimates. Left: Same matrices at each date ($p = 3$, $N = 25$, $T = 3$, 1000 Trials, $\rho_0 = 0.70$, $\alpha = 0.30$, $\beta = 0.10$). Right: Different matrices at each date ($p = 3$, $N = 25$, $T = 3$, 1000 Trials, $\rho_1 = 0.08$, $\rho_2 = 0.90$, $\rho_3 = 0.10$, $\alpha = 0.30$, $\beta = 0.10$). The textures are different at each date for both settings.

3.2.5 Statistical properties

Asymptotic distribution of statistics in Elliptical case

We have the following result:

Proposition 3.2.9. *The distribution of $2 \log(\hat{\Lambda}_{\mathcal{CE}}^g)$ under H_0 is asymptotically equivalent to a $\chi_{(T-1)p^2}^2$ distribution.*

Proof. Property inherited from Wilks theorem. □

This asymptotic result is a direct application of the Wilks theorem presented in 1.4.4 and allows to obtain an approximation when the sample size is large, which can be limited in practice.

This asymptotic behavior has been studied through simulations as well for the case of a Student-t distribution associated with the Student-t GLRT. Figure 3.6 shows the empirical distribution obtained through a small sample size compared to the theoretical one from proposition 3.2.9. It can be seen that the approximation is not accurate which is expected since the result is only asymptotic.

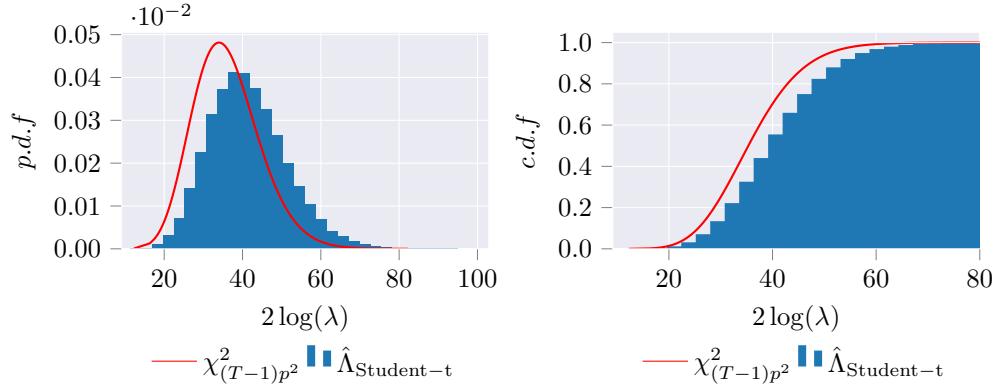


Figure 3.6: Distribution of statistic under small sample size ($p = 3, T = 5, N = 9, 10^5$ Monte-Carlo Trials)

Figure 3.7 shows the same plots with large sample size. In this figure, the approximation is accurate.

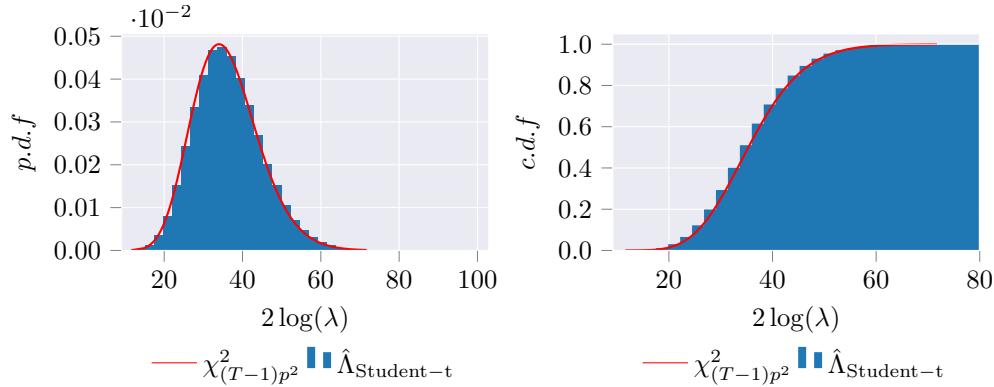


Figure 3.7: Distribution of statistic under large sample size ($p = 3, T = 5, N = 100, 10^5$ Monte-Carlo Trials)

For the compound-Gaussian derived statistics including scale parameters, the Wilks theorem does not apply since the dimension of the parameter space grows with the number of samples. Obtaining the distribution of the tests is no longer evident in this case. We will thus first show that these statistics have the CFAR property.

Study of the CFAR property in deterministic compound-Gaussian case

We consider the CFAR property which is primordial if we want to apply the statistic in a decision scheme where the significance level is important. We have the following propositions:

Proposition 3.2.10. $\hat{\Lambda}_{\text{MT}}$ is CFAR texture and matrix for Problem (3.9).

Proof. See Appendix 3.A.9. \square

Proposition 3.2.11. $\hat{\Lambda}_{\text{Mat}}$ is CFAR texture and matrix for Problem (3.10).

Proof. The same arguments as used in Proposition 3.2.10 are applied here. \square

Proposition 3.2.12. $\hat{\Lambda}_{\text{Tex}}$ is CFAR texture but is not CFAR matrix for Problem (3.11).

Proof. See Appendix 3.A.10. \square

The CFAR texture and matrix behavior of the new statistics have been tested in simulation using the same methodology as in 3.2.4. To this end, a time series has been generated under the H_0 regime of problem (3.9) which also corresponds to H_0 for the other problems. The statistics have been computed in MC trials to generate the plots shown at Figure 3.8. The Gaussian-derived statistics have also been computed. The plots show that these Gaussian statistics vary when the texture changes and thus, have not the texture CFAR property. In contrast, the new statistics, do not vary for any texture parameter tested, which is an improvement. In this regards, $\hat{\Lambda}^{\text{Mat}}$ is the most robust one since the statistic does not vary even if the texture equality between the dates is not respected.

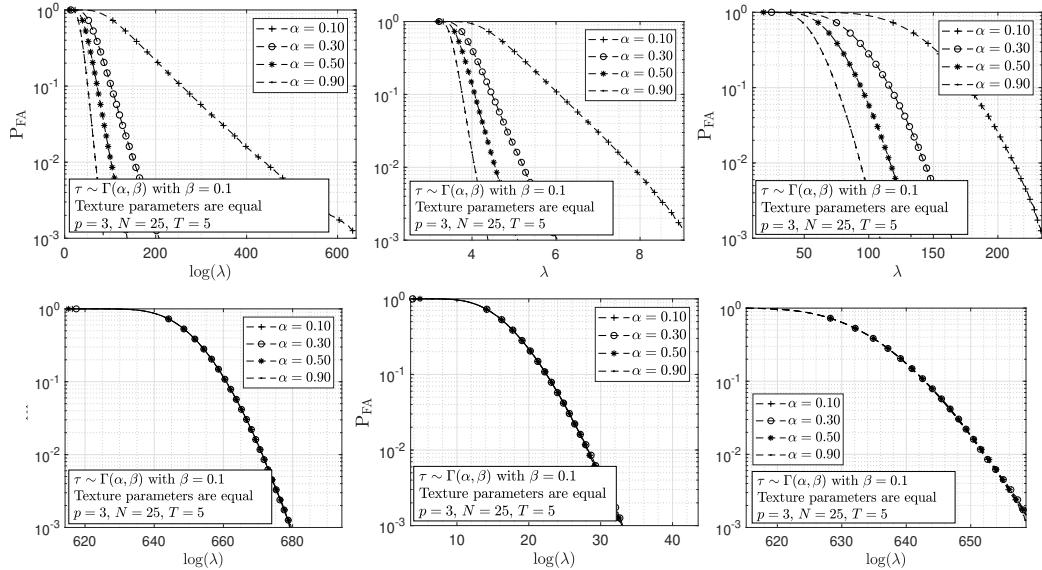


Figure 3.8: Texture CFAR behaviour. Top-left: $\hat{\Lambda}_G$. Top-center: $\hat{\Lambda}_{t_1}$. Top-right: $\hat{\Lambda}_{\text{Wald}}$. Bottom-left: $\hat{\Lambda}_{\text{MT}}$. Bottom-center: $\hat{\Lambda}_{\text{Mat}}$. Bottom-Right: $\hat{\Lambda}_{\text{Tex}}$. $\rho = 0.3$ at each date for all the curves.

Next, the matrix CFAR behaviour is tested using $\forall k, \forall t, \tau_k^t = 1$. Figure 3.9 presents plots of Monte Carlo trials where the coefficients for the covariance matrix vary. The plots

show that the Gaussian statistics are CFAR which was demonstrated in [Ciuonzo et al., 2017]. It shows that $\hat{\Lambda}_{\text{MT}}$, $\hat{\Lambda}_{\text{Mat}}$ have the CFAR matrix behaviour while $\hat{\Lambda}_{\text{Tex}}$ has not. This result is coherent with the theoretical analysis.

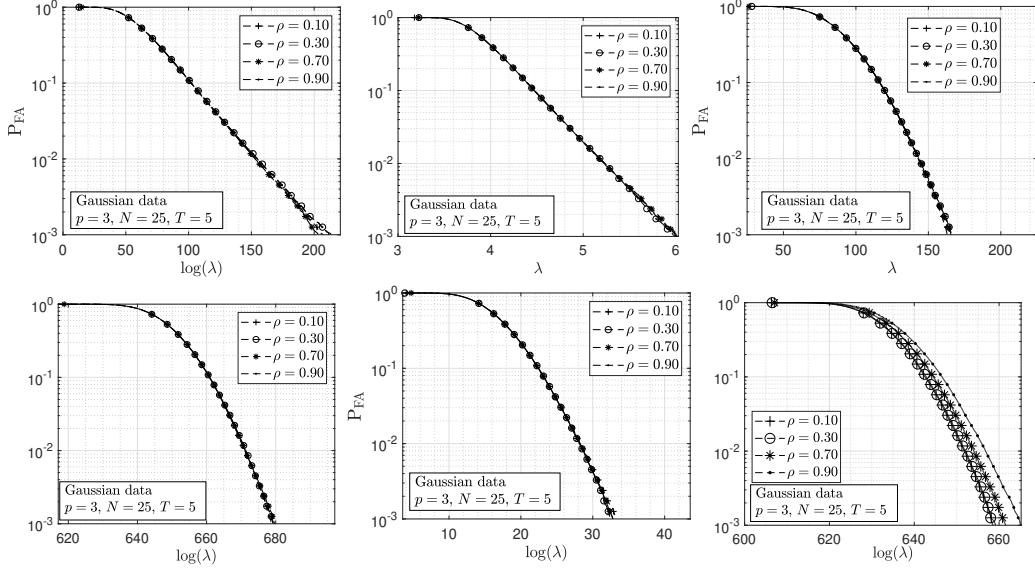


Figure 3.9: Matrix CFAR behaviour. Top-left: $\hat{\Lambda}_G$. Top-center: $\hat{\Lambda}_{t_1}$. Top-right: $\hat{\Lambda}_{\text{Wald}}$. Bottom-left: $\hat{\Lambda}_{\text{MT}}$. Bottom-center: $\hat{\Lambda}_{\text{Mat}}$. Bottom-Right: $\hat{\Lambda}_{\text{Tex}}$.

Bias and Consistency of new estimates

As a first step in analysing the distribution of $\hat{\Lambda}_{\text{MT}}$ and $\hat{\Lambda}_{\text{Tex}}$, we consider analysing the statistical properties of bias and consistency of the estimates $\hat{\xi}_0^{\text{MT}}$ and $\hat{\xi}_t^{\text{Tex}}$. Indeed, well known methods, such as delta method [Oehlert, 1992], rely on those assumptions in order to obtain results about asymptotic distributions of estimates which in turn can be used in the analysis of the asymptotic distribution of the statistics. The following propositions can be obtained:

Proposition 3.2.13. $\hat{\xi}_0^{\text{MT}}$ and $\hat{\xi}_t^{\text{Tex}}$ are unbiased estimators of ξ_0 , the true covariance under null hypothesis.

Proof. See Appendix 3.A.11 □

Proposition 3.2.14. $\hat{\xi}_0^{\text{MT}}$ and $\hat{\xi}_t^{\text{Tex}}$ are consistent estimators of ξ_0 , the true covariance under null hypothesis.

Proof. See Appendix 3.A.12 □

We consider the consistency properties in simulation as well. We generated a time series under deterministic compound-Gaussian model using the same methodology as in 3.2.4, where we ensured that $\forall t \in \llbracket 1, T \rrbracket$, $\tau_k^t = \tau_k^0$. For $\hat{\xi}_0^{\text{MT}}$ and $\hat{\xi}_0^{\text{Mat}}$, a single covariance matrix Σ_0 has been used for all the dates, while for $\hat{\xi}_t^{\text{Tex}}$, each were different. Figure 3.10 presents a

Monte Carlo simulation where the Mean Square Error (MSE) of estimation is plotted against the number of observations. The plots show the decrease at a constant rate of the MSE for all the different estimates and confirm the consistency of those estimates in practice. We also notice that in this situation $\hat{\xi}_0^{\text{MT}}$ is a better estimate than $\hat{\xi}_0^{\text{Mat}}$. This is not surprising since $\hat{\xi}_0^{\text{MT}}$ takes into account the equality of texture parameters between the dates.

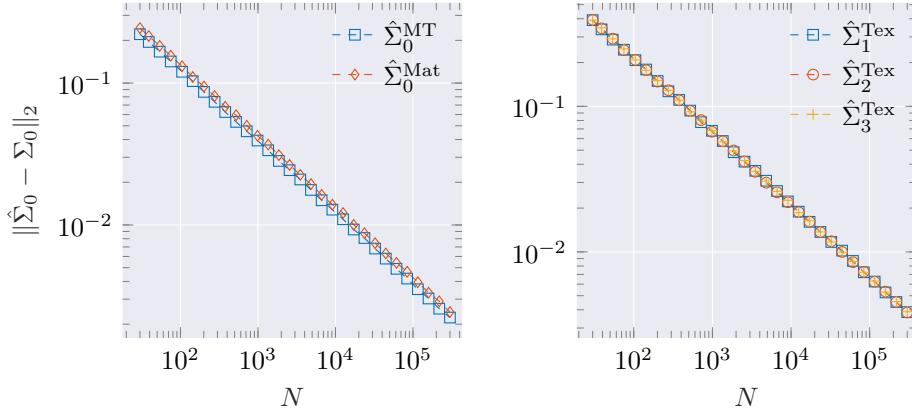


Figure 3.10: Asymptotic consistency of estimates. Left: $p = 3$, $T = 3$, 1000 Trials, $\rho_0 = 0.70$, $\alpha = 0.3$, $\beta = 0.1$, same textures for each date. Right: $p = 3$, $T = 3$, 1000 Trials, $\rho_1 = 0.29$, $\rho_2 = 0.56$, $\rho_3 = 0.58$, $\alpha = 0.3$, $\beta = 0.1$, same textures for each date.

These results are interesting since they can be used in an analysis of the null distributions for the compound-Gaussian derived statistics. This work has not yet been considered due to time constraints but will be treated as an extension at a later time.

3.2.6 Performance study of new statistics

Elliptical case

In this section, in order to compare the performance of detection, we choose to compute the probability of detection P_D , through means of M-C trials, on a simple situation where there is only one change at date t_C . Before the change, a covariance Σ_0 is chosen and after the change a covariance Σ_1 is used. We use covariance of the form $(\Sigma_{\epsilon=0,1})_{i,j} = \tau_\epsilon \rho_\epsilon^{|i-j|}$.

The Bartlett distance [Frery et al., 2014] on covariance matrices is used as a measure of the amplitude of the change:

$$d_B(\Sigma_1, \Sigma_2) = \log \left(\frac{|\Sigma_1 + \Sigma_2|^2}{|\Sigma_1| |\Sigma_2|} \right) - 2p \log(2). \quad (3.18)$$

This quantity will be considered as an equivalent Signal to Noise Ratio (SNR) for the covariance equality testing problem.

We compare the following statistics: $\hat{\Lambda}_G$, $\hat{\Lambda}_{t_1}$, The Student-t GLRT (obtained using $g(t) = (1 + 2t/\nu)^{-(2p+\nu)/2}$) supposing $\nu = 2$, $\hat{\Lambda}_{\text{MT}}$, $\hat{\Lambda}_{\mathcal{CAE}}$, $\hat{\Lambda}_{\text{Tex}}$. The Wald statistic proved to be numerically unstable when used on heavy tailed distributions and was thus omitted for this simulation. The thresholds for guaranteeing the P_{FA} are computed numerically through 10^4 Monte-Carlo simulations.

We consider three separate setups:

- Gaussian case with either a change in scale or shape:

The results plotted in Figure 3.11, show that for a change in shape matrix, the Gaussian statistic yields the best detection performance while the other robust statistics have slightly worse performance. This result is expected since usually robustness comes with a slight trade-off in performance in Gaussian case. $\hat{\Lambda}_{\text{Tex}}$ doesn't yield any detection since it is not sensitive to a change in shape.

For a change in scale, it is interesting to see that both $\hat{\Lambda}_{t_1}$ and $\hat{\Lambda}_{\text{Tex}}$ slightly outperforms $\hat{\Lambda}_G$. There is no theoretical explanation about that fact. The Student-t detector yields poorer performance since for $\nu = 2$, it expects the distribution of the data to be heavy-tailed while it is not. Finally, $\hat{\Lambda}_{\text{CAE}}$, which is not sensitive to a change in scale, does not yield a detection.

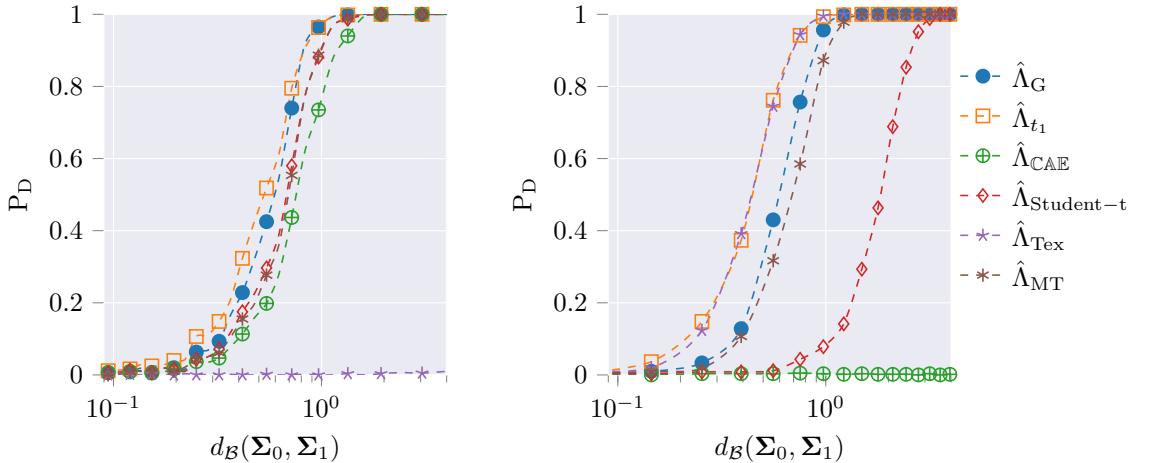


Figure 3.11: P_D as a function of $d_B(\Sigma_0, \Sigma_1)$ for a Gaussian case. Left: Change in Shape only. Right: Change in scale only. Parameters: $p = 5$, $N = 15$, $T = 10$, $P_{FA} = 10^{-3}$. Curves are computed with 600 Monte Carlo trials.

- Student-t case with either a change in scale or shape:

The results, plotted in Figure 3.11, shows that for a change in shape matrix, $\hat{\Lambda}_{\text{CAE}}$ and $\hat{\Lambda}_{\text{Student}-t}$ have much better detection compared to others. This is expected since their respective model assumption is respected while for other statistics, it is not resulting in much higher thresholds of detection.

For a change in scale, the best detector is $\hat{\Lambda}_{\text{Student}-t}$ while $\hat{\Lambda}_{\text{CAE}}$, which is not sensitive to scale, doesn't yield a detection. For other scale sensitive detectors, the same conclusions as for shape explain their relative lower detection rate. In this case, the Gaussian GLRT outperforms the t_1 statistic. This might indicate better robustness of this detector with regards to non-Gaussianity.

- K-distribution with either a change in scale or shape:

In this case, the performance of $\hat{\Lambda}_{\text{CAE}}$ with regards to shape doesn't vary while has a comparatively lower detection rate. This is expected since $\hat{\Lambda}_{\text{CAE}}$ is adapted to all

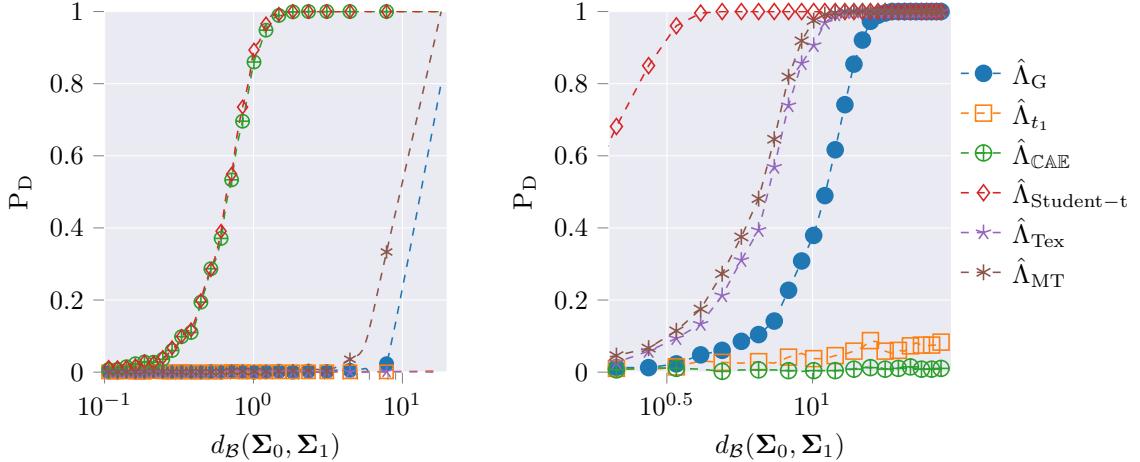


Figure 3.12: P_D as a function of $d_B(\Sigma_0, \Sigma_1)$ for a Student-t with degrees of freedom $\nu = 2$. Left: Change in Shape only. Right: Change in scale only. Parameters: $p = 5$, $N = 15$, $T = 10$, $P_{FA} = 10^{-3}$. Curves are computed with 600 Monte Carlo trials.

elliptical distributions making it more robust. Again The Gaussian GLRT appears to be better than t_1 statistic, which is coherent with results in Student-t case.

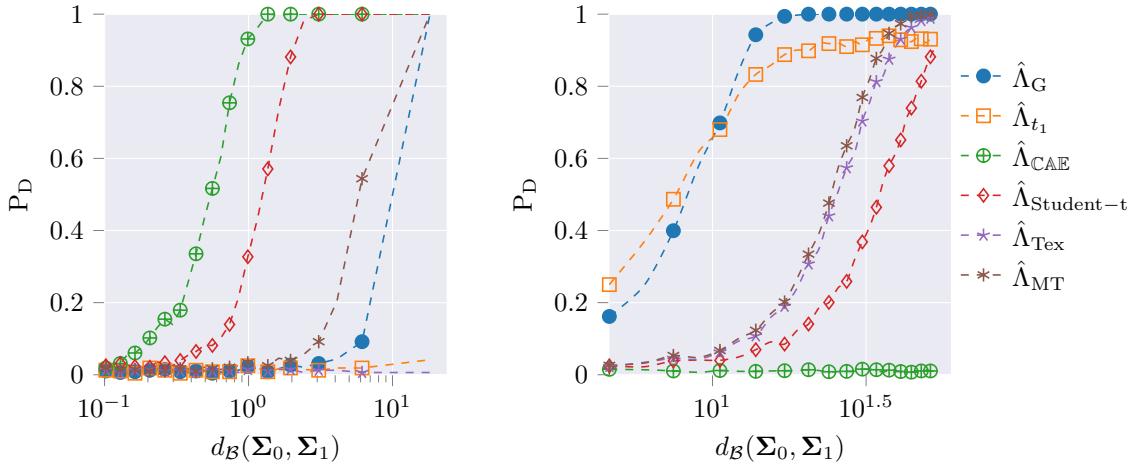


Figure 3.13: P_D as a function of $d_B(\Sigma_0, \Sigma_1)$ for a K-distributed data with shape parameter $\nu = 0.1$. Left: Change in Shape only. Right: Change in scale only. Parameters: $p = 5$, $N = 15$, $T = 10$, $P_{FA} = 10^{-3}$. Curves are computed with 600 Monte Carlo trials.

Concerning scale, it appears that Gaussian statistics outperforms the other statistics. There is no real insight to this result since for all detectors, their model assumption is not respected.

Deterministic compound-Gaussian case

For this case, the same methodology as section 3.2.4 about convergence is used: we consider realisations of random variables $\mathbf{x} = \sqrt{\tau} \mathbf{z}$ where τ follows a Γ -distribution with shape parameter α and scale parameter β . \mathbf{z} is generated through a Gaussian realization with covariance matrix chosen to be Toeplitz of the form:

$$\boldsymbol{\xi} = (\sigma_{m,n})_{\substack{1 \leq m \leq p \\ 1 \leq n \leq p}},$$

where : $\sigma_{m,n} = \rho^{|m-n|}.$

We consider analyzing the theoretical performance of the compound-Gaussian derived statistics with regards to Gaussian ones. The statistic derived under elliptical model is omitted here since from a practical point of view, the performance vary greatly in case of mismatch as seen in previous simulation results. We consider a time series with $T = 10$, $p = 3$, $N = 7$ with a change at $t = 5$ and plot Receiver Operating Characteristic (ROC) curves for each problem since the SNR is difficult to define in this case.

- Problem (3.9): Before change, the covariance matrix is associated with $\rho = 0.1$. The textures are generated with $\alpha = 0.3, \beta = 0.1$ and are equal for each date. After the change, we have $\rho = 0.8$ and $\alpha = 0.3, \beta = 0.3$ and the textures are equal for each date.
- Problem (3.10): Before change, the covariance matrix is associated with $\rho = 0.1$. After the change, we have $\rho = 0.8$. For any date, the textures are generated using $\alpha = 0.3, \beta = 0.1$ and are different for each date.
- Problem (3.11): Before and after the change, the covariance matrix is obtained using a random value for ρ and the textures are generated using $\alpha = 0.3, \beta = 0.3$. Before change, the textures are equal and after the change, they are different.
- Gaussian problem: the textures are all fixed to one. Before the change, the covariance matrix is associated with $\rho = 0.1$. After the change, the covariance matrix is associated with $\rho = 0.8$.

Figure 3.14 gives the results obtained by MC trials. The thresholds for a given P_{Fa} are computed numerically using the H_0 regime of the problem considered. Although not realistic on real images, this allows, on synthetic data, to have an experimental threshold that matches the objective P_{Fa} even if the test is not CFAR for the problem considered.

For each problem, the derived statistic yields the best result. The Gaussian statistics have poorer performance than $\hat{\Lambda}_{MT}$ and $\hat{\Lambda}_{Tex}$ for testing a change in the texture. $\hat{\Lambda}_{Mat}$ performs the best when there is only a change in the covariance matrix shape. For the third problem, since there is no change in the matrices, it is not surprising that the detection rate is low. $\hat{\Lambda}_{MT}$ appears to be the best option for testing changes on the textures since for both problems (3.9) and (3.11), the performance are good. This is explained by the fact that the distribution under H_0 of the statistic is less sensitive to a violation of the matrix equality assumption than it would be from a texture one. Since in Problem (3.11) the textures are equal before the change, the threshold to guarantee the P_{Fa} is still low enough to guarantee good performance. Finally, when the data is strictly Gaussian, $\hat{\Lambda}_{t_1}$ and $\hat{\Lambda}_G$ have better results than the robust statistics. This result is expected since there is a trade-off between robustness and performance when considering robust methods. Among the new statistics, $\hat{\Lambda}_{Tex}$ does not allow to detect a change in the shape so its results are expected to be lowest.

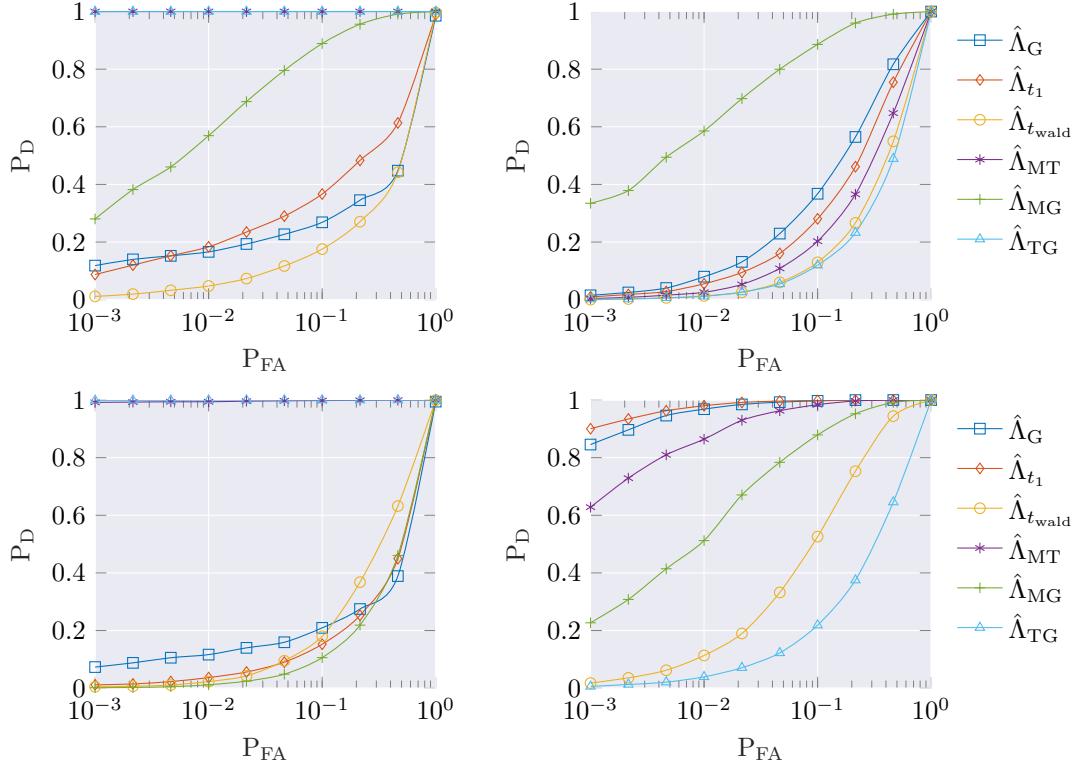


Figure 3.14: ROC curves obtained on synthetic data (10^4 Monte-Carlo trials). Top-left: Problem (3.9). Top-right: Problem (3.10). Bottom-left: Problem (3.11). Bottom-right: Gaussian setting.

3.3 APPLICATION OF ROBUST STATISTICS TO SAR CHANGE DETECTION

In this section, we will only consider the performance of the compound-Gaussian-derived statistics and compare them to the classic Gaussian ones. Indeed, since for elliptical-derived statistics need the knowledge of the density generator function g , they are impractical for real datasets.

3.3.1 Data description

The proposed statistics have been tested on real images coming from two different datasets: SDMS (Courtesy AFRL/RYA) [Scarborough et al., 2010] and UAVSAR (Courtesy NASA/JPL-Caltech). From SDMS, three images of the same scene, presented in Figure 3.16, are used. The ground truth is obtained from [Carotenuto et al., 2016] for the two dates and [Ciuonzo et al., 2017] for the three dates. From UAVSAR, two scenes with two images each are used. They are presented in Figure 3.15. The ground truth is collected from [Ratha et al., 2017, Nascimento et al., 2019]. Table 3.1 gives an overall perspective of the scenes used in the study.

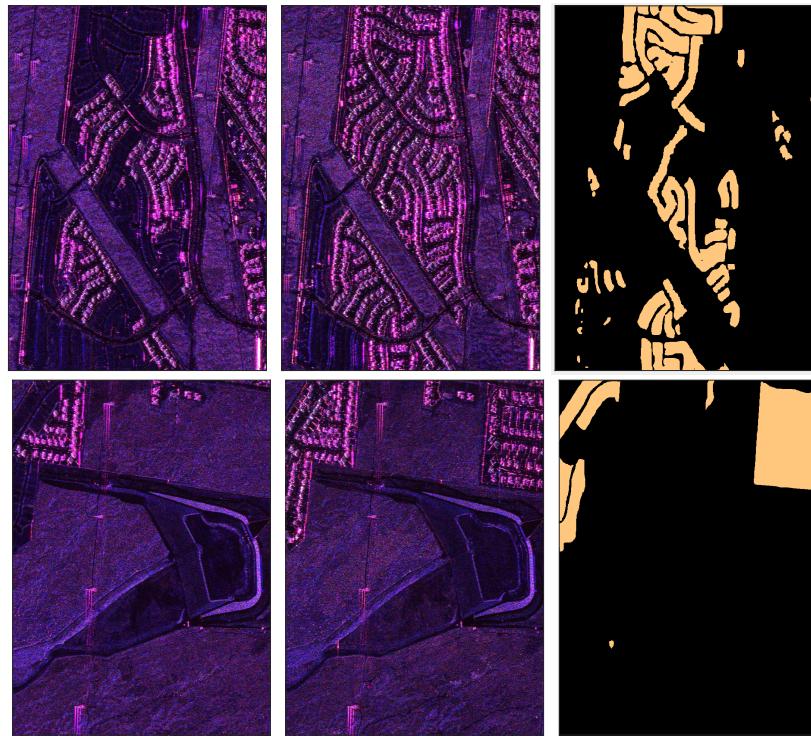


Figure 3.15: UAVSAR Dataset in Pauli representation. Left: April 23, 2009. Middle: May 15, 2011. Right: Ground Truth. Top: Scene 1. Bottom: Scene 2.

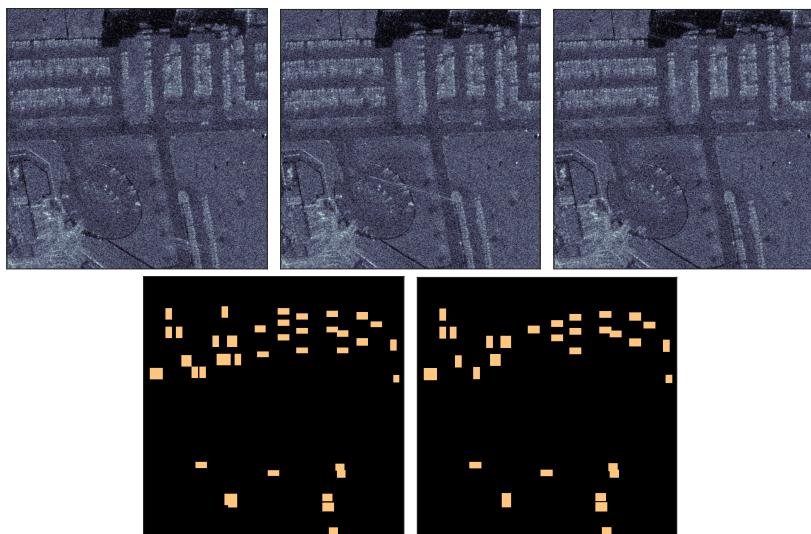


Figure 3.16: SDMS Dataset. Top-left: FP0120. Top-middle: FP0121 . Top-right: FP0124. Bottom-left: Ground Truth FP0120-FP0121-FP0124. Bottom-right: Ground Truth FP0121-FP0124.

Table 3.1: Description of SAR data used

Dataset	SDMS CCD challenge	UAVSAR SanAnd 26524 03 Segment 4
URL	https://www.sdms.afrl.af.mil/	https://uavstar.jpl.nasa.gov
Tim interval	Within a day	April 23, 2009 - May 11, 2015
Resolution	0.2m by 0.2m	1.67m by 0.6m
Scenes	FP0121 and FP0120-FP0.121-FP0124	Scene 1 and Scene 2
p	3	3
T	3	2
Size	1000px × 1000px	2300px × 600px

3.3.2 First analysis

We first try the various Gaussian and new statistics on the three dates of SDMS and UAVSAR dataset. Figure 3.17 presents the results relative to the statistics for SDMS data and Figure 3.18 for UAVSAR. The values for $\hat{\Lambda}_{t_1}$ and $\hat{\Lambda}_{\text{Wald}}$ are omitted since they have similar behaviour than $\hat{\Lambda}_G$. Qualitatively, each statistic is high at the location of the changes given by the ground truth. For both dataset, $\hat{\Lambda}_{\text{Mat}}$ seems to have poorer performance since the values of the statistic are not much higher on the changes compared to the background. For UAVSAR data scene 1, a linear pattern appears in the bottom-right corner and responds highly for all detectors except for $\hat{\Lambda}_{\text{MT}}$ and $\hat{\Lambda}_{\text{Tex}}$. However, it is difficult to conclude solely on those qualitative terms.

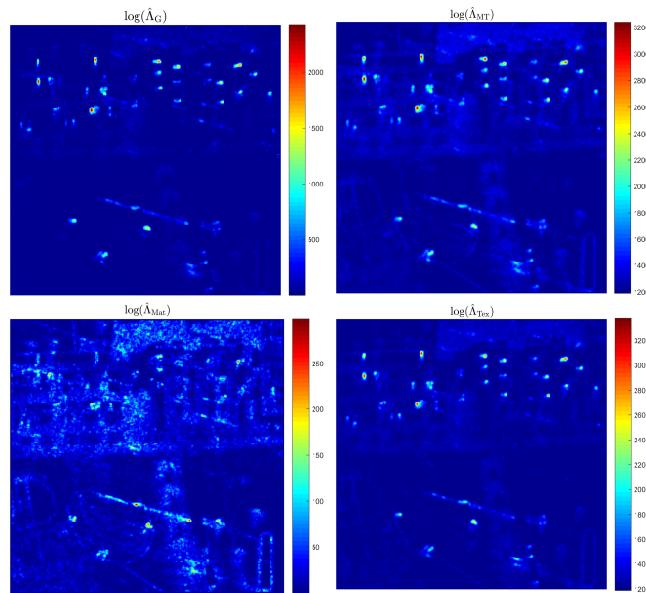


Figure 3.17: Value of the different statistics for SDMS FP0120-FP0121-FP0124. $p = 3$, $N_1 = N_2 = 11$.

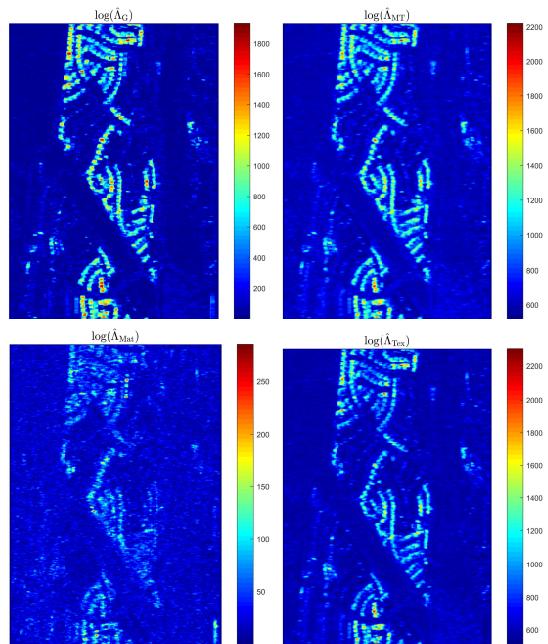


Figure 3.18: Value of the different statistics for UAVSAR Scene 1. $p = 3$, $N_1 = N_2 = 11$.

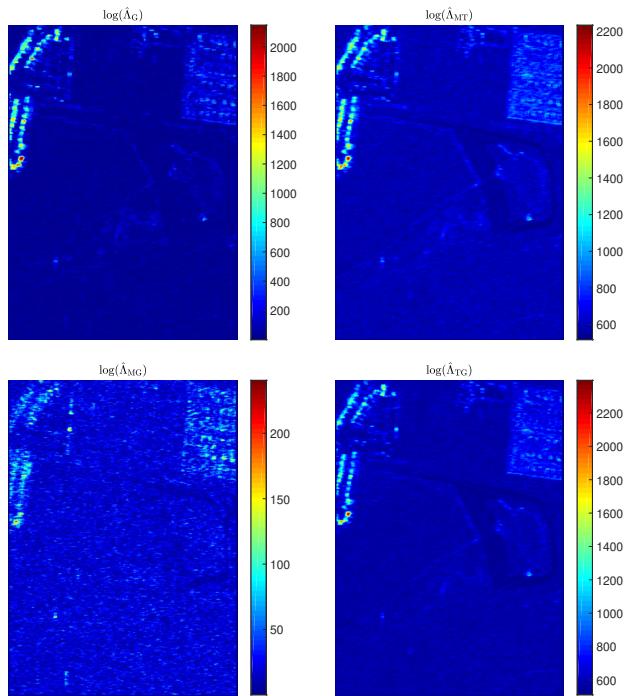


Figure 3.19: Value of the different statistics for UAVSAR Scene 2. $p = 3$, $N_1 = N_2 = 11$.

ROC plots

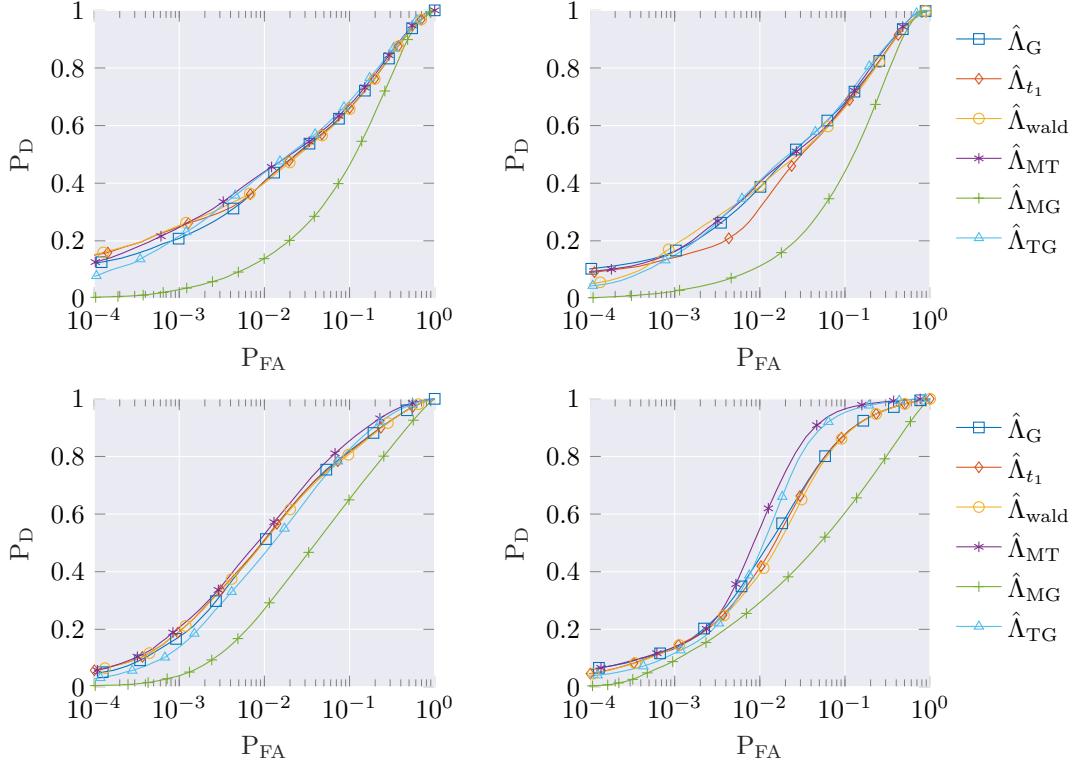


Figure 3.20: P_D versus P_{FA} on real data. Top-left: SDMS FP0121-FP0124. Top-right: FP0120-FP0121-FP0124. Bottom-left: UAVSAR Scene1. Bottom-right: UAVSAR Scene 2. For all images, $p = 3$, $N_1 = N_2 = 11$.

To quantify the performance of the statistics, experimental ROC curves are plotted using the Ground truth, denoted $\mathbb{1}_{\text{Gt}}(x, y)$ associated with spatial coordinates (x, y) , by computing the following:

- Probability of false alarm:

$$P_{FA} = N_{FD}/N_{NC},$$

where: $N_{FD} = \sum_{x,y} (\hat{\Lambda}(x, y) \geq \lambda) \times (1 - \mathbb{1}_{\text{Gt}}(x, y))$,

$$N_{NC} = \sum_{x,y} (1 - \mathbb{1}_{\text{Gt}}(x, y)).$$

- Probability of detection:

$$P_D = N_{GD}/N_C,$$

where: $N_{GD} = \sum_{x,y} (\hat{\Lambda}(x, y) \geq \lambda) \times \mathbb{1}_{\text{Gt}}(x, y)$,

$$N_C = \sum_{x,y} \mathbb{1}_{\text{Gt}}(x, y).$$

Figure 3.20 shows the results for each dataset and a size of analysis window of $N_1 = N_2 = 11$. It appears that $\hat{\Lambda}_{MT}$ has the overall best performance: it has similar results to $\hat{\Lambda}_G$ on SDMS dataset but performs better on UAVSAR dataset. Each Gaussian-derived statistic has similar performance but $\hat{\Lambda}_G$ appears to have better results than $\hat{\Lambda}_{t_1}$ and $\hat{\Lambda}_{Wald}$.

These results can be interpreted as follows: on the SDMS dataset, while the resolution is high, the images are globally homogeneous. In fact, much of the details are not visible and the objects appear to be blurry. This means that in practice, on a small local neighborhood, the Gaussian model is accurate and thus that the Gaussian-derived statistics perform well. Nonetheless, the new statistics, except $\hat{\Lambda}_{Mat}$, do not have lower performance and can still have better performance when the size of the neighborhood chosen is high as will be shown afterward. On the other hand, the objects are better resolved on the UAVSAR. The transitions are sharper which means that a heterogeneous model is more accurate and thus that the new statistics will perform better. The difference of performance for UAVSAR scene 2 can be explained by the fact that dynamic between the darker zones and the bright ones is much higher than in scene 1.

For the datasets used here, $\hat{\Lambda}_{Mat}$ does not perform well. This is due to the fact that the detection omits the texture parameters which are responsible for the power. In these datasets, the ground truth corresponds to the arrival or disappearance of strong scatterers and thus, the power has an important role. As explained before, $\hat{\Lambda}_{Mat}$ allows detecting changes which are focused on the correlation structures and is not appropriate for those kinds of change.

Increasing the size of window

In order to test the impact of the size of the analysis window, we fix an experimental $P_{FA} = 10^{-2}$ and plot the P_D against the size of the window. Figure 3.21 gives the results for all datasets.

By increasing the size of the window, the detection rate improves. It can be explained by the fact that the estimation step has been performed on more data and is thus more precise. The drawback is that the detection is obtained with a lower spatial resolution.

When increasing the size of the window, $\hat{\Lambda}_{MT}$ and $\hat{\Lambda}_{Tex}$ perform better than the Gaussian statistics, especially on UAVSAR Scene 2. This is expected, since increasing the size of the window means that the data are spread over a large spatial leading to an increase of the heterogeneity due to the presence of many scatterers in the scene.

Increasing the dimension of pixels

Finally, we consider the performance if the size of vector p increases. To this end, we exploit the wavelet decomposition method presented in the chapter 2. Using this decomposition on all polarimetric canals of SDMS dataset allows having an image with $p = 27$. The decomposition is not performed on UAVSAR dataset since it does not exhibit a physical diversity using the wavelet decomposition. Figure 3.22 gives the result of the CD for all the statistics. When compared to the performance using solely polarimetric information, it appears that using this method, the performance is lower when the P_{FA} is very low, while they are improved for $P_{FA} > 10^{-1}$. The case of $\hat{\Lambda}_{Tex}$, which has significantly better performance, highlights again that the texture parameter plays a significant role in CD applications.

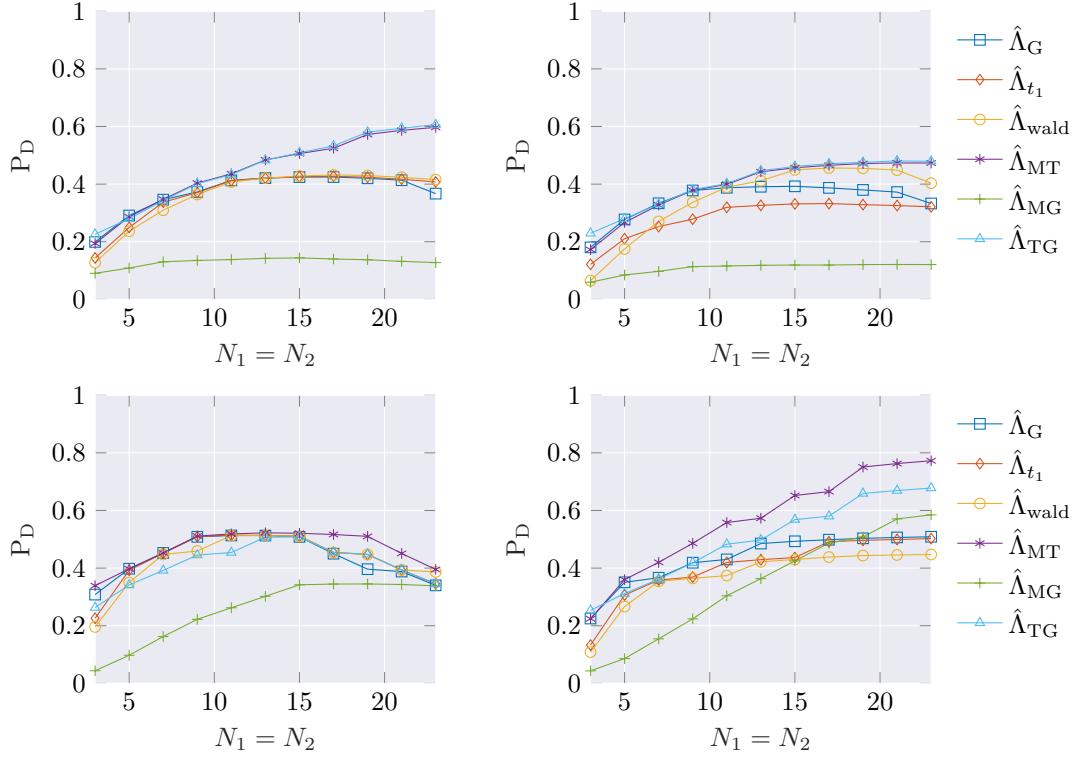


Figure 3.21: P_D as a function of window size at $P_{FA} = 10^{-2}$. Top-left: SDMS FP0121-FP0124. Top-right: FP0120-FP0121-FP0124. Bottom-left: UAVSAR Scene1. Bottom-right: UAVSAR Scene2. For all images, $p = 3$.

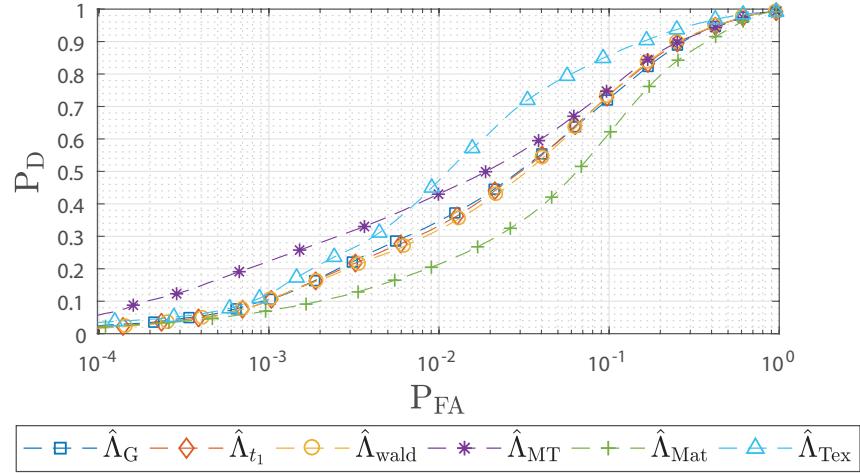


Figure 3.22: P_D versus P_{FA} on SDMS FP0121-FP0124 with wavelet decomposition. $N_1 = N_2 = 11$.

3.4 CONCLUSIONS

We have derived new statistics under both elliptical and compound-Gaussian models in order to bring robustness to the covariance equality problem. We have both studied the usability of those statistics (convergence properties of estimates) as well as the statistical properties (namely distribution under the null hypothesis for the elliptical case). Performance has also been tested through simulations and real datasets for which improvement has been shown with regards to Gaussian-derived statistics.

The model (3.10) was the first one considered in our work since it allows to test the shape matrix without being influenced by the scale. This has proved, however, to be insufficient with regards to change detection since it discards the relative power between the images which accounts for the changes. However, when the images are not well calibrated or when they are acquired at different elevation angles, there can be an arbitrary change of this power between the acquisitions which is not significative. In that case, this approach is still interesting.

When we want to take into account the scale as well, it lead us to the model (3.9) which test both scale and shape matrix. It is the equivalent of the Gaussian one in this robust context. In practice this one has been shown to work the best.

The model (3.11) is only sensitive to the scale and discard the shape matrix information which makes it less preformant than the (3.9) so there is no real interest to this approach.

The exact or asymptotic distribution under null for compound-Gaussian statistic has not yet been achieved but the consistency properties of estimates are a first step which can be exploited to that end. Indeed, since asymptotically the estimates are consistent, it would be possible to obtain an asymptotic approximation by replacing the estimates by the true matrix and analyze the distribution of quadratic terms.

3.A APPENDIX

3.A.1 Proof of proposition 3.2.1 at p. 86

Proposition. *The GLRT under problem (3.3) is the following quantity:*

$$\hat{\Lambda}_{\mathbb{C}\mathcal{E}^g} = \frac{\left| \hat{\Sigma}_0^M \right|^{TN}}{\prod_{t=1}^T \left| \hat{\Sigma}_t^M \right|^N} \prod_{k=1}^{k=N} \frac{g(q(\hat{\Sigma}_t^M, \mathbf{x}_k^t))}{g(q(\hat{\Sigma}_0^M, \mathbf{x}_k^t))} \begin{array}{c} \text{H}_1 \\ \lessgtr \\ \text{H}_0 \end{array} \lambda, \quad (3.19)$$

where:

$$\dot{\Sigma}_t^M = f_t(\hat{\Sigma}_t^M), \quad \dot{\Sigma}_0^M = \frac{1}{T} \sum_{t=1}^T f_t(\hat{\Sigma}_0^M) \quad \text{and}$$

$$f_t(\Sigma) = \frac{1}{N} \sum_{k=1}^N \frac{-g'(q(\Sigma, \mathbf{x}_k^t))}{g(q(\Sigma, \mathbf{x}_k^t))} \mathbf{x}_k^t \mathbf{x}_k^t \quad (3.20)$$

Proof. The GLRT in this problem reads:

$$\hat{\Lambda} = \frac{\max_{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}; \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T)}{\max_{\boldsymbol{\theta}_0} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}; \boldsymbol{\theta}_0)} \quad (3.21)$$

where $\boldsymbol{\theta}_0 = \{\Sigma_0\}$ and $\forall t \in \llbracket 1, T \rrbracket$, $\boldsymbol{\theta}_t = \{\Sigma_t\}$.

Given that the observations are independent, we can write:

$$\frac{\max_{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T} \prod_{t=1}^T \prod_{k=1}^{k=N} p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{E}}(\mathbf{x}_k^t; \boldsymbol{\theta}_t, g)}{\max_{\boldsymbol{\theta}_0} \prod_{t=1}^T \prod_{k=1}^{k=N} p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{E}}(\mathbf{x}_k^t; \boldsymbol{\theta}_0, g)}. \quad (3.22)$$

This expression can be computed by optimizing the numerator and denominator separately. Then, the idea is to estimate each unknown parameter separately and plugging back the estimates. Thus we can compute:

$$\hat{\Lambda} = \frac{\mathcal{L}_1(\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_T)}{\mathcal{L}_0(\hat{\boldsymbol{\theta}}_0)}, \quad (3.23)$$

where

$$\begin{aligned}\mathcal{L}_1(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T) &= \prod_{\substack{k=1 \\ t=1}}^{k=N \\ t=T} p_{\mathbf{x}_k^t}^{\mathbb{CE}}(\mathbf{x}_k^t; \boldsymbol{\theta}_t, g), \\ \mathcal{L}_0(\boldsymbol{\theta}_0) &= \prod_{\substack{k=1 \\ t=1}}^{k=N \\ t=T} p_{\mathbf{x}_k^t}^{\mathbb{CE}}(\mathbf{x}_k^t; \boldsymbol{\theta}_0, g), \\ \hat{\boldsymbol{\theta}}_0 &= \operatorname{argmax}_{\boldsymbol{\theta}_0} \mathcal{L}_0(\boldsymbol{\theta}_0), \\ \forall t \in \llbracket 1, T \rrbracket, \hat{\boldsymbol{\theta}}_t &= \operatorname{argmax}_{\boldsymbol{\theta}_t} \mathcal{L}_1(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T).\end{aligned}$$

The optimizations can be done by obtaining the stationary point of the log-likelihood. Indeed, it is shown in [Ollila and Tyler, 2014] that the negative of the log-likelihood is g-covex towards the scatter matrix given that $\rho = -\log g$ is differentialbe, $u = -g'/g$ is a function having values on \mathbb{R}^+ and $\phi(t) = tu(t)$ is non-decreasing.

- Considering the numerator, we have:

$$\log \mathcal{L}_1 \approx -N \sum_{k=1}^N \log |\boldsymbol{\Sigma}_t| + \sum_{k=1}^N \rho(q(\boldsymbol{\Sigma}_t, \mathbf{x}_k^t))$$

Using chain-rule and recalling complex differentiation results [Gini and Greco, 2002]:

$$\begin{aligned}\frac{\partial \log |\boldsymbol{\xi}|}{\partial \boldsymbol{\xi}} &= \boldsymbol{\xi}^{-1}, \\ \frac{\partial q(\boldsymbol{\xi}, \mathbf{x}_k^t)}{\partial \boldsymbol{\xi}} &= -\mathbf{S}_k^t \boldsymbol{\xi}^{-2},\end{aligned}\tag{3.24}$$

we have:

$$\begin{aligned}\frac{\partial \log \mathcal{L}_1}{\partial \boldsymbol{\Sigma}_t} &= -N \boldsymbol{\Sigma}_t^{-1} + \sum_{k=1}^N \frac{1}{g(q(\boldsymbol{\Sigma}_t, \mathbf{x}_k^t))} \left(\frac{dg}{d\alpha} \right)_{\alpha=q(\boldsymbol{\xi}, \mathbf{x}_k^t)} \left(\frac{\partial q(\boldsymbol{\Sigma}_t, \mathbf{x}_k^t)}{\partial \boldsymbol{\Sigma}_t} \right) \\ &= -N \boldsymbol{\Sigma}_t^{-1} + \sum_{k=1}^N -\frac{g'(q(\boldsymbol{\Sigma}_t, \mathbf{x}_k^t))}{g(q(\boldsymbol{\Sigma}_t, \mathbf{x}_k^t))} \mathbf{S}_k^t\end{aligned}$$

Solving $\frac{\partial \log \mathcal{L}_1}{\partial \boldsymbol{\Sigma}_t} = \mathbf{0}_{p^2}$ yields:

$$\hat{\boldsymbol{\Sigma}}_t = \frac{1}{N} \sum_{k=1}^N -\frac{g'(q(\hat{\boldsymbol{\Sigma}}_t, \mathbf{x}_k^t))}{g(q(\boldsymbol{\Sigma}_0, \mathbf{x}_k^t))} \mathbf{S}_k^t\tag{3.25}$$

- Considering the denominator, we have:

$$\log \mathcal{L}_0 \approx -TN \sum_{t=1}^T \sum_{k=1}^N \log |\boldsymbol{\Sigma}_0| + \sum_{t=1}^T \sum_{k=1}^N \rho(q(\boldsymbol{\Sigma}_0, \mathbf{x}_k^t))$$

Using the same optimization procedure as for the numerator yields:

$$\hat{\Sigma}_0 = \frac{1}{TN} \sum_{t=1}^T \sum_{k=1}^N -\frac{g'(q(\hat{\Sigma}_0, \mathbf{x}_k^t))}{g(q(\hat{\Sigma}_0, \mathbf{x}_k^t))} \mathbf{S}_k^t \quad (3.26)$$

Finally, plugging the estimates in the expression of the GLRT at eq. (3.A.1) yields:

$$\hat{\Lambda}_{\mathbb{C}\mathcal{E}^g} = \frac{\left| \hat{\Sigma}_0^M \right|^{TN}}{\prod_{t=1}^T \left| \hat{\Sigma}_t^M \right|^N} \prod_{k=1}^N \frac{g\left(q(\hat{\Sigma}_t^M, \mathbf{x}_k^t)\right)}{g\left(q(\hat{\Sigma}_0^M, \mathbf{x}_k^t)\right)} \begin{array}{c} \text{H}_1 \\ \gtrless \\ \text{H}_0 \end{array}$$

□

3.A.2 Proof of proposition 3.2.2 at p. 88

Proposition. *The GLRT under problem (3.6) is the following quantity:*

$$\hat{\Lambda}_{\mathbb{C}\mathcal{A}\mathcal{E}} = \frac{\left| \hat{\xi}_0^M \right|^{TN}}{\prod_{t=1}^T \left| \hat{\xi}_t^{\text{TE}} \right|^N} \prod_{k=1}^N \frac{\left(q(\hat{\xi}_0^{\text{TE}}, \mathbf{x}_k^t) \right)^p}{\left(q(\hat{\xi}_t^{\text{TE}}, \mathbf{x}_k^t) \right)^p} \begin{array}{c} \text{H}_1 \\ \gtrless \\ \text{H}_0 \end{array} \quad (3.27)$$

where:

$$\hat{\xi}_t^{\text{TE}} = f_t^{\text{TE}}(\hat{\xi}_t^{\text{TE}}), \hat{\xi}_0^{\text{TE}} = \frac{1}{T} \sum_{t=1}^T f_t^{\text{TE}}(\hat{\xi}_0^{\text{TE}}) \text{ and}$$

$$f_t^{\text{TE}}(\xi) = \frac{p}{N} \sum_{k=1}^N \frac{\mathbf{x}_k^t \mathbf{x}_k^t \mathbf{H}}{q(\xi, \mathbf{x}_k^t)} \quad (3.28)$$

Proof. The expression of the GLRT can be obtained using the expression of the GLRT under $\mathbb{C}\mathcal{E}$ at eq. (3.4) with the density generator $g = 1/x^p$. Although the conditions mentioned to guarantee the convexity of the log-likelihood are not respected (mainly the non-decreasing property of $\phi(t) = tg'(t)/g(t)$), it has been shown in [Ollila and Tyler, 2014] that the well known Tyler estimator [Tyler, 1987] is the MLE of the shape matrix.

□

3.A.3 Proof of proposition 3.2.3 at p. 89

Proposition. *The GLRT ratio under hypotheses of Problem (3.9) is the following:*

$$\hat{\Lambda}_{\text{MT}} = \frac{\left| \hat{\xi}_0^{\text{MT}} \right|^{TN}}{\prod_{t=1}^T \left| \hat{\xi}_t^{\text{TE}} \right|^N} \prod_{k=1}^N \frac{\left(\sum_{t=1}^T q\left(\hat{\xi}_0^{\text{MT}}, \mathbf{x}_k^t\right) \right)^{Tp}}{T^{Tp} \prod_{t=1}^T \left(q\left(\hat{\xi}_t^{\text{TE}}, \mathbf{x}_k^t\right) \right)^p} \begin{array}{c} \text{H}_1 \\ \gtrless \\ \text{H}_0 \end{array} \quad (3.29)$$

where

$$\hat{\boldsymbol{\xi}}_0^{\text{MT}} = f_{N,T}^{\text{MT}}(\hat{\boldsymbol{\xi}}_0^{\text{MT}}) = \frac{p}{N} \sum_{k=1}^N \frac{\sum_{t=1}^T \mathbf{s}_k^t}{\sum_{t=1}^T q(\hat{\boldsymbol{\xi}}_0^{\text{MT}}, \mathbf{x}_k^t)}. \quad (3.30)$$

Proof. In this problem, we test a change in both texture and covariance parameters. Thus, the GLRT for this problem has the following form:

$$\hat{\Lambda} = \frac{\max_{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}; \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T)}{\max_{\boldsymbol{\theta}_0} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}; \boldsymbol{\theta}_0)} \quad (3.31)$$

where $\boldsymbol{\theta}_0 = \{\tau_1, \dots, t_N, \boldsymbol{\xi}_0\}$ and $\forall t \in \llbracket 1, T \rrbracket$, $\boldsymbol{\theta}_t = \{\tau_1^t, \dots, \tau_N^t, \boldsymbol{\xi}_t\}$.

Using the assumption that all observations are independent, we can rewrite:

$$\hat{\Lambda} = \frac{\max_{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T} \prod_{k=1}^{N_T} \prod_{t=1}^{T_k} p_{\mathbf{x}_k^t}^{\mathcal{CN}}(\mathbf{x}_k^t; \boldsymbol{\theta}_t)}{\max_{\boldsymbol{\theta}_0} \prod_{k=1}^{N_T} \prod_{t=1}^{T_k} p_{\mathbf{x}_k^t}^{\mathcal{CN}}(\mathbf{x}_k^t; \boldsymbol{\theta}_0)}.$$

This expression can be computed by optimizing the numerator and denominator separately. Then, the idea is to estimate each unknown parameter separately and plugging back the estimates. Indeed, as we show in 3.2.4, the negative log of the likelihood functions considered here are jointly g-convex with regards to the covariance and texture parameters. In this case, each stationary-point of the negative log-likelihood corresponds to a global minima which in turn corresponds to the global maxima of the likelihoods. Thus we can compute:

$$\hat{\Lambda} = \frac{\mathcal{L}_1(\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_T)}{\mathcal{L}_0(\hat{\boldsymbol{\theta}}_0)}, \quad (3.32)$$

where

$$\begin{aligned} \mathcal{L}_1(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T) &= \prod_{k=1}^{N_T} \prod_{t=1}^{T_k} p_{\mathbf{x}_k^t}^{\mathcal{CN}}(\mathbf{x}_k^t; \boldsymbol{\theta}_t), \\ \mathcal{L}_0(\boldsymbol{\theta}_0) &= \prod_{k=1}^{N_T} \prod_{t=1}^{T_k} p_{\mathbf{x}_k^t}^{\mathcal{CN}}(\mathbf{x}_k^t; \boldsymbol{\theta}_0), \\ \hat{\boldsymbol{\theta}}_0 &= \operatorname{argmax}_{\boldsymbol{\theta}_0} \mathcal{L}_0(\boldsymbol{\theta}_0), \\ \forall t \in \llbracket 1, T \rrbracket, \hat{\boldsymbol{\theta}}_t &= \operatorname{argmax}_{\boldsymbol{\theta}_t} \mathcal{L}_1(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T). \end{aligned}$$

We optimize \mathcal{L}_0 and \mathcal{L}_1 separately:

- Consider

$$\log \mathcal{L}_0 = -\pi^{TNp} - T N \log |\boldsymbol{\xi}_0| - T p \sum_{k=1}^N \log(\tau_k) - \sum_{\substack{k=1 \\ k=1}}^{t=T} \frac{q(\boldsymbol{\xi}_0, \mathbf{x}_k^t)}{\tau_k}.$$

Let $k \in \llbracket 1, N \rrbracket$, we solve:

$$\frac{\partial \log \mathcal{L}_0}{\partial \tau_k} = -T p \sum_{k=1}^N \frac{1}{\tau_k} + \sum_{t=1}^T \frac{q(\boldsymbol{\xi}_0, \mathbf{x}_k^t)}{\tau_k^2} = 0,$$

which leads to:

$$\hat{\tau}_k = \frac{1}{T p} \sum_{t=1}^T q(\boldsymbol{\xi}_0, \mathbf{x}_k^t). \quad (3.33)$$

Now we consider the optimization with regards to $\boldsymbol{\xi}_0$. We solve:

$$\frac{\partial \log \mathcal{L}_0}{\partial \boldsymbol{\xi}_0} = -T N \boldsymbol{\xi}_0^{-1} + \sum_{\substack{k=1 \\ k=1}}^{t=T} \frac{\mathbf{S}_k^t}{\tau_k} \boldsymbol{\xi}_0^{-2} = \mathbf{0}_{p^2},$$

which yields:

$$\hat{\boldsymbol{\xi}}_0 = \frac{1}{T N} \sum_{\substack{k=1 \\ k=1}}^{t=T} \frac{\mathbf{S}_k^t}{\tau_k}. \quad (3.34)$$

Then by plugging back the estimates of textures at eq. (3.33) in eq. (3.34), we obtain:

$$\hat{\boldsymbol{\xi}}_0 = \frac{p}{N} \sum_{k=1}^N \frac{\sum_{t=1}^T \mathbf{S}_k^t}{\sum_{t=1}^T q(\hat{\boldsymbol{\xi}}_0^{\text{MT}}, \mathbf{x}_k^t)}, \quad (3.35)$$

that we denote $\hat{\boldsymbol{\xi}}_0^{\text{MT}}$.

- For \mathcal{L}_1 , we consider the same procedure and optimize alternatively for each τ_k^t and $\boldsymbol{\xi}_t$. We have:

$$\log \mathcal{L}_1 = -\pi^{TNp} - N \sum_{t=1}^T \log |\boldsymbol{\xi}_t| - p \sum_{k=1}^N \log(\tau_k^t) - \sum_{\substack{k=1 \\ k=1}}^{t=T} \frac{q(\boldsymbol{\xi}_t, \mathbf{x}_k^t)}{\tau_k^t}.$$

Let $k \in \llbracket 1, N \rrbracket, t \in \llbracket 1, T \rrbracket$, solving

$$\frac{\partial \log \mathcal{L}_1}{\partial \tau_k^t} = 0,$$

yields:

$$\hat{\tau}_k^t = \frac{1}{p} q(\boldsymbol{\xi}_t, \mathbf{x}_k^t). \quad (3.36)$$

Let $t \in \llbracket 1, T \rrbracket$, we have to solve:

$$\frac{\partial \log \mathcal{L}_1}{\partial \boldsymbol{\xi}_t} = N \boldsymbol{\xi}_t^{-1} + \sum_{k=1}^{k=N} \frac{\mathbf{S}_k^t}{\tau_k} \boldsymbol{\xi}_t^{-2} = \mathbf{0}_{p^2}$$

which yields:

$$\hat{\boldsymbol{\xi}}_t = \frac{1}{N} \sum_{k=1}^{k=N} \frac{\mathbf{S}_k^t}{\tau_k^t}. \quad (3.37)$$

Then by plugging estimates of eq. (3.36) in (3.37), we obtain:

$$\hat{\boldsymbol{\xi}}_t = \frac{p}{N} \sum_{k=1}^{k=N} \frac{\mathbf{S}_k^t}{q(\hat{\boldsymbol{\xi}}_t, \mathbf{x}_k^t)}, \quad (3.38)$$

that we denote $\hat{\boldsymbol{\xi}}_t^{\text{TE}}$.

Finally, we have to compute:

$$\begin{aligned} \hat{\Lambda} &= \frac{\mathcal{L}_1(\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_T)}{\mathcal{L}_0(\hat{\boldsymbol{\theta}}_0)} \\ &= \frac{\prod_{\substack{k=1 \\ t=1}}^{k=N} \frac{1}{\pi^p |\hat{\boldsymbol{\xi}}_t^{\text{TE}}| (\hat{\tau}_k^t)^p} \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t)}{\hat{\tau}_k^t} \right\}}{\prod_{\substack{k=1 \\ t=1}}^{k=N} \frac{1}{\pi^p |\hat{\boldsymbol{\xi}}_0^{\text{MT}}| (\hat{\tau}_k^t)^p} \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_0^{\text{MT}}, \mathbf{x}_k^t)}{\hat{\tau}_k^t} \right\}} \\ &= \frac{\left| \hat{\boldsymbol{\xi}}_0^{\text{MT}} \right|^{TN} \prod_{\substack{k=1 \\ t=1}}^{k=N} \frac{(\hat{\tau}_k^t)^p}{(\hat{\tau}_k^t)^p} \exp \left\{ -\sum_{\substack{k=1 \\ t=1}}^{k=N} \frac{q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t)}{\hat{\tau}_k^t} \right\}}{\prod_{t=1}^T \left| \hat{\boldsymbol{\xi}}_t^{\text{TE}} \right|^N \prod_{\substack{k=1 \\ t=1}}^{k=N} \frac{(\hat{\tau}_k^t)^p}{(\hat{\tau}_k^t)^p} \exp \left\{ -\sum_{\substack{k=1 \\ t=1}}^{k=N} \frac{q(\hat{\boldsymbol{\xi}}_0^{\text{MT}}, \mathbf{x}_k^t)}{\hat{\tau}_k^t} \right\}} \end{aligned}$$

Now, if we replace the texture estimates by their expression at eq. (3.33) and eq. (3.36), we

have:

$$\begin{aligned}\hat{\Lambda} &= \frac{\left|\hat{\xi}_0^{\text{MT}}\right|^{TN}}{\prod_{t=1}^T \left|\hat{\xi}_t^{\text{TE}}\right|^N} \prod_{k=1}^N \frac{\left(\sum_{t=1}^T q\left(\hat{\xi}_0^{\text{MT}}, \mathbf{x}_k^t\right)\right)^{Tp}}{T^{Tp} \prod_{t=1}^T \left(q\left(\hat{\xi}_t^{\text{TE}}, \mathbf{x}_k^t\right)\right)^p} \frac{\exp \left\{-p \sum_{k=1}^N \sum_{t=1}^T \frac{q\left(\hat{\xi}_t^{\text{TE}}, \mathbf{x}_k^t\right)}{q\left(\hat{\xi}_0^{\text{MT}}, \mathbf{x}_k^t\right)}\right\}}{\exp \left\{-Tp \sum_{k=1}^N \sum_{t=1}^T \frac{q\left(\hat{\xi}_0^{\text{MT}}, \mathbf{x}_k^t\right)}{q\left(\hat{\xi}_0^{\text{MT}}, \mathbf{x}_k^t\right)}\right\}} \\ &= \frac{\left|\hat{\xi}_0^{\text{MT}}\right|^{TN}}{\prod_{t=1}^T \left|\hat{\xi}_t^{\text{TE}}\right|^N} \prod_{k=1}^N \frac{\left(\sum_{t=1}^T q\left(\hat{\xi}_0^{\text{MT}}, \mathbf{x}_k^t\right)\right)^{Tp}}{T^{Tp} \prod_{t=1}^T \left(q\left(\hat{\xi}_t^{\text{TE}}, \mathbf{x}_k^t\right)\right)^p}.\end{aligned}$$

Since the covariance estimates are solution to fixed-point equations, we do not replace them and have the final form of the statistic. \square

3.A.4 Proof of proposition 3.2.4 at p. 89

Proposition. *The GLRT ratio under hypotheses of problem (3.10) is the following:*

$$\hat{\Lambda}_{\text{Mat}} = \frac{\left|\hat{\xi}_0^{\text{Mat}}\right|^{TN}}{\prod_{t=1}^T \left|\hat{\xi}_t^{\text{TE}}\right|^N} \prod_{k=1}^N \frac{\left(q\left(\hat{\xi}_0^{\text{Mat}}, \mathbf{x}_k^t\right)\right)^p}{\left(q\left(\hat{\xi}_t^{\text{TE}}, \mathbf{x}_k^t\right)\right)^p} \stackrel{\text{H}_1}{\gtrless} \lambda, \quad (3.39)$$

where

$$\hat{\xi}_0^{\text{Mat}} = f_{N,T}^{\text{Mat}}\left(\hat{\xi}_0^{\text{Mat}}\right) = \frac{p}{TN} \sum_{k=1}^N \sum_{t=1}^T \frac{\mathbf{S}_k^t}{q\left(\hat{\xi}_0^{\text{Mat}}, \mathbf{x}_k^t\right)}.$$

Proof. In this problem, we test a change in the covariance shape only. Thus, the GLRT for this problem has the following form:

$$\hat{\Lambda} = \frac{\max_{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}; \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T)}{\max_{\boldsymbol{\theta}_0, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}; \boldsymbol{\theta}_0, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T)} \quad (3.40)$$

where $\boldsymbol{\theta}_0 = \{\xi_0\}$, $\forall t \in \llbracket 1, T \rrbracket$, $\boldsymbol{\theta}_t = \{\xi_t\}$ and $\forall t \in \llbracket 1, T \rrbracket$, $\boldsymbol{\Phi}_t = \{\tau_1^t, \dots, \tau_N^t\}$.

Using the assumption that all observations are independent, we can rewrite:

$$\hat{\Lambda} = \frac{\max_{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T} \prod_{k=1}^N \prod_{t=1}^T p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^t; \boldsymbol{\theta}_t, \boldsymbol{\Phi}_t)}{\max_{\boldsymbol{\theta}_0, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T} \prod_{k=1}^N \prod_{t=1}^T p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^t; \boldsymbol{\theta}_0, \boldsymbol{\Phi}_T)}.$$

This expression can be computed by optimizing the numerator and denominator separately just as done in the previous derivations at section 3.A.3 and computing:

$$\hat{\Lambda} = \frac{\mathcal{L}_1(\hat{\theta}_1, \dots, \hat{\theta}_T, \hat{\Phi}_1^1, \dots, \hat{\Phi}_T^1)}{\mathcal{L}_0(\hat{\theta}_0, \hat{\Phi}_1^0, \dots, \hat{\Phi}_T^0)}, \quad (3.41)$$

where

$$\begin{aligned} \mathcal{L}_1(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T) &= \prod_{\substack{k=1 \\ t=1}}^{k=N \\ t=T} p_{\mathbf{x}_k^t}^{\mathcal{CN}}(\mathbf{x}_k^t; \boldsymbol{\theta}_t, \boldsymbol{\Phi}_t), \\ \mathcal{L}_0(\boldsymbol{\theta}_0, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T) &= \prod_{\substack{k=1 \\ t=1}}^{k=N \\ t=T} p_{\mathbf{x}_k^t}^{\mathcal{CN}}(\mathbf{x}_k^t; \boldsymbol{\theta}_0, \boldsymbol{\Phi}_t), \\ \hat{\boldsymbol{\theta}}_0 &= \operatorname{argmax}_{\boldsymbol{\theta}_0} \mathcal{L}_0(\boldsymbol{\theta}_0, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T), \\ \forall t \in \llbracket 1, T \rrbracket, \hat{\boldsymbol{\Phi}}_t^0 &= \operatorname{argmax}_{\boldsymbol{\Phi}_t} \mathcal{L}_0(\boldsymbol{\theta}_0, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T), \\ \forall t \in \llbracket 1, T \rrbracket, \hat{\boldsymbol{\theta}}_t &= \operatorname{argmax}_{\boldsymbol{\theta}_t} \mathcal{L}_1(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T), \\ \forall t \in \llbracket 1, T \rrbracket, \hat{\boldsymbol{\Phi}}_t^1 &= \operatorname{argmax}_{\boldsymbol{\Phi}_t} \mathcal{L}_1(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_T, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_T). \end{aligned}$$

Here, the optimization towards $\boldsymbol{\theta}_t$ and $\boldsymbol{\Phi}_t^1$ is exactly the same as done in section 3.A.3 where the parameters $\boldsymbol{\Phi}_t$ were compromised in the $\boldsymbol{\theta}_t$. Thus we will omit them here and only remind the results:

$$\begin{aligned} \forall t \in \llbracket 1, T \rrbracket, \hat{\tau}_k^t &= \hat{\tau}_{1k}^t = \frac{1}{p} q(\boldsymbol{\xi}_t, \mathbf{x}_k^t), \\ \forall t \in \llbracket 1, T \rrbracket, \hat{\boldsymbol{\xi}}_t &= \frac{p}{N} \sum_{k=1}^N \frac{\mathbf{S}_k^t}{q(\hat{\boldsymbol{\xi}}_t, \mathbf{x}_k^t)}. \end{aligned}$$

Concerning the others estimation problems, we have:

$$\log \mathcal{L}_0 = -\pi^{TNp} - TN \log |\boldsymbol{\xi}_0| - p \sum_{\substack{k=1 \\ t=1}}^{k=N \\ t=T} \log(\tau_k^t) - \sum_{\substack{k=1 \\ t=1}}^{k=N \\ t=T} \frac{q(\boldsymbol{\xi}_0, \mathbf{x}_k^t)}{\tau_k^t}.$$

The optimization towards each τ_k^t leads to:

$$\forall k \in \llbracket 1, N \rrbracket, \forall t \in \llbracket 1, T \rrbracket, \hat{\tau}_k^t = \hat{\tau}_{0k}^t = \frac{1}{p} q(\boldsymbol{\xi}_t, \mathbf{x}_k^t) \quad (3.42)$$

The optimization towards $\boldsymbol{\xi}_0$ was solved using the same procedure that led to eq. (3.34) gives:

$$\hat{\boldsymbol{\xi}}_0 = \frac{1}{TN} \sum_{\substack{k=1 \\ t=1}}^{k=N \\ t=T} \frac{\mathbf{S}_k^t}{\tau_{0k}^t}. \quad (3.43)$$

By plugging back eq. (3.42) in eq. (3.43), we obtain:

$$\hat{\boldsymbol{\xi}}_0 = \frac{p}{TN} \sum_{k=1}^{k=N} \sum_{t=1}^{t=T} \frac{\mathbf{s}_k^t}{q(\hat{\boldsymbol{\xi}}_t, \mathbf{x}_k^t)}, \quad (3.44)$$

that we denote $\hat{\boldsymbol{\xi}}_0^{\text{Mat}}$.

Finally, we have to compute:

$$\begin{aligned} \hat{\Lambda} &= \frac{\mathcal{L}_1(\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_T, \hat{\Phi}_1^1, \dots, \hat{\Phi}_T^1)}{\mathcal{L}_0(\hat{\boldsymbol{\theta}}_0, \hat{\Phi}_1^0, \dots, \hat{\Phi}_T^0)}, \\ &= \frac{\prod_{k=1}^{k=N} \prod_{t=1}^{t=T} \frac{1}{\pi^p |\hat{\boldsymbol{\xi}}_t^{\text{TE}}| (\hat{\tau}_{1k}^t)^p} \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t)}{\hat{\tau}_{1k}^t} \right\}}{\prod_{k=1}^{k=N} \prod_{t=1}^{t=T} \frac{1}{\pi^p |\hat{\boldsymbol{\xi}}_0^{\text{Mat}}| (\hat{\tau}_{0k}^t)^p} \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_0^{\text{Mat}}, \mathbf{x}_k^t)}{\hat{\tau}_{0k}^t} \right\}} \\ &= \frac{\left| \hat{\boldsymbol{\xi}}_0^{\text{Mat}} \right|^{TN} \prod_{k=1}^{k=N} \prod_{t=1}^{t=T} \frac{(\hat{\tau}_{0k}^t)^p}{(\hat{\tau}_{1k}^t)^p} \frac{\exp \left\{ -\sum_{k=1}^{k=N} \frac{q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t)}{\hat{\tau}_{1k}^t} \right\}}{\exp \left\{ -\sum_{k=1}^{k=N} \frac{q(\hat{\boldsymbol{\xi}}_0^{\text{Mat}}, \mathbf{x}_k^t)}{\hat{\tau}_{0k}^t} \right\}}}{\prod_{t=1}^T \left| \hat{\boldsymbol{\xi}}_t^{\text{TE}} \right|^N} \end{aligned}$$

When replacing the texture estimates by their expression, we obtain:

$$\hat{\Lambda} = \frac{\left| \hat{\boldsymbol{\xi}}_0^{\text{Mat}} \right|^{TN} \prod_{k=1}^{k=N} \prod_{t=1}^{t=T} \frac{\left(q(\hat{\boldsymbol{\xi}}_0^{\text{Mat}}, \mathbf{x}_k^t) \right)^p}{\left(q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t) \right)^p}}{\prod_{t=1}^T \left| \hat{\boldsymbol{\xi}}_t^{\text{TE}} \right|^N} \quad (3.45)$$

□

3.A.5 Proof of proposition 3.2.5 at p. 90

Proposition. *The GLRT ratio under hypotheses of problem (3.11) is the following:*

$$\hat{\Lambda}_{\text{Tex}} = \prod_{t=1}^T \frac{\left| \hat{\boldsymbol{\xi}}_t^{\text{Tex}} \right|^N}{\left| \hat{\boldsymbol{\xi}}_t^{\text{TE}} \right|^N} \prod_{k=1}^N \frac{\left(\sum_{t=1}^T q(\hat{\boldsymbol{\xi}}_t^{\text{Tex}}, \mathbf{x}_k^t) \right)^{Tp}}{T^{Tp} \prod_{t=1}^T \left(q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t) \right)^p} \stackrel{H_1}{\gtrless} \stackrel{H_0}{\lessgtr} \lambda, \quad (3.46)$$

where

$$\hat{\xi}_t^{\text{Tex}} = f_{N,T,t}^{\text{Tex}} \left(\hat{\xi}_1^{\text{Tex}}, \dots, \hat{\xi}_T^{\text{Tex}} \right), \quad (3.47)$$

$$= \frac{T p}{N} \sum_{k=1}^N \frac{\mathbf{S}_k^t}{\sum_{t'=1}^T q \left(\hat{\xi}_{t'}^{\text{Tex}}, \mathbf{x}_k^t \right)}. \quad (3.48)$$

Proof. In this problem, we test a change in the texture parameters only. Thus, the GLRT for this problem has the following form:

$$\hat{\Lambda} = \frac{\max_{\theta_1, \dots, \theta_T, \Phi_1, \dots, \Phi_T} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}; \theta_1, \dots, \theta_T, \Phi_1, \dots, \Phi_T)}{\max_{\theta_0, \Phi_1, \dots, \Phi_T} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}; \theta_0, \Phi_1, \dots, \Phi_T)} \quad (3.49)$$

where $\theta_0 = \{\tau_1, \dots, \tau_N\}$, $\forall t \in \llbracket 1, T \rrbracket$, $\theta_t = \{\tau_1^t, \dots, \tau_N^t\}$ and $\forall t \in \llbracket 1, T \rrbracket$, $\Phi_t = \{\xi_t\}$.

Using the assumption that all observations are independent, we can rewrite:

$$\hat{\Lambda} = \frac{\max_{\theta_1, \dots, \theta_T, \Phi_1, \dots, \Phi_T} \prod_{t=1}^{k=N} p_{\mathbf{x}_k^t}^{\mathcal{CN}}(\mathbf{x}_k^t; \theta_t, \Phi_t)}{\max_{\theta_0, \Phi_1, \dots, \Phi_T} \prod_{t=1}^{k=N} p_{\mathbf{x}_k^t}^{\mathcal{CN}}(\mathbf{x}_k^t; \theta_0, \Phi_T)}.$$

This expression can be computed by optimizing the numerator and denominator separately just as done in the previous derivations at section 3.A.3 and computing:

$$\hat{\Lambda} = \frac{\mathcal{L}_1 \left(\hat{\theta}_1, \dots, \hat{\theta}_T, \hat{\Phi}_1^1, \dots, \hat{\Phi}_T^1 \right)}{\mathcal{L}_0 \left(\hat{\theta}_0, \hat{\Phi}_1^0, \dots, \hat{\Phi}_T^0 \right)}, \quad (3.50)$$

where

$$\begin{aligned} \mathcal{L}_1(\theta_1, \dots, \theta_T, \Phi_1, \dots, \Phi_T) &= \prod_{t=1}^{k=N} p_{\mathbf{x}_k^t}^{\mathcal{CN}}(\mathbf{x}_k^t; \theta_t, \Phi_t), \\ \mathcal{L}_0(\theta_0, \Phi_1, \dots, \Phi_T) &= \prod_{t=1}^{k=N} p_{\mathbf{x}_k^t}^{\mathcal{CN}}(\mathbf{x}_k^t; \theta_0, \Phi_t), \\ \hat{\theta}_0 &= \underset{\theta_0}{\operatorname{argmax}} \mathcal{L}_0(\theta_0, \Phi_1, \dots, \Phi_T), \\ \forall t \in \llbracket 1, T \rrbracket, \hat{\Phi}_t^0 &= \underset{\Phi_t}{\operatorname{argmax}} \mathcal{L}_0(\theta_0, \Phi_1, \dots, \Phi_T), \\ \forall t \in \llbracket 1, T \rrbracket, \hat{\theta}_t &= \underset{\theta_t}{\operatorname{argmax}} \mathcal{L}_1(\theta_1, \dots, \theta_T, \Phi_1, \dots, \Phi_T), \\ \forall t \in \llbracket 1, T \rrbracket, \hat{\Phi}_t^1 &= \underset{\Phi_t}{\operatorname{argmax}} \mathcal{L}_1(\theta_1, \dots, \theta_T, \Phi_1, \dots, \Phi_T). \end{aligned}$$

Here, the optimization towards $\boldsymbol{\theta}_t$ and $\boldsymbol{\Phi}_t^1$ is exactly the same as done in section 3.A.3 where the parameters $\boldsymbol{\Phi}_t$ were compromised in the $\boldsymbol{\theta}_t$. Thus we will omit them here and only remind the results:

$$\begin{aligned}\forall t \in \llbracket 1, T \rrbracket, \hat{\tau}_k^t &= \hat{\tau}_{1k}^t = \frac{1}{p} q(\boldsymbol{\xi}_t, \mathbf{x}_k^t), \\ \forall t \in \llbracket 1, T \rrbracket, \hat{\boldsymbol{\xi}}_t &= \frac{p}{N} \sum_{k=1}^N \frac{\mathbf{S}_k^t}{q(\hat{\boldsymbol{\xi}}_t, \mathbf{x}_k^t)}.\end{aligned}$$

Concerning the others estimation problems, we have:

$$\log \mathcal{L}_0 = -\pi^{TNp} - N \sum_{t=1}^{t=T} \log |\boldsymbol{\xi}_t| - p \sum_{k=1}^{k=N} \log (\tau_k) - \sum_{t=1}^{t=T} \frac{q(\boldsymbol{\xi}_t, \mathbf{x}_k^t)}{\tau_k}.$$

The optimization towards each τ_k^t leads to:

$$\forall k \in \llbracket 1, N \rrbracket, \hat{\tau}_k = \hat{\tau}_{0k} = \frac{1}{pT} \sum_{t=1}^{t=T} q(\boldsymbol{\xi}_t, \mathbf{x}_k^t). \quad (3.51)$$

The optimization towards each $\boldsymbol{\xi}_t$ gives:

$$\hat{\boldsymbol{\xi}}_0 = \frac{1}{N} \sum_{k=1}^{k=N} \frac{\mathbf{S}_k^t}{\tau_{0k}}. \quad (3.52)$$

And by plugging back eq. (3.42) in eq. (3.43), we obtain:

$$\hat{\boldsymbol{\xi}}_t = \frac{Tp}{N} \sum_{t'=1}^{t=T} \frac{\mathbf{S}_k^t}{\sum_{t'=1}^{t=T} q(\hat{\boldsymbol{\xi}}_{t'}, \mathbf{x}_k^{(t')})}, \quad (3.53)$$

that we denote $\hat{\boldsymbol{\xi}}_t^{\text{Tex}}$.

Finally, we have to compute:

$$\begin{aligned}\hat{\Lambda} &= \frac{\mathcal{L}_1(\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\Phi}}_1^1, \dots, \hat{\boldsymbol{\Phi}}_T^1)}{\mathcal{L}_0(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\Phi}}_1^0, \dots, \hat{\boldsymbol{\Phi}}_T^0)}, \\ &= \frac{\prod_{k=1}^{k=N} \frac{1}{\pi^p |\hat{\boldsymbol{\xi}}_t^{\text{TE}}| (\hat{\tau}_{1k}^t)^p} \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t)}{\hat{\tau}_{1k}^t} \right\}}{\prod_{k=1}^{k=N} \frac{1}{\pi^p |\hat{\boldsymbol{\xi}}_t^{\text{Tex}}| (\hat{\tau}_{0k}^t)^p} \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_t^{\text{Tex}}, \mathbf{x}_k^t)}{\hat{\tau}_{0k}^t} \right\}}\end{aligned}$$

$$= \prod_{t=1}^T \frac{\left| \hat{\boldsymbol{\xi}}_t^{\text{Tex}} \right|^N}{\left| \hat{\boldsymbol{\xi}}_t^{\text{TE}} \right|^N} \prod_{k=1}^N \frac{(\hat{\tau}_{0k}^t)^p}{(\hat{\tau}_{1k}^t)^p} \frac{\exp \left\{ - \sum_{\substack{k=1 \\ t=1}}^{k=N} \frac{q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t)}{\hat{\tau}_{1k}^t} \right\}}{\exp \left\{ - \sum_{\substack{k=1 \\ t=1}}^{k=N} \frac{q(\hat{\boldsymbol{\xi}}_0^{\text{Mat}}, \mathbf{x}_k^t)}{\hat{\tau}_{0k}^t} \right\}}$$

When replacing the texture estimates by their expression, we obtain:

$$\hat{\Lambda}_{\text{Tex}} = \prod_{t=1}^T \frac{\left| \hat{\boldsymbol{\xi}}_t^{\text{Tex}} \right|^N}{\left| \hat{\boldsymbol{\xi}}_t^{\text{TE}} \right|^N} \prod_{k=1}^N \frac{\left(\sum_{t=1}^T q(\hat{\boldsymbol{\xi}}_t^{\text{Tex}}, \mathbf{x}_k^t) \right)^{Tp}}{T^{Tp} \prod_{t=1}^T \left(q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t) \right)^p} \stackrel{H_1}{\gtrless} \lambda, \quad (3.54)$$

□

3.A.6 Proof of proposition 3.2.6 at p. 91

Proposition. $\hat{\boldsymbol{\xi}}_0^{\text{MT}}$, $\hat{\boldsymbol{\xi}}_0^{\text{Mat}}$ and $\hat{\boldsymbol{\xi}}_t^{\text{Tex}}$ are the unique arguments of the global maxima of their respective log-likelihood cost functions over the observations.

Proof. We will consider here only the proof for $\hat{\boldsymbol{\xi}}_0^{\text{MT}}$, since the same procedure can be applied to show the property for the others. In other words, we will show that $\hat{\tau}_k$ at eq. (3.33) and $\hat{\boldsymbol{\xi}}_0^{\text{MT}}$ at eq. (3.35) are the global maxima of the following log-likelihood function:

$$\begin{aligned} \log \mathcal{L}(\tau_1, \dots, \tau_N, \boldsymbol{\xi}_0) &= -\pi^{TNp} - TN \log |\boldsymbol{\xi}_0| - \\ &\quad Tp \sum_{k=1}^N \log(\tau_k) - \sum_{\substack{k=1 \\ t=1}}^{k=N} \frac{q(\boldsymbol{\xi}_0, \mathbf{x}_k^t)}{\tau_k}. \end{aligned} \quad (3.55)$$

We first recall two useful results:

Lemma 1. Any local minimum of a g-convex function over a manifold \mathcal{M} is a global minimum.

Lemma 2. Consider the manifold $\mathbb{S}_{\mathbb{H}}^p$ and the following geodesic:

$$\boldsymbol{\xi}_t^{\boldsymbol{\xi}_0, \boldsymbol{\xi}_1} = \boldsymbol{\xi}_0^{\frac{1}{2}} \left(\boldsymbol{\xi}_0^{-\frac{1}{2}} \boldsymbol{\xi}_1 \boldsymbol{\xi}_0^{-\frac{1}{2}} \right)^t \boldsymbol{\xi}_0^{\frac{1}{2}}, \quad t \in [0, 1],$$

between two points $\boldsymbol{\xi}_0, \boldsymbol{\xi}_1$.

Let $\mathbf{h}_i \in \mathbb{C}^p$, $a \in \pm 1$, $a' \in \pm 1$ for $i = 1, \dots, m$ and $\mathbf{H}_i \in \mathbb{C}^{q,p}$ for $i = 1, \dots, n$. The function

$$\mathcal{L}(\boldsymbol{\xi}) = \log \left| \sum_{i=1}^n \mathbf{H}_i \boldsymbol{\xi}^a \mathbf{H}_i^H \right| + \sum_{i'=1}^m \mathbf{h}_{i'}^H \boldsymbol{\xi}^{a'} \mathbf{h}_{i'}, \quad (3.56)$$

is strictly g-convex in $\boldsymbol{\xi} \in \mathbb{S}_{\mathbb{H}}^p$.

When looking at the negative of function $\log \mathcal{L}$ in eq. (3.55), straightforward application of Lemma 2 allows to conclude that it is jointly g-convex in $\boldsymbol{\xi}_0$ and for all τ_k :

- the g-convexity for each τ_k is obtained by rewriting the negative of eq. (3.55) in the form of eq. (3.56), if we take³:

$$\begin{aligned}\boldsymbol{\xi} &= \tau_k, \\ a &= 1, \\ a' &= -1, \\ \{\mathbf{h}_i | i \in \llbracket 1, m \rrbracket\} &= \left\{ \mathbf{x}_k^t \boldsymbol{\xi}_0^{-\frac{1}{2}} | k \in \llbracket 1, N \rrbracket, t \in \llbracket 1, T \rrbracket \right\}, \\ \mathbf{H}_i &= \tau_i \delta_{ik}.\end{aligned}$$

Here, $\boldsymbol{\xi}$ and \mathbf{H}_i are taken as matrices of size 1×1 .

- the g-convexity in $\boldsymbol{\xi}_0$ is obtained by rewriting the negative of eq. (3.55) in the form of eq. (3.56), if we take⁴:

$$\begin{aligned}\boldsymbol{\xi} &= \boldsymbol{\xi}_0, \\ a &= 1, \\ a' &= -1, \\ \{\mathbf{h}_i | i \in \llbracket 1, m \rrbracket\} &= \left\{ \frac{\mathbf{x}_k^t}{\sqrt{\tau_k}} | k \in \llbracket 1, N \rrbracket, t \in \llbracket 1, T \rrbracket \right\}, \\ \mathbf{H}_i &= \delta_{i1}.\end{aligned}$$

Here, \mathbf{H}_i are taken as matrices of size 1×1 .

So we have the strict g-convexity, application of Lemma 1 allows us to conclude that the estimates correspond to unique global maxima.

□

3.A.7 Proof of proposition 3.2.7 at p. 91

Proposition. $\hat{\boldsymbol{\xi}}_0^{\text{MT}}$, $\hat{\boldsymbol{\xi}}_0^{\text{Mat}}$ and $\hat{\boldsymbol{\xi}}_t^{\text{Tex}}$ are the arguments to the global minima obtained inside $\mathbb{S}_{\mathbb{H}}^p$.

Proof. Again, we will only consider the case of $\hat{\boldsymbol{\xi}}_0^{\text{MT}}$, since the same considerations lead to the result for the others. Up to now we have only shown that the negative log-likelihood $-\log \mathcal{L}$ is g-convex. To show that it has a unique minimum in $\mathbb{S}_{\mathbb{H}}^p$, and thus that the fixed-point equation to $\hat{\boldsymbol{\xi}}_0^{\text{MT}}$ admits a unique solution within the manifold, it suffices to show that the minimum of $\log \mathcal{L}$ occurs in the interior of $\mathbb{S}_{\mathbb{H}}^p$. To this end we have to show that $\log \mathcal{L}(\boldsymbol{\xi}) \rightarrow \infty$ as $\boldsymbol{\xi} \rightarrow \text{Bound}(\mathbb{S}_{\mathbb{H}}^p)$, the boundary of $\mathbb{S}_{\mathbb{H}}^p$.

Let $\lambda_1(\boldsymbol{\xi}), \dots, \lambda_p(\boldsymbol{\xi})$, be the ordered eigenvalues of $\boldsymbol{\xi}$. We can rewrite $-\log \mathcal{L}$ as⁵:

$$T N \sum_{j=1}^p \log \lambda_j(\boldsymbol{\xi}) + \sum_{k=1}^{t=T} \frac{q(\boldsymbol{\xi}, \mathbf{x}_k^t)}{\tau_k}.$$

³Considering solely the terms involving the considered texture parameter.

⁴Considering solely the terms involving the covariance matrix.

⁵Omitting the constants with regards to $\boldsymbol{\xi}_0$

Now, decomposing $\boldsymbol{\xi}$ as $\boldsymbol{\xi}_{\text{EVD}} = \mathbf{P}^H \mathbf{D} \mathbf{P}$ we can write $q(\boldsymbol{\xi}, \mathbf{x}_k^t)$ as:

$$\sum_{j=1}^p |[\mathbf{y}_k^t]_j|^2 / \lambda_j(\boldsymbol{\xi}),$$

where $[\mathbf{y}_k^t]_j$ is the j -th element of $\mathbf{y}_k^t = \mathbf{P}^H \mathbf{x}_k^t$. Then we have:

$$-\log \mathcal{L}(\boldsymbol{\xi}) \geq \sum_{\substack{j=p \\ k=1 \\ j=1}}^{t=T} \frac{|[\mathbf{y}_k^t]_j|^2}{\lambda_j(\boldsymbol{\xi}) \tau_k} + T N \sum_{j=1}^p \log \lambda_j(\boldsymbol{\xi}),$$

Finally, $\boldsymbol{\xi} \rightarrow \text{Bound}(\mathbb{S}_{\mathbb{H}}^p)$ if and only if $\lambda_1(\boldsymbol{\xi}) \rightarrow \infty$ and/or $\lambda_p(\boldsymbol{\xi}) \rightarrow 0$. Under both regimes the right-hand side of the previous equation goes to ∞ , which concludes the proof. \square

3.A.8 Proof of proposition 3.2.8 at p. 92

Proposition. Let $\{\mathbf{x}_k^t | k \in \llbracket 1, N \rrbracket, t \in \llbracket 1, T \rrbracket\}$ be a set of observations. Let us define vectors $\mathbf{v}_i \in \mathbb{R}^p$ such that $\forall k, \forall t, \mathbf{v}_{(T-1)*N+k} = (\Re(\mathbf{x}_k^t)^T, \Im(\mathbf{x}_k^t)^T)^T$ and $\mathbf{v}_{(2T-1)*N+k} = (-\Im(\mathbf{x}_k^t)^T, \Re(\mathbf{x}_k^t)^T)^T$. Let $\mathbb{P}_{2TN}(\bullet)$ be the empirical distribution of samples $\{\mathbf{v}_i | i \in \llbracket 1, 2TN \rrbracket\}$. Then the fixed-point algorithms $(\boldsymbol{\xi}_0^{\text{MT}})_{k+1} = f_{N,T}^{\text{MT}}((\boldsymbol{\xi}_0^{\text{MT}})_k)$ and $(\boldsymbol{\xi}_0^{\text{Mat}})_{k+1} = f_{N,T}^{\text{Mat}}((\boldsymbol{\xi}_0^{\text{Mat}})_k)$ converge to unique solutions up to a scale factor if and only if the following condition is respected:

(C1) $\mathbb{P}_{2TN}(\{\mathbf{0}\}) = 0$ and for all linear subspaces $V \subset \mathbb{R}^{2p}$, we have $\mathbb{P}_{2TN}(V) < \dim(V)/2p$.

Proof. The proof is done in three steps: we first go from the complex dataset to an equivalent real one. Then, we prove the sufficient implication and then the necessary one.

- Let us define, as done in [Mahot et al., 2013], the mapping from $\mathbb{C}^{p \times p}$ to $\mathbb{R}^{2p \times 2p}$ as a function denoted $f_{\mathbb{CR}} : \mathbb{S}_{\mathbb{H}}^p \rightarrow \mathbb{S}_{++}^{2p}$, whose definition is:

$$f_{\mathbb{CR}}(\boldsymbol{\xi}) = \frac{1}{2} \begin{bmatrix} \Re(\boldsymbol{\xi}) & -\Im(\boldsymbol{\xi}) \\ \Im(\boldsymbol{\xi}) & \Re(\boldsymbol{\xi}) \end{bmatrix}. \quad (3.57)$$

Given the observations $\{\mathbf{x}_k^t | k \in \llbracket 1, N \rrbracket, t \in \llbracket 1, T \rrbracket\}$, we define $\mathbf{v}_i \in \mathbb{R}^p$ such that $\forall k, \forall t, \mathbf{v}_{(T-1)*N+k} = (\Re(\mathbf{x}_k^t)^T, \Im(\mathbf{x}_k^t)^T)^T$ and $\mathbf{v}_{(2T-1)*N+k} = (-\Im(\mathbf{x}_k^t)^T, \Re(\mathbf{x}_k^t)^T)^T$. Using identities of Theorem 1 in [Ollila et al., 2012a], we can show that $f_{\mathbb{CR}}(\hat{\boldsymbol{\xi}}_0^{\text{MT}})$ is solution to the following fixed-point equation:

$$\boldsymbol{\xi} = \frac{p}{2N} \sum_{k=1}^{2N} \frac{\sum_{t=1}^T \mathbf{v}_i^t \mathbf{v}_i^{t^T}}{\sum_{t=1}^T \mathbf{v}_i^{t^T} \boldsymbol{\xi}^{-1} \mathbf{v}_i^t}. \quad (3.58)$$

Since there is an equivalence between \mathbb{R}^{2p} and \mathbb{C}^p , we can consider the convergence of $f_{\mathbb{CR}}(\hat{\boldsymbol{\xi}}_0^{\text{MT}})$ and the result will apply to the complex case.

2. Now we prove that if the condition **C1** is respected, the fixed-point algorithm to compute the solution of eq. (3.58) converges. Since the transformation $\mathbf{x} \rightarrow \mathbf{M}\mathbf{x}$, for any non-singular matrix \mathbf{M} , leads to the transformation $\boldsymbol{\xi} \rightarrow \mathbf{M}\boldsymbol{\xi}\mathbf{M}^T$ in eq. (3.58), we can assume $\boldsymbol{\xi} = \mathbf{I}_{2p}$ without loss of generality. We will show the convergence in three steps:

Step 1. Let $\lambda_1(\boldsymbol{\xi}_k), \dots, \lambda_{2p}(\boldsymbol{\xi}_k)$, be the ordered eigenvalues of $\boldsymbol{\xi}_k$ the matrix at iteration k of the algorithm. We remark that for any $\mathbf{x} \in \mathbb{R}^{2p}$, $\mathbf{x}^T \boldsymbol{\xi}_k^{-1} \mathbf{x} \geq \lambda_1(\boldsymbol{\xi}_k)^{-1} \mathbf{x}^T \mathbf{x}$ so we can write the following inequality:

$$\boldsymbol{\xi}_{k+1} \leq \lambda_1(\boldsymbol{\xi}_k)(2N)^{-1} p \sum_{i=1}^{2N} \frac{\sum_{t=1}^T \mathbf{v}_i^t \mathbf{v}_i^{t^T}}{\sum_{t=1}^T \mathbf{v}_i^t \mathbf{v}_i^t} = \lambda_1(\boldsymbol{\xi}_k) \mathbf{I}_p,$$

where the ordering refers to the partial ordering of symmetric matrices. Similarly, we can write $\boldsymbol{\xi}_{k+1} \geq \lambda_{2p}(\boldsymbol{\xi}_k) \mathbf{I}_{2p}$. These two inequalities imply that

$$\lambda_1(\boldsymbol{\xi}_{k+1}) \leq \lambda_1(\boldsymbol{\xi}_k) \text{ and } \lambda_{2p}(\boldsymbol{\xi}_{k+1}) \geq \lambda_{2p}(\boldsymbol{\xi}_k). \quad (3.59)$$

Step 2. Let \mathbf{P}_k be the $2p \times 2p$ symmetric idempotent matrix which projects orthogonally into the eigenspace $E_k = \{\mathbf{x} \in \mathbb{R}^{2p} | \boldsymbol{\xi}_k \mathbf{x} = \lambda_1(\boldsymbol{\xi}_k) \mathbf{x}\}$. We will show here that if $\lambda_1(\boldsymbol{\xi}_{k+1}) = \lambda_1(\boldsymbol{\xi}_k)$ then

$$E_{k+1} \subset E_k, \text{ with equality only if } E_k = \mathbb{R}^{2p}. \quad (3.60)$$

To this end, we use the fixed-point equation multiplied by \mathbf{P}_{k+1} , the inequality $\mathbf{v} \in \mathbb{R}^{2p}$, $\mathbf{v}^T \boldsymbol{\xi}_k^{-1} \mathbf{v} \geq \lambda_1(\boldsymbol{\xi}_k)^{-1} \mathbf{v}^T \mathbf{v}$ and the fact that $\boldsymbol{\xi} = \mathbf{I}_{2p}$ is a solution to the fixed point equation, to show that $\lambda_1(\boldsymbol{\xi}_{k+1}) \mathbf{P}_{k+1} \leq \lambda_1(\boldsymbol{\xi}_k) \mathbf{P}_{k+1}$. Equality is obtained if and only if the following condition is respected:

$$\forall k, \forall t, \mathbf{P}_{k+1} \mathbf{v}_k^t = \mathbf{0}_{2p} \text{ or } \mathbf{P}_k \mathbf{v}_k^t = \mathbf{v}_k^t. \quad (3.61)$$

Assume that this condition is respected and thus $\lambda_1(\boldsymbol{\xi}_{k+1}) = \lambda_1(\boldsymbol{\xi}_k)$. Then multiplying eq. (3.58) by \mathbf{P}_{k+1} and $(\mathbf{I}_{2p} - \mathbf{P}_k)$ and using (3.61) leads to $\lambda_1(\boldsymbol{\xi}_{k+1}) \mathbf{P}_{k+1} (\mathbf{I}_{2p} - \mathbf{P}_k) = \mathbf{0}_{2p,2p}$. Thus, we have: $\mathbf{P}_{k+1} = \mathbf{P}_k \mathbf{P}_{k+1}$. This means that $E_{k+1} \subset E_k$. If equality holds, then (3.61) implies $\mathbb{P}_{2TN}(E_k) + \mathbb{P}_{2TN}(E_k^\perp) = 1$. This contradicts condition (**C1**) unless $E_k = \mathbb{R}^{2p}$.

Step 3. Statement (3.59) implies that $\lambda_1(\boldsymbol{\xi}_k) \rightarrow \lambda_1$ and $\lambda_{2p}(\boldsymbol{\xi}_k) \rightarrow \lambda_{2p}$ for some λ_1 and λ_2 such that $0 < \lambda_{2p} \leq \lambda_1 < \infty$. It just suffices to be shown at this point that $\lambda_1 = \lambda_{2p}$ which implies $\boldsymbol{\xi}_k \rightarrow \lambda_1 \mathbf{I}_{2p}$.

Since $\{\boldsymbol{\xi}_k\}$ has bounded eigenvalues, there exist a convergent subsequence, for example, $\boldsymbol{\xi}_{k(2p)+1} \rightarrow \mathbf{A}_0 \in \mathbb{S}_{++}^{2p}$, which implies

$$\boldsymbol{\xi}_{k(2p)+1} \rightarrow \mathbf{A}_1 = (2N)^{-1} p \sum_{k=1}^N \left(\sum_{t=1}^T \mathbf{v}_k^t \mathbf{v}_k^{t^T} \right) / \left(\sum_{t=1}^T \mathbf{v}_k^t \mathbf{v}_k^t \right).$$

The largest and smallest eigenvalues of both \mathbf{A}_0 and \mathbf{A}_1 must be λ_1 and λ_{2p} . Let $E_{\infty,k} = \{\mathbf{x} \in \mathbb{R}^{2p} | \mathbf{A}_k \mathbf{x} = \lambda_1 \mathbf{x}\}$. We can assume without loss of generality that $\dim(E_{\infty,1}) \geq \dim(E_{\infty,0})$. If this assumption is not respected, we can replace the subsequence $\{\xi_{k(2p)}\}$ by $\{\xi_{k(2p)+1}\}$ and so on until the assumption is met. This condition on the dimension together with (3.60), implies that $E_{\infty,1} = \mathbb{R}^{2p}$. Thus, $\xi_{k(2p)} \rightarrow \lambda_1 \mathbf{I}_{2p}$ which implies $\lambda_1 = \lambda_{2p}$.

3. Finally we show that if a solution to eq. (3.58) exists, then condition **(C1)** is respected.

Without loss of generality, we assume again that $\xi = \mathbf{I}_{2p}$. Let S be a proper subspace and \mathbf{Q} be the idempotent matrix which projects orthogonally into S . Letting $\xi = \mathbf{I}_{2p}$, then multiplying both sides of eq. (3.58) by $(\mathbf{I}_{2p} - \mathbf{Q})$ and finally taking the trace yields:

$$2p - \dim(S) = (2N)^{-1} p \sum_{k=1}^N \left(\sum_{t=1}^T \mathbf{v}_k^t \mathbf{v}_k^t \right) / \left(\sum_{t=1}^T \mathbf{v}_k^t \mathbf{v}_k^t \right). \quad (3.62)$$

Since $\mathbf{x}^T (\mathbf{I}_{2p} - \mathbf{Q}) \mathbf{x} = 0$ for $\mathbf{x} \in S$ and $\mathbf{x}^T (\mathbf{I}_{2p} - \mathbf{Q}) \mathbf{x} \leq \mathbf{x}^T \mathbf{x}$ for $\mathbf{x} \notin S$, it follows from eq. (3.62) that $2p - \dim(S) \leq 2p(1 - \mathbb{P}_{2TN}(S))$ which is equivalent to condition **(C1)**. \square

3.A.9 Proof of proposition 3.2.10 at p. 95

Proposition. $\hat{\Lambda}_{MT}$ is CFAR texture and matrix for Problem (3.9).

Proof. We consider separately the texture and matrix properties:

- *Texture CFAR:* First, $\hat{\xi}_0^{MT}$ and $\hat{\xi}_t^{MT}$ are invariant by the substitution $\mathbf{x}_k^t \rightarrow \mathbf{x}_k^t / \tau_k^{(0)}$. Then, the different

$$\frac{\left(\sum_{t=1}^T q(\hat{\xi}_0^{MT}, \mathbf{x}_k^t) \right)^{Tp}}{T^{Tp} \prod_{t=1}^T \left(q(\hat{\xi}_t^{MT}, \mathbf{x}_k^t) \right)^p}$$

terms are also invariant by the same substitution. This means that the values of $\{\tau_k^{(0)} | k \in [1, N]\}$ do not affect the statistic of $\hat{\Lambda}_{MT}$, which is the definition of texture CFAR property in this problem.

- *Matrix CFAR:* As said in the discussion of 1.4.8, the estimates of matrices are subject to an indetermination which is resolved by an appropriate normalization. For any estimate $\hat{\xi} \in \{\hat{\xi}_0^{MT}, \hat{\xi}_1^{MT}, \dots, \hat{\xi}_T^{MT}\}$, when replacing $\hat{\xi}$ by $p \hat{\xi} / \text{Tr}(\hat{\xi})$ in eq. (3.12), the trace terms simplify in the expression of $\hat{\Lambda}_{MT}$. Thus, the statistic is homogeneous by the normalization constraint.

Then, the statistic is invariant for the group of transformation:

$$\mathcal{G} = \{ \mathbf{G} \mathbf{x}_k^t | t \in [1, T], k \in [1, N], \mathbf{G} \in \mathbb{S}_{\mathbb{H}}^p \}.$$

Indeed, we can write:

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

where all terms $\mathbf{G} \hat{\boldsymbol{\xi}}_0^{\text{MT}} \mathbf{G}$, $\mathbf{G} \hat{\boldsymbol{\xi}}_t^{\text{TE}} \mathbf{G}$, $q(\hat{\boldsymbol{\xi}}_0^{\text{MT}}, \mathbf{G} \mathbf{x}_k^t)$, $q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{G} \mathbf{x}_k^t)$ can be written as functions of $\{\mathbf{G} \mathbf{x}_k^t | t \in [1, T], k \in [1, N]\}$.

Finally by taking $\mathbf{G} = \boldsymbol{\xi}_0^{-1/2}$, the statistic is a function of $\{\boldsymbol{\xi}_0^{-1/2} \mathbf{x}_k^t | t \in [1, T], k \in [1, N]\}$ where $\boldsymbol{\xi}_0^{-1/2} \mathbf{x}_k^t \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$. It follows that the statistic is independent of $\boldsymbol{\xi}_0$ that ends the proof.

The same arguments of invariance are used for $\hat{\Lambda}_{\text{MT}}^{\text{marg}}$. \square

3.A.10 Proof of proposition 3.2.12 at p. 95

Proposition. $\hat{\Lambda}_{\text{Tex}}$ is CFAR texture but is not CFAR matrix for Problem (3.11).

Proof. The texture CFAR property is done using the same procedure as propositions 3.2.10 and 3.2.11.

The matrix CFAR property cannot be ensured due to the trace normalization. For any estimate $\hat{\boldsymbol{\xi}} \in \{\hat{\boldsymbol{\xi}}_1^{\text{TE}}, \dots, \hat{\boldsymbol{\xi}}_T^{\text{TE}}, \hat{\boldsymbol{\xi}}_1^{\text{Tex}}, \dots, \hat{\boldsymbol{\xi}}_T^{\text{Tex}}\}$, when replacing $\hat{\boldsymbol{\xi}}$ by $p \hat{\boldsymbol{\xi}} / \text{Tr}(\hat{\boldsymbol{\xi}})$ in eq. (3.15), we have:

$$\begin{aligned} \hat{\Lambda}_{\text{Tex}} &= \prod_{t=1}^T \frac{\left| \hat{\boldsymbol{\xi}}_t^{\text{Tex}} \right|}{\left| \hat{\boldsymbol{\xi}}_t^{\text{TE}} \right| \text{Tr}(\hat{\boldsymbol{\xi}}_t^{\text{Tex}})} \times \\ &\quad \prod_{k=1}^N \frac{\left(\sum_{t=1}^T \text{Tr}(\hat{\boldsymbol{\xi}}_t^{\text{Tex}}) q(\hat{\boldsymbol{\xi}}_t^{\text{Tex}}, \mathbf{x}_k^t) \right)^{Tp}}{T^{Tp} \prod_{t=1}^T \left(q(\hat{\boldsymbol{\xi}}_t^{\text{TE}}, \mathbf{x}_k^t) \right)^p}. \end{aligned}$$

In this expression, the trace terms do not simplify. \square

3.A.11 Proof of proposition 3.2.13 at p. 96

We first consider the following Lemma:

Lemma 3. Be $\boldsymbol{\xi} \in \mathbb{S}_{\mathbb{H}}^p$,

$\forall \mathbf{U} \in \mathbb{C}^{q,p}$ so that \mathbf{U} is a unitary matrix, $\mathbf{U} \boldsymbol{\xi} \mathbf{U}^H = \boldsymbol{\xi} \Leftrightarrow \exists \alpha \in \mathbb{R}, \boldsymbol{\xi} = \alpha \mathbf{I}_p$.

Proof.

- If $\boldsymbol{\xi} = \alpha \mathbf{I}_p$, $\forall \mathbf{U} \in \mathbb{C}^{q,p}$ so that \mathbf{U} is a unitary matrix, $\mathbf{U} \boldsymbol{\xi} \mathbf{U}^H = \mathbf{U} \alpha \mathbf{I}_p \mathbf{U}^H = \alpha \mathbf{I}_p = \boldsymbol{\xi}$.

- Since ξ is in $\mathbb{S}_{\mathbb{H}}^p$, $\exists \Delta \in \mathbb{D}^p \exists \mathbf{P} \in \mathbb{C}^{p,p}$, unitary so that: $\xi = \mathbf{P} \Delta \mathbf{P}^H$. By taking $\mathbf{U} = \mathbf{P}$, it follows that $\xi \in \mathbb{D}^p$.

Now take $\mathbf{U} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & & & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{bmatrix}$, which is a unitary matrix.

We have $\forall i \in \{1, \dots, p-1\}$, $\{\mathbf{U}\xi\mathbf{U}^H\}_{i,i} = \xi_{i+1,i+1}$ and $\{\mathbf{U}\xi\mathbf{U}^H\}_{p,p} = \xi_{1,1}$. Using $\mathbf{U}\xi\mathbf{U}^H = \xi$, it follows that $\forall (i,j) \in \{1, \dots, p\}^2$, $\xi_{i,i} = \xi_{j,j}$. In conclusion, $\exists \alpha \in \mathbb{R}$, $\xi = \alpha \mathbf{I}_p$.

□

Then we can consider the proposition at hand:

Proposition. $\hat{\xi}_0^{\text{MT}}$ and $\hat{\xi}_t^{\text{Tex}}$ are unbiased estimators of ξ_0 , the true covariance under null hypothesis.

Proof. We will consider here only the case of $\hat{\xi}_0^{\text{MT}}$ since the same analysis will lead to the result for $\hat{\xi}_t^{\text{Tex}}$.

Define $\forall k \forall t$, $\mathbf{y}_k^t = \xi_0^{-\frac{1}{2}} \frac{\mathbf{x}_k^t}{\tau_k^0}$, and $\hat{\Xi}_0^{\text{MT}} = \xi_0^{-\frac{1}{2}} \hat{\xi}_0^{\text{MT}} \xi_0^{-\frac{1}{2}}$, where ξ_0 is the true shape matrix and $\{\tau_k^0 | 1 \leq k \leq N\}$ are the true texture parameters. We have:

$$\hat{\Xi}_0^{\text{MT}} = \frac{p}{N} \sum_{k=1}^N \frac{\sum_{t=1}^T \mathbf{y}_k^t \mathbf{y}_k^t H}{\sum_{t=1}^T \mathbf{y}_k^t H \{\hat{\Xi}_0^{\text{MT}}\}^{-1} \mathbf{y}_k^t}$$

Let $\mathbf{U} \in \mathbb{C}^{q,p}$ so that \mathbf{U} is a unitary matrix, we define $\forall k \forall t$, $\zeta_k^t = \mathbf{U} \mathbf{y}_k^t$.

We have:

$$\mathbf{U} \hat{\Xi}_0^{\text{MT}} \mathbf{U}^H = \frac{p}{N} \sum_{k=1}^N \frac{\sum_{t=1}^T \zeta_k^t \zeta_k^t H}{\sum_{t=1}^T \zeta_k^t H \{\mathbf{U} \hat{\Xi}_0^{\text{MT}} \mathbf{U}^H\}^{-1} \zeta_k^t}$$

Since $\forall k \forall t$, $\mathbf{y}_k^t \sim \mathcal{CN}(\mathbf{0}_p, \mathbf{I}_p)$, we have: $\forall k \forall t$, $\zeta_k^t \sim \mathcal{CN}(\mathbf{0}_p, \mathbf{U} \mathbf{I}_p \mathbf{U}^H)$. And \mathbf{U} is unitary so $\forall k \forall t$, $\zeta_k^t \sim \mathcal{CN}(\mathbf{0}_p, \mathbf{I}_p)$. From this, we have that $\mathbf{U} \hat{\Xi}_0^{\text{MT}} \mathbf{U}^H$ follows the same distribution as $\hat{\Xi}_0^{\text{MT}}$ and thus:

$$\mathbb{E} \left\{ \mathbf{U} \hat{\Xi}_0^{\text{MT}} \mathbf{U}^H \right\} = \mathbf{U} \mathbb{E} \left\{ \hat{\Xi}_0^{\text{MT}} \right\} \mathbf{U} = \mathbb{E} \left\{ \hat{\Xi}_0^{\text{MT}} \right\}.$$

Using Lemma 3, it follows that $\exists \alpha \in \mathbb{R}$, $\mathbb{E} \left\{ \hat{\Xi}_0^{\text{MT}} \right\} = \alpha \mathbf{I}_p$.

Moreover, $\mathbb{E} \left\{ \hat{\Xi}_0^{\text{MT}} \right\} = \xi_0^{-\frac{1}{2}} \mathbb{E} \left\{ \Xi_0^{\text{MT}} \right\} \xi_0^{-\frac{1}{2}}$. From this, we can write $\mathbb{E} \left\{ \hat{\Xi}_0^{\text{MT}} \right\} = \alpha \xi_{(0)}$.

Finally we use the normalization constraint: $\text{Tr}(\xi_0^{-1} \hat{\Xi}_0^{\text{MT}}) = p$, which guarantees that $\alpha = 1$.

□

3.A.12 Proof of proposition 3.2.14 at p. 96

We first give a useful lemma to demonstrate our proposition:

Lemma 4. [Vaart, 1998, p. 46, Th. 5.9]

Let Ψ_N be random vector-valued functions and let Ψ a fixed valued vector of $\boldsymbol{\theta} \in \Theta$, such that for every $\epsilon > 0$:

$$\begin{aligned} (C1): \quad & \max_{\boldsymbol{\theta} \in \Theta} \|\Psi_N(\boldsymbol{\theta}) - \Psi(\boldsymbol{\theta})\| \xrightarrow[N \rightarrow +\infty]{\text{Pr}} 0, \\ (C2): \quad & \min_{d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) > \epsilon} \|\Psi(\boldsymbol{\theta})\| > 0 = \|\Psi(\boldsymbol{\theta}_0)\|. \end{aligned} \quad (3.63)$$

Then any sequence $\hat{\boldsymbol{\theta}}_N$ such that $\Psi_N(\hat{\boldsymbol{\theta}}_N) = o_P(1)$ converges in probability to $\boldsymbol{\theta}_0$.

Then, we have:

Proposition. $\hat{\xi}_0^{\text{MT}}$ and $\hat{\xi}_t^{\text{Tex}}$ are consistent estimators of ξ_0 , the true covariance under null hypothesis.

Proof. Again, we will consider here only the case of $\hat{\xi}_0^{\text{MT}}$ since the same analysis will lead to the result for $\hat{\xi}_t^{\text{MC}}$.

To show the property, we use the Lemma 4. In our situation,

$$\Theta = \mathbb{S}_{\mathbb{H}}^p, d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2, \hat{\boldsymbol{\theta}}_N = f_{1,N}(f_{1,N}), \boldsymbol{\theta}_0 = \xi_0, \\ \Psi_N(\xi) : \begin{cases} \mathbb{S}_{\mathbb{H}}^p \rightarrow \mathbb{S}_{\mathbb{H}}^p \\ \xi \rightarrow \xi - f_{1,N}(f_{1,N}) \end{cases}, \Psi(\xi) : \begin{cases} \mathbb{S}_{\mathbb{H}}^p \rightarrow \mathbb{S}_{\mathbb{H}}^p \\ \xi \rightarrow \xi - \mathbb{E}\{f_1(\xi)\} \end{cases},$$

$$f_{1,N} : \begin{cases} \mathbb{S}_{\mathbb{H}}^p \rightarrow \mathbb{S}_{\mathbb{H}}^p \\ \xi \rightarrow \frac{1}{N} \sum_{k=1}^N f_k(\xi) \end{cases}, \text{ where } f_k(\xi) = p \frac{\sum_{t=1}^T \mathbf{x}_k^t \mathbf{x}_k^{t \text{ H}}}{\sum_{t=1}^T \mathbf{x}_k^{t \text{ H}} \xi^{-1} \mathbf{x}_k^t}.$$

- By construction, $\Psi_N(\hat{\boldsymbol{\theta}}_N) = \mathbf{0}$. Thus, $\Psi_N(\hat{\boldsymbol{\theta}}_N) \xrightarrow[N \rightarrow +\infty]{\text{Pr}} \mathbf{0}$, which is the definition of $o_P(1)$.
- First (C1) is ensured using the Strong Law of Large Numbers (SLLN):
Since $f_{1,N}(\xi) \xrightarrow[N \rightarrow +\infty]{\text{a.s.}} \mathbb{E}\{f_1(\xi)\}$, $\forall \xi \in \mathbb{S}_{\mathbb{H}}^p$, $\|\Psi_N(\xi) - \Psi(\xi)\| \xrightarrow[N \rightarrow +\infty]{\text{Pr}} 0$.
- Then (C2) is ensured by using proposition 3.2.13 (unbiasedness):
We have $\mathbb{E}\{f_1(\hat{\xi}_0^{\text{MT}})\} = \xi_0$, which means that if $\xi \neq \xi_0$, $\Psi(\xi) \neq 0$ and thus $\|\Psi(\xi)\| > 0$.

Since all the conditions of the theorem are respected, we can conclude that $\hat{\xi}_0^{\text{MT}}$ is a consistent estimate of ξ_0 . \square

3.B DETECTION METHODS BASED ON STRUCTURED COVARIANCE MATRICES FOR MULTIVARIATE SAR IMAGES PROCESSING

Detection Methods Based on Structured Covariance Matrices for Multivariate SAR Images Processing

R. Ben Abdallah, A. Mian^{ID}, A. Breloy^{ID}, A. Taylor, M. N. El Korso, and D. Lautru

Abstract—Testing the similarity of covariance matrices (CMs) from groups of observations has been shown to be a relevant approach for change and/or anomaly detection in synthetic aperture radar images. Although the term “similarity” usually refers to equality or proportionality, we explore the testing of shared properties in the structure of low rank (LR) plus identity CM, which are appropriate for radar processing. Specifically, we derive two new generalized likelihood ratio tests to infer: 1) on the equality of the LR signal component of CMs and 2) on the proportionality of the LR signal component of CMs. The formulation of the second test involves nontrivial optimization problems for which we tailor efficient majorization–minimization algorithms. Eventually, the proposed detection methods enjoy interesting properties that are illustrated on simulations and on an application to real data for change detection.

Index Terms—Change detection, covariance testing, generalized likelihood ratio test (GLRT), low-rank (LR) structure, synthetic aperture radar (SAR).

I. INTRODUCTION

STATISTICAL testing of covariance matrix (CM) equality (or proportionality) has received increasing interest in the context of synthetic aperture radar (SAR) image processing. Indeed, this well-established hypothesis test has been successfully studied and applied to change/anomaly detection and classification in SAR images. Notably, equality testing has been proposed for change detection in SAR in [1]–[8]. A clear overview and statistical analysis of this topic is proposed in [9]. The extension to proportionality testing has been proposed in [10] and [11]. In this scope, elliptical noise modeling has also been studied to develop robust CM-based SAR image processing in [12]–[15].

In this letter, we propose new statistical tests in the context of structured CM. Indeed, the CM of radar measurements usually exhibits inherent structures. In a very general case, the samples can be modeled as a realization of a low-rank (LR) signal component plus white Gaussian (thermal) noise. This leads to a CM structured as $\Sigma = \Sigma_R + \sigma^2 \mathbf{I}$, where Σ_R is the

Manuscript received August 28, 2018; revised November 11, 2018 and December 14, 2018; accepted December 22, 2018. This work was supported by ANR-ASTRID MARGARITA under Grant ANR-17-ASTR-0015. (*Corresponding author: A. Mian*)

R. Ben Abdallah, A. Breloy, M. N. El Korso, and D. Lautru are with LEME (EA4416), University Paris Nanterre, 92000 Nanterre, France.

A. Mian is with SONDRA, CentraleSupélec, Université Paris-Saclay, 91190 Gif-sur-Yvette, France (e-mail: ammar.mian@centralesupelec.fr).

A. Taylor is with DEMR, ONERA, Université Paris Saclay, 91123 Palaiseau, France.

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author.

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

Digital Object Identifier 10.1109/LGRS.2018.2890155

LR signal CM. Taking this prior knowledge in the detection process offers several advantages.

- 1) Introducing relevant prior information (here, the LR CM structure) in the model improves detection performances.
- 2) LR-structured models allow dealing with low-sample-support issues since fewer samples are required to estimate the CM. In particular, it allows performing tests even when the sample CM (SCM) is not invertible, while this condition is restrictive for standard tests.
- 3) The considered formulation can go beyond equality and proportionality testing. For example, consider local power fluctuations of ground response modeled as $\Sigma_i = \tau_i \Sigma_R + \sigma^2 \mathbf{I}$, where $\tau_i \in \mathbb{R}^+$ and i denotes the index of a homogeneous set. Such model leads to $\Sigma_i \neq \Sigma_j$ as well as $\Sigma_i \not\propto \Sigma_j$ for $i \neq j$. If the goal is to detect a signal anomaly in Σ_R , detectors based on CM equality/proportionality testing may lead to an excessive number of false alarms.

Specifically, we propose two novel generalized likelihood ratio tests (GLRTs) that account for the considered LR structure, with assumed known rank. Formally, these detectors reexpress the equality and proportionality tests, but only on the signal LR component Σ_R of the total CM. The derivation of the second test requires solving some nontrivial optimization problems, for which we tailor appropriate majorization–minimization (MM) algorithms in the Supplementary material attached to this letter. It is worth mentioning that the proposed formulations and optimization methods can also be adapted to design various other tests—such as eigenvectors (or principal subspace) equality testing—for other specific applications.

Finally, the performance of the proposed detectors is illustrated on the simulated data, where they exhibit interesting properties. Furthermore, the benefits of the proposed methods are also illustrated for a change detection application on a UAVSAR data set (courtesy of NASA/JPL-Caltech, <https://uavstar.jpl.nasa.gov>). For this application, our conclusions are as the following.

- 1) Incorporating the LR structure in CM equality testing offers an improvement of the detection performance (especially for small local windows) with a small increase in the computational cost.
- 2) Testing the CM proportionality (either with LR or full rank model) requires more computational time, which does not appear to be beneficial to obtain a high probability detection (PD). This is due to the fact that these detectors are designed to be insensitive to power fluctuations, while this phenomenon is relevant in change detection. However, these detectors are still interesting

for a number of other purposes, such as ensuring a low probability of false alarm (PFA) in local anomaly detection [10].

- 3) Both of the LR methods allow increasing the spatial resolution of the detection process, as they are defined for lower sample support.

Notations: Italic type indicates a scalar quantity, lower case boldface indicates a vector quantity, and upper case boldface a matrix. The transpose conjugate operator is H . $\text{Tr}\{\cdot\}$ and $|\cdot|$ are, respectively, the trace and the determinant operators. $\text{etr}\{\cdot\}$ is the exponential of trace operator. $\{w_n\}_{n \in \llbracket 1, N \rrbracket}$ denotes the set of elements w_n , with $n \in \llbracket 1, N \rrbracket$, often contracted in $\{w_n\}$. Definition of the needed eigenvalue decomposition will be through the equality symbol $\stackrel{\text{EVD}}{=}$. \mathcal{H}_M^{++} denotes the set of $M \times M$ Hermitian positive definite matrices. \propto stands for “proportional to.” $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a complex-valued random Gaussian vector of mean $\boldsymbol{\mu}$ and CM $\boldsymbol{\Sigma}$. $x \sim \Gamma(v, \xi)$ is a random variable following a Gamma distribution of shape v and scale ξ .

II. MODEL AND PROBLEM STATEMENT

A. Signal Model

In the following, \mathbf{z}_k^i denotes a sample, in which the superscript $i \in \llbracket 0, I \rrbracket$ refers to the index of a set of independent identically distributed variables, and $k \in \llbracket 1, K_i \rrbracket$ to the index of the sample in this set (of size K_i). Depending on the context, i may stand either for the index of a local patch or for the index of a time series. For a given sample set $\{\mathbf{z}_k^i\}$ of multivariate pixels, we consider the following data model:

$$\mathbf{z}_k^i = \mathbf{s}_k^i + \mathbf{n}_k^i \quad (1)$$

where:

- 1) \mathbf{s}_k^i is the ground response, which consists of a mixture of LR signal contributions. The resulting observation is modeled as $\mathbf{s}_k^i \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_i)$ with unknown LR CM \mathbf{C}_i . As commonly assumed in the literature [16], [17], the rank R is considered known, or already preestablished¹
- 2) $\mathbf{n}_k^i \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$ is the thermal noise of known variance σ^2 . The extension of proposed algorithms to unknown σ^2 is trivial and is tested on real data in Section V.

Consequently, the samples are distributed as $\mathbf{z}_k^i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ where the total CM has a LR plus identity structure. To reflect this structure, we consider the following parameterization:

$$\boldsymbol{\Sigma}_i = \tau_i \mathbf{V}_i \boldsymbol{\Lambda}_i \mathbf{V}_i^H + \sigma^2 \mathbf{I} \stackrel{\Delta}{=} \tau_i \boldsymbol{\Sigma}_R^i + \sigma^2 \mathbf{I} \quad (2)$$

where τ_i is a positive scaling factor, \mathbf{V}_i is a $M \times R$ unitary matrix, and $\boldsymbol{\Lambda}_i$ is an $R \times R$ positive diagonal matrix. Eventually, the likelihood of the data set is

$$\mathcal{L}(\{\mathbf{z}_k^i\} | \boldsymbol{\theta}) = \prod_{i=0}^I \frac{\text{etr}\{-\mathbf{S}_i \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta})\}}{|\boldsymbol{\Sigma}_i(\boldsymbol{\theta})|^{K_i}} \quad (3)$$

¹ Indeed, the proposed results can still be applied using plug-in rank estimates or by integrating physical prior knowledge on this parameter [18]. About rank estimation, the reader is referred to the overview [19] and recent methods using shrinkage [20] or random matrix theory [21].

with $\mathbf{S}_i = \sum_{k=1}^{K_i} \mathbf{z}_k^i \mathbf{z}_k^i H$ and where $\boldsymbol{\theta}$ denotes the set of parameters defining the $\boldsymbol{\Sigma}_i$ (specified in the following). Note that this model generalizes the LR compound Gaussian plus white Gaussian noise distribution, which is a realistic model for radar measurements embedded in thermal noise [17], [22]. The latter corresponds to the special case $K_i = 1, \forall i \in \llbracket 0, I \rrbracket$ in our setting.

B. Problem Statement

For the general model in (2) and (3), we turn to the problem of testing whether the CM of the sample set under test $i = 0$ shares some common properties with the secondary sets $i \in \llbracket 1, I \rrbracket$. These properties are related to the parameters of the decomposition in (2) (i.e., τ_i , \mathbf{V}_i , and $\boldsymbol{\Lambda}_i$) and will be specified depending on the proposed test. This problem is relevant to detect, e.g., a local anomaly in the patch with respect to adjacent patches, or a temporal change in the last sample of a time series.

III. STATE OF THE ART: EXISTING GLRTs

A. Equality Testing

The standard hypothesis test [9] reads

$$\begin{cases} \mathcal{H}_0: \boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}, & \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} \quad \forall i \in \llbracket 1, I \rrbracket \\ \mathcal{H}_1: \boldsymbol{\Sigma}_0 \neq \boldsymbol{\Sigma}, & \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} \quad \forall i \in \llbracket 1, I \rrbracket. \end{cases} \quad (4)$$

The GLRT for this hypothesis test, denoted t_{glr}^E , reads

$$|\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_0}| / (|\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_1}^0|^{\rho_0} |\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_1}^*|^{\rho_*}) \stackrel{\mathcal{H}_1}{\gtrless}_{\mathcal{H}_0} \delta_{\text{glr}}^E \quad (5)$$

with the quantities $K = \sum_{i=0}^I K_i$, $K_* = K - K_0$, the ratios $\rho_0 = K_0/K$, $\rho_* = K_*/K$, and the SCMs $\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_0} = \sum_{i=0}^I \mathbf{S}_i / K$, $\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_1}^0 = \mathbf{S}_0 / K_0$ and $\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_1}^* = \sum_{i=1}^I \mathbf{S}_i / K_*$.

B. Proportionality Testing

The classical hypothesis test [10] is

$$\begin{cases} \mathcal{H}_0: \boldsymbol{\Sigma}_0 = \beta_0 \boldsymbol{\Sigma}, & \boldsymbol{\Sigma}_i = \beta_i \boldsymbol{\Sigma} \quad \forall i \in \llbracket 1, I \rrbracket \\ \mathcal{H}_1: \boldsymbol{\Sigma}_0 \neq \beta_0 \boldsymbol{\Sigma}, & \boldsymbol{\Sigma}_i = \beta_i \boldsymbol{\Sigma} \quad \forall i \in \llbracket 1, I \rrbracket. \end{cases} \quad (6)$$

The GLRT on proportionality, denoted by t_{glr}^P , is given as

$$\left(\frac{|\hat{\beta}_{\mathcal{H}_0}^0 \hat{\boldsymbol{\Sigma}}_{\mathcal{H}_0}^{\text{gfp}}|}{|\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_1}^0|} \right)^{K_0} \prod_{i=1}^I \left(\frac{|\hat{\beta}_{\mathcal{H}_0}^i \hat{\boldsymbol{\Sigma}}_{\mathcal{H}_0}^{\text{gfp}}|}{|\hat{\beta}_{\mathcal{H}_1}^i \hat{\boldsymbol{\Sigma}}_{\mathcal{H}_1}^{\text{gfp}}|} \right)^{K_i} \stackrel{\mathcal{H}_1}{\gtrless}_{\mathcal{H}_0} \delta_{\text{glr}}^P \quad (7)$$

where $\{\hat{\beta}_{\mathcal{H}_0}^i\}$ and $\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_0}^{\text{gfp}}$ are the proportionality coefficients and shape matrix estimated with the generalized fixed point estimator (GFPE) [11] applied on the set $\{\mathbf{S}_i\}_{i \in \llbracket 0, I \rrbracket}$, $\{\hat{\beta}_{\mathcal{H}_1}^i\}$ and $\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_1}^{\text{gfp}}$ are obtained from the GFPE on the set $\{\mathbf{S}_i\}_{i \in \llbracket 1, I \rrbracket}$, and where $\hat{\boldsymbol{\Sigma}}_{\mathcal{H}_1}^0$ is the SCM defined above.

IV. PROPOSED DETECTORS

A. GLRT for LR CM Equality Testing

In this section, we develop a GLRT that is sensitive to a variation of any parameter of the LR signal CM in the set $i = 0$. Thus, for the CM model in (2), consider the following hypothesis test:

$$\begin{cases} \mathcal{H}_0: \tau_i \boldsymbol{\Sigma}_R^i = \boldsymbol{\Sigma}_R & \forall i \in \llbracket 0, I \rrbracket \\ \mathcal{H}_1: \tau_i \boldsymbol{\Sigma}_R^i = \boldsymbol{\Sigma}_R & \forall i \in \llbracket 1, I \rrbracket \\ \tau_0 \boldsymbol{\Sigma}_R^0 \neq \boldsymbol{\Sigma}_R \end{cases} \quad (8)$$

that reads almost identical to the standard equality testing of Section III, except that the CMs belong to the set of LR plus identity structured matrices $\mathcal{S}_{\text{LR}}^{++}$. Hence, the hypothesis test can be recasted as

$$\begin{cases} \mathcal{H}_0 : & \Sigma_i = \Sigma_{R, \mathcal{H}_0} + \sigma^2 \mathbf{I} \stackrel{\Delta}{=} \Sigma_{\mathcal{H}_0} \in \mathcal{S}_{\text{LR}}^{++} \quad \forall i \in \llbracket 0, I \rrbracket \\ \mathcal{H}_1 : & \begin{cases} \Sigma_0 = \Sigma_{R, \mathcal{H}_1}^0 + \sigma^2 \mathbf{I} \stackrel{\Delta}{=} \Sigma_{\mathcal{H}_1}^0 \in \mathcal{S}_{\text{LR}}^{++} \\ \Sigma_i = \Sigma_{R, \mathcal{H}_1}^* + \sigma^2 \mathbf{I} \stackrel{\Delta}{=} \Sigma_{\mathcal{H}_1}^* \in \mathcal{S}_{\text{LR}}^{++} \quad \forall i \in \llbracket 1, I \rrbracket \end{cases} \end{cases}$$

The expression of the GLRT is therefore

$$\frac{\max_{\theta_{\mathcal{H}_1}^{\text{lrE}}} \mathcal{L}(\{\mathbf{z}_k^i\} | \mathcal{H}_1, \theta_{\mathcal{H}_1}^{\text{lrE}})}{\max_{\theta_{\mathcal{H}_0}^{\text{lrE}}} \mathcal{L}(\{\mathbf{z}_k^i\} | \mathcal{H}_0, \theta_{\mathcal{H}_0}^{\text{lrE}})} \stackrel{\mathcal{H}_1}{\gtrless} \delta_{\text{glr}}^{\text{lrE}} \quad (9)$$

with sets $\theta_{\mathcal{H}_1}^{\text{lrE}} = \{\Sigma_{\mathcal{H}_1}^0, \Sigma_{\mathcal{H}_1}^*\}$ and $\theta_{\mathcal{H}_0}^{\text{lrE}} = \{\Sigma_{\mathcal{H}_0}\}$. In the context of Gaussian data, the maximum-likelihood estimator (MLE) of LR-structured CM is obtained by thresholding the eigenvalues of the SCM [16] with the operator \mathcal{T}_R . This operator associates to any Hermitian matrix $\Sigma \stackrel{\text{EVD}}{=} \mathbf{V} \Lambda \mathbf{V}^H$ the regularization $\mathcal{T}_R\{\Sigma\} \stackrel{\text{EVD}}{=} \mathbf{V} \tilde{\Lambda} \mathbf{V}^H$ with

$$[\tilde{\Lambda}]_{i,i} = \begin{cases} \max([\Lambda]_{i,i}, \sigma^2), & i \leq R \\ \sigma^2, & i > R. \end{cases} \quad (10)$$

Therefore, it can be easily shown that

$$\hat{\theta}_{\mathcal{H}_0}^{\text{lrE}} = \{\mathcal{T}_R\{\hat{\Sigma}_{\mathcal{H}_0}\}\} \text{ and } \hat{\theta}_{\mathcal{H}_1}^{\text{lrE}} = \{\mathcal{T}_R\{\hat{\Sigma}_{\mathcal{H}_1}^0\}, \mathcal{T}_R\{\hat{\Sigma}_{\mathcal{H}_1}^*\}\}$$

with $\hat{\Sigma}_{\mathcal{H}_0}$, $\hat{\Sigma}_{\mathcal{H}_1}^0$, and $\hat{\Sigma}_{\mathcal{H}_1}^*$ defined in III. Finally, the GLRT for testing the equality of LR structured matrices, denoted $t_{\text{glr}}^{\text{lrE}}$, reads as

$$\mathcal{L}(\{\mathbf{z}_k^i\} | \mathcal{H}_1, \hat{\theta}_{\mathcal{H}_1}^{\text{lrE}}) / \mathcal{L}(\{\mathbf{z}_k^i\} | \mathcal{H}_0, \hat{\theta}_{\mathcal{H}_0}^{\text{lrE}}) \stackrel{\mathcal{H}_1}{\gtrless} \delta_{\text{glr}}^{\text{lrE}}. \quad (11)$$

To evaluate this test, three singular value decompositions (SVDs) of SCMs are required. In comparison, $t_{\text{glr}}^{\text{E}}$ requires to compute the determinant of the same three SCMs. The proposed $t_{\text{glr}}^{\text{lrE}}$ is, therefore, slightly computationally more expensive.

B. GLRT for LR CM Proportionality Testing

In this section, we derive a GLRT to infer on the proportionality of the LR signal component of the CM. Note that this test differs from strict proportionality testing. Indeed, scaling fluctuations should only apply on the LR part of the signal CM, and not to the identity, related to the thermal noise. For the CM model in (2), this leads to the following hypothesis test:

$$\begin{cases} \mathcal{H}_0 : & \Sigma_R^i = \Sigma_R \quad \forall i \in \llbracket 0, I \rrbracket \\ \mathcal{H}_1 : & \begin{cases} \Sigma_R^i = \Sigma_R \quad \forall i \in \llbracket 1, I \rrbracket \\ \Sigma_R^0 \neq \Sigma_R \end{cases} \end{cases} \quad (12)$$

which reads as signals sharing the same CM structure, but with fluctuating power τ_i with respect to sample set i . The test (12) can be recasted as

$$\begin{cases} \mathcal{H}_0 : & \Sigma_i = \tau_i \mathbf{V}_{\mathcal{H}_0} \Lambda_{\mathcal{H}_0} (\mathbf{V}_{\mathcal{H}_0})^H + \sigma^2 \mathbf{I} \quad \forall i \in \llbracket 0, I \rrbracket \\ \mathcal{H}_1 : & \begin{cases} \Sigma_0 = \mathbf{V}_{\mathcal{H}_1}^0 \mathbf{A}_{\mathcal{H}_1}^0 (\mathbf{V}_{\mathcal{H}_1}^0)^H + \sigma^2 \mathbf{I} \\ \Sigma_i = \tau_i \mathbf{V}_{\mathcal{H}_1}^* \Lambda_{\mathcal{H}_1}^* (\mathbf{V}_{\mathcal{H}_1}^*)^H + \sigma^2 \mathbf{I} \quad \forall i \in \llbracket 1, I \rrbracket \end{cases} \end{cases} \quad (13)$$

where we collapsed a redundant parameter as $\mathbf{A}_0 = \tau_0 \Lambda_0$ (that is still diagonal). The expression of the corresponding GLRT, denoted $t_{\text{glr}}^{\text{lrP}}$ is given by

$$\frac{\max_{\theta_{\mathcal{H}_1}^{\text{lrP}}} \mathcal{L}(\{\mathbf{z}_k^i\} | \mathcal{H}_1, \theta_{\mathcal{H}_1}^{\text{lrP}})}{\max_{\theta_{\mathcal{H}_0}^{\text{lrP}}} \mathcal{L}(\{\mathbf{z}_k^i\} | \mathcal{H}_0, \theta_{\mathcal{H}_0}^{\text{lrP}})} \stackrel{\mathcal{H}_1}{\gtrless} \delta_{\text{glr}}^{\text{lrP}} \quad (14)$$

with sets of parameters $\theta_{\mathcal{H}_0}^{\text{lrP}} = \{\{\tau_i\}_{i \in \llbracket 0, I \rrbracket}, \mathbf{V}_{\mathcal{H}_0}, \Lambda_{\mathcal{H}_0}\}$ and $\theta_{\mathcal{H}_1}^{\text{lrP}} = \{\{\tau_i\}_{i \in \llbracket 1, I \rrbracket}, \mathbf{V}_{\mathcal{H}_1}^*, \Lambda_{\mathcal{H}_1}^*, \mathbf{V}_{\mathcal{H}_1}^0, \mathbf{A}_{\mathcal{H}_1}^0\}$. It is clear that the maximization of the likelihood function is not trivial, due notably to the unitary constraints on the eigenvectors. To overcome this issue, we propose the use of the block-MM algorithm [23]. This methodology can be applied to our problem by generalizing some results of [22]. Due to space constraints, the full derivation of the proposed algorithms is left in the Supplementary material attached to this letter. Eventually, these algorithms allow to evaluate the MLEs $\hat{\theta}_{\mathcal{H}_0}^{\text{lrP}}$ and $\hat{\theta}_{\mathcal{H}_1}^{\text{lrP}}$ and the GLRT as

$$\mathcal{L}(\{\mathbf{z}_k^i\} | \mathcal{H}_1, \hat{\theta}_{\mathcal{H}_1}^{\text{lrP}}) / \mathcal{L}(\{\mathbf{z}_k^i\} | \mathcal{H}_0, \hat{\theta}_{\mathcal{H}_0}^{\text{lrP}}) \stackrel{\mathcal{H}_1}{\gtrless} \delta_{\text{glr}}^{\text{lrP}}. \quad (15)$$

In terms of computational cost, it is noted that each update of our proposed MM algorithm is obtained in closed form. Empirically, this algorithm converges quite fast, and we used only 10 iterations in our application to the real data. The main bottleneck is in the update of the eigenvectors, which requires to compute thin-SVD of an $M \times R$ matrix. Thus, this test is more computationally expensive than $t_{\text{glr}}^{\text{E}}$ or $t_{\text{glr}}^{\text{lrE}}$. However, it is on the same scale as its full rank counterpart $t_{\text{glr}}^{\text{P}}$, which involves fixed-point iterations of SCMs inversions.

V. SIMULATIONS AND APPLICATION

A. Numerical Simulations

In this section, the performance of the aforementioned detectors is illustrated through simulations. We use as a criterion the receiver operating characteristic (ROC) curve which displays the PD versus the PFA.

1) Simulation Setup: We consider $M = 20, R = 5, K_i = 25 \forall i$ with $I = 3$. The sample set $\{\mathbf{z}_k^i\}$ is generated according to $\mathbf{z}_k^i \sim \mathcal{CN}(\mathbf{0}, \Sigma_i)$ with Σ_i given in (2): \mathbf{V}_i are the first R eigenvectors of the Toeplitz matrix $[\Sigma_T]_{i,j} = \rho^{|i-j|}$ with $\rho = 0.9(1 + \sqrt{-1})/\sqrt{2}$ and $[\Lambda_i]_{r,r} = \alpha(R + 1 - r)$, where α is set so that the signal to noise ratio fits $\text{SNR} = \text{Tr}\{\Lambda_i\}/R\sigma^2 = 15\text{dB}$ with $\sigma^2 = 1$. The variable τ_i is specified in the following for each scenario. Under \mathcal{H}_1 , the anomaly in the LR signal CM of the set $i = 0$ is generated by either: 1) reversing the eigenvalues, i.e., $[\Lambda_0]_{r,r} = \alpha r$ ("structure change") or 2) changing one of the eigenvectors in \mathbf{V}_i ("subspace change"). Note that the structure change is more challenging since a subspace change is easier to detect at high SNR. In order to compute the ROC curves, 10^4 Monte-Carlo runs are performed under both \mathcal{H}_0 and \mathcal{H}_1 and the PD and PFA are computed with respect to a threshold grid for each detector.

2) Compared Detectors: We compare the following detection statistics: the GLRT for equality testing $t_{\text{glr}}^{\text{E}}$ from Section III-A, the GLRT for proportionality testing $t_{\text{glr}}^{\text{P}}$ from Section III-B, the proposed GLRT for LR equality testing

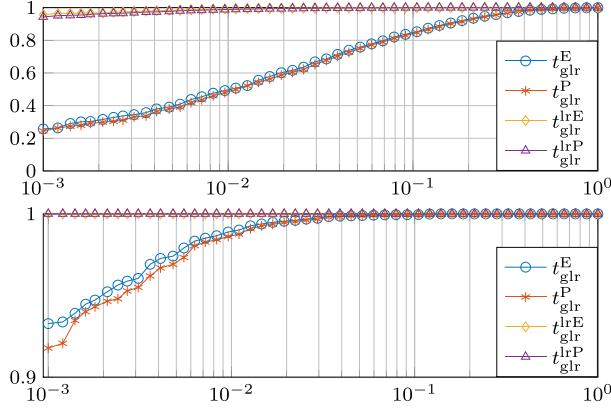


Fig. 1. ROC curves for the model discussed in Section IV-A: \mathcal{H}_1 is either a (Top) structure change or a (Bottom) subspace change.

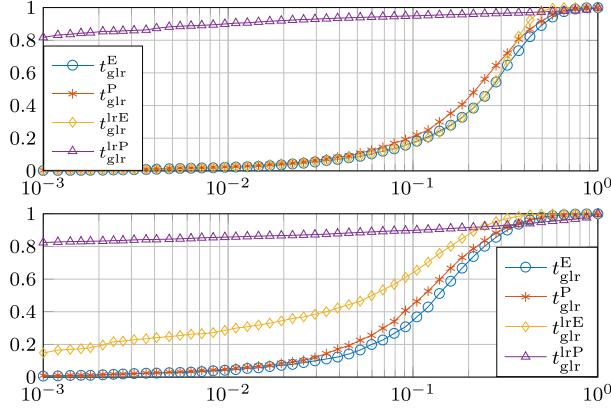


Fig. 2. ROC curves for the model discussed in Section IV-B. \mathcal{H}_1 is either a (Top) structure change or a (Bottom) subspace change.

$t_{\text{glr}}^{\text{lrE}}$ from Section IV-A, and the proposed GLRT for LR proportionality testing $t_{\text{glr}}^{\text{lrP}}$ from Section IV-B.

3) *Results:* Fig. 1 displays the ROC curves of the detectors for the signal model from Section IV-A (homogeneous power), i.e., $\tau_i = 1 \forall i$. Under this settings, the performances of $t_{\text{glr}}^{\text{E}}$ and $t_{\text{glr}}^{\text{P}}$ are almost identical. The same observation is made for $t_{\text{glr}}^{\text{lrE}}$ and $t_{\text{glr}}^{\text{lrP}}$, which both outperform their full-rank counterparts since they exploit the LR structure of the CM appropriately. Fig. 2 displays the ROC curves of the detectors for the signal model from Section IV-B (fluctuating power) with $\tau_i \sim \Gamma(\nu, 1/\nu)$ and $\nu = 1$. Under these settings, $t_{\text{glr}}^{\text{lrP}}$ exhibits the best performance, as expected. This is mainly due to a high false alarm rate of the other detectors, as the fluctuation of the signal power generates CMs that are not equal, nor proportional, even under \mathcal{H}_0 . Hence, $t_{\text{glr}}^{\text{lrP}}$ appears interesting for reducing the false alarms rate when the signal has a varying power over the observed sets.

B. Change Detection on Real Data

In this section, the performance of the proposed detectors is illustrated for change detection on a UAVSAR data set.

1) *Setup:* The considered data set is SanAnd_26524_03 Segment 4, of coordinates (top left pixel) [2891, 28 891], with two acquisition dates: April 23, 2009 and May 11, 2015. For one acquisition, the initial datacube size is $2360 \times 600 \times 3$ and is preprocessed using

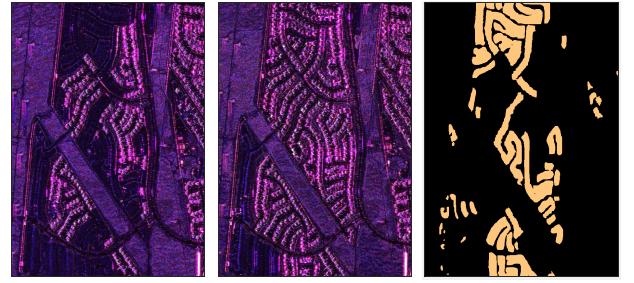


Fig. 3. UAVSAR data set in Pauli representation. (Left) April 23, 2009. (Middle) May 15, 2011. (Right) Ground truth for change detection.

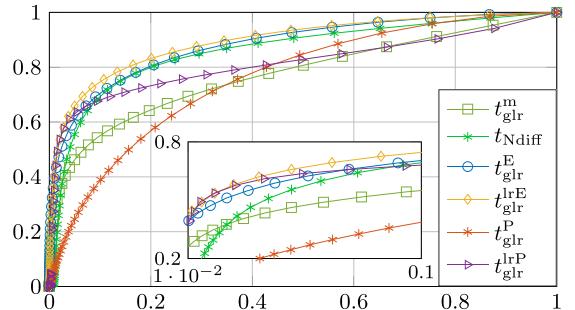


Fig. 4. ROC curves (PD versus PFA) of the different detectors on the UAVSAR data set.

the wavelet-decomposition transform presented in [24]. This transformation, which allows decomposing a SAR image into canals corresponding to a physical behavior of the scatterers, has been shown to increase the detection performance [24]. This transformation increases the depth of the datacube from $M = 3$ to $M = 12$. To form local patches, we use a 5×5 sliding window centered around each pixel. As $I = 2$, these patches provide two sets: $\{\mathbf{z}_k^0\}$ and $\{\mathbf{z}_k^1\}$ with $k \in [1, 25]$. The different presented detectors are then applied on these two sets to test a change in the properties/parameters of the CM between $i = 0$ and $i = 1$.

The ground truth for change detection is taken from [25] and presented in Fig. 3. This provides observations under both \mathcal{H}_0 and \mathcal{H}_1 , which allows us to compute the ROC curves empirically.

2) *Compared Detectors:* We compare the same detectors as in Section V-A.2. The proposed LR detection methods are applied with $R = 1$, as a rank one signal component can be assumed by analyzing the spectrum of the data matrix. This simplification still allows obtaining interesting results in average, and the use of local adaptive rank selection is left as a potential extension. The noise variance σ^2 is estimated locally with the mean of the $(M - 1)$ lowest eigenvalues computed with an SVD of the SCM of all samples $\{\mathbf{z}_k^i\}$ in the patch. To show the benefits of the multivariate setting, we also compare the results with a monovariate detector applied to the summed entries of each pixel, denoted t_m , as well as the so-called normalized mean-differences detector, denoted t_{Ndiff} [26].

3) *Results:* The ROC curves of the different detectors are displayed in Fig. 4. In this example, $t_{\text{glr}}^{\text{lrE}}$ offers an improvement of the detection performance compared to the standard equality testing $t_{\text{glr}}^{\text{E}}$. This improvement is obtained for only a slight increase in the computational time, highlighting the interest of

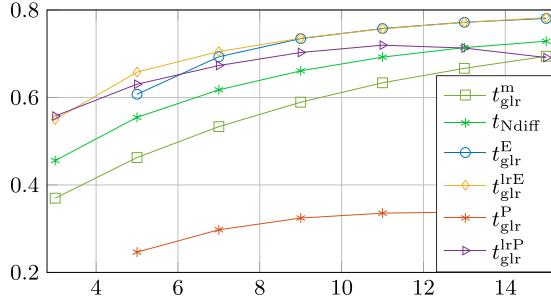


Fig. 5. PD versus spatial window size \sqrt{K} for fixed PFA = 5% of the different detectors on the UAVSAR data set.

the proposed LR formulation. $t_{\text{glr}}^{\text{lrP}}$ offers similar performance, but only for low false alarm rate. Thus, for this application, this test may not be worth the increase in the computational time. This is also the case for its counterpart $t_{\text{glr}}^{\text{P}}$, which exhibits the lowest performance in this context. Intuitively, a power fluctuation seems interesting to be captured when it comes to change detection. Therefore, proportionality testing (not sensitive to this change) may not be the most appropriate for this application, as illustrated by the performance of $t_{\text{glr}}^{\text{P}}$. However, this detector is still interesting for other purposes, such as local anomaly detection [10].

For each detector, Fig. 5 displays the PD with respect to the spatial window size \sqrt{K} with fixed PFA = 5%. Up to a reasonable window size, the detection performance increases with respect to K . However, this is at the detriment of the spatial resolution of the process. This figure, hence, illustrates the interest of the proposed LR methods, as they offer an improvement of the performance/resolution tradeoff. Notably, the proposed methods allow for $K < M$, where other standard covariance-based detectors are not defined (due to noninvertible SCMs).

VI. CONCLUSION

This letter proposed two new detectors for CM-based detection process. These detectors extend, respectively, the equality and proportionality testing to LR-structured CM models, with a mild increase in the computational cost with respect to their corresponding full-rank counterparts. Numerical simulations illustrated their properties and interest depending on the context. An application to real data for change detection in SAR images time-series showed the interest of the LR approach. Specifically, the LR equality testing offers a gain in detection performance, while the LR proportionality testing does, but only for low PFA. Most notably, both of the proposed LR methods allow increasing the spatial resolution of the detection process, as they require fewer samples than the size of the data.

REFERENCES

- [1] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 1, pp. 4–19, Jan. 2003.
- [2] L. M. Novak, "Change detection for multi-polarization multi-pass SAR (invited paper)," *Proc. SPIE*, vol. 5808, no. 1, pp. 234–247, May 2005.
- [3] S. W. Chen, X. S. Wang, and M. Sato, "PolInSAR complex coherence estimation based on covariance matrix similarity test," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4699–4710, Nov. 2012.
- [4] M. Liu, H. Zhang, C. Wang, and F. Wu, "Change detection of multilook polarimetric SAR images using heterogeneous clutter models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7483–7494, Dec. 2014.
- [5] V. Carotenuto, A. De Maio, C. Clemente, and J. Soraghan, "Unstructured versus structured GLRT for multipolarization SAR change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1665–1669, Aug. 2015.
- [6] A. A. Nielsen, K. Conradsen, and H. Skriver, "Change detection in full and dual polarization, single- and multifrequency SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 4041–4048, Aug. 2015.
- [7] A. A. Nielsen, K. Conradsen, and H. Skriver, "Omnibus test for change detection in a time sequence of polarimetric SAR data," *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 3398–3401.
- [8] A. De Maio, D. Orlando, L. Pallotta, and C. Clemente, "A multifamily GLRT for oil spill detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 63–79, Jan. 2017.
- [9] D. Ciuonzo, V. Carotenuto, and A. De Maio, "On multiple covariance equality testing with application to SAR change detection," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 5078–5091, Oct. 2017.
- [10] A. Taylor, H. Orlot, P. Forster, F. Brigui, L. Savy, and F. Daout, "Reducing false alarm rate by testing proportionality of covariance matrices," in *Proc. Int. RADAR Conf.*, 2017, pp. 1–4.
- [11] A. Taylor, P. Forster, F. Daout, H. M. Orlot, and L. Savy, "A generalization of the fixed point estimate for packet-scaled complex covariance matrix estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5393–5405, Oct. 2017.
- [12] G. Vasile, J.-P. Ovarlez, F. Pascal, and C. Tison, "Coherency matrix estimation of heterogeneous clutter in high-resolution polarimetric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1809–1826, Apr. 2010.
- [13] P. Formont, F. Pascal, G. Vasile, J.-P. Ovarlez, and L. Ferro-Famil, "Statistical classification for heterogeneous polarimetric SAR images," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 567–576, Jun. 2011.
- [14] M. Liu, H. Zhang, and C. Wang, "Change detection in urban areas of high-resolution polarization SAR images using heterogeneous clutter models," in *Proc. 3rd Int. Asia-Pacific Conf. Synth. Aperture Radar (APSAR)*, 2011, pp. 1–4.
- [15] A. Mian, J.-P. Ovarlez, G. Ginolhac, and A. M. Atto, "A robust change detector for highly heterogeneous multivariate images," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2018, pp. 3429–3433.
- [16] B. Kang, V. Monga, and M. Rangaswamy, "Rank-constrained maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 501–515, Jan. 2014.
- [17] O. Besson, "Bounds for a mixture of low-rank compound-Gaussian and white Gaussian noises," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5723–5732, Nov. 2016.
- [18] N. A. Goodman and J. M. Stiles, "On clutter rank observed by arbitrary arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 178–186, Jan. 2007.
- [19] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
- [20] L. Huang and H. C. So, "Source enumeration via MDL criterion based on linear shrinkage estimation of noise subspace covariance matrix," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4806–4821, Oct. 2013.
- [21] E. Terreux, J.-P. Ovarlez, and F. Pascal. (Oct. 2017). "Robust model order selection in large dimensional elliptically symmetric noise." [Online]. Available: <https://arxiv.org/abs/1710.06735>
- [22] Y. Sun, A. Breloy, P. Babu, D. P. Palomar, F. Pascal, and G. Ginolhac, "Low-complexity algorithms for low rank clutter parameters estimation in radar systems," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1986–1998, Apr. 2016.
- [23] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.
- [24] A. Mian, J. P. Ovarlez, G. Ginolhac, and A. Atto, "Multivariate change detection on high resolution monovariate SAR image using linear time-frequency analysis," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, 2017, pp. 1942–1946.
- [25] D. Ratha, S. De, T. Celik, and A. Bhattacharya, "Change detection in polarimetric SAR images using a geodesic distance between scattering mechanisms," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1066–1070, Jul. 2017.
- [26] P. Villa and G. Lechi, "Normalized difference reflectance: An approach to quantitative change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2007, pp. 2366–2369.

3.C. ROBUST LOW-RANK CHANGE DETECTION FOR SAR IMAGE TIME SERIES

3.C ROBUST LOW-RANK CHANGE DETECTION FOR SAR IMAGE TIME SERIES

ROBUST LOW-RANK CHANGE DETECTION FOR SAR IMAGE TIME SERIES

A. Mian^{*,†}, A. Breloy[‡], G. Ginolhac^{*,†} and J.-P. Ovarlez^{*,¶}

* SONDRA, CentraleSupélec, Université Paris-Saclay, F-91190, Gif-sur-Yvette, France

† LISTIC, Université Savoie Mont-Blanc, F-74944, Annecy le Vieux, France

‡ LEME, Université Paris-Nanterre, IUT Ville d'Avray, LEME-EA4416, France

¶ DEMR, ONERA, Université Paris-Saclay, F-91123, Palaiseau, France

ABSTRACT

This paper considers the problem of detecting changes in multivariate Synthetic Aperture Radar image time series. Classical methodologies based on covariance matrix analysis are usually built upon the Gaussian assumption, as well as an unstructured signal model. Both of these hypotheses may be inaccurate for high-dimension/resolution images, where the noise can be heterogeneous (non-Gaussian) and where all channels are not always informative (low-rank structure). In this paper, we tackle these two issues by proposing a new detector assuming a robust low-rank model. Analysis of the proposed method on a UAVSAR dataset shows promising results.

Index Terms— Change detection; Synthetic aperture Radar; Low Rank; Compound Gaussian;

1. INTRODUCTION

Analysis of Synthetic Aperture Radar (SAR) Image Time Series (ITS) has become a popular topic of study since it has many practical applications for Earth monitoring such as disaster assessment or land-cover analysis. Developing reliable methodologies for Change Detection (CD) in SAR-ITS is thus an active topic of research. The CD problem is challenging due to the lack of ground truths, which does not allow to apply 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 unreliable. Under these conditions, unsupervised methodologies, often based on statistical tools, have been popular approaches in recent decades [1].

The CD problem can be seen as designing a distance. Among popular methodologies, Coherent Change Detection (CCD) [2] and the log-ratio operator [3] have received noticeable attention. However, these methodologies are limited to pairs of one-dimensional images, while modern sensors allow obtaining multidimensional ones (using e.g., polarimetric

The work was partially supported by PHOENIX ANR-15-CE23-0012 and MARGARITA ANR-17-ASTR-0015 grants of the French *National Agency of Research*.

or spectro-angular channels [4, 5]). Exploiting this diversity allows an improvement of performance in terms of CD applications for SAR-ITS.

For multivariate data, the covariance matrix has been shown to be a relevant feature so as to assess changes [6]. Assuming a complex Gaussian model, [7] has considered statistical information theory to design a distance, while [8] have adapted covariance homogeneity tests from the statistical literature [9], such as the Generalized Likelihood Ratio Test (GLRT). These methodologies allow to reach good performance, but suffer nonetheless from two issues encountered in high-dimension/resolution images:

- i) The Gaussian model has been shown to be inaccurate in recent radar clutter analysis [10] due to the inherent heterogeneity of these images. In order to be robust to this non-Gaussianity, [11] proposed various GLRTs, assuming a Compound Gaussian distribution.
- ii) Standard detectors are derived assuming unstructured covariance matrices, while the signal of interest usually lies in a low-dimensional subspace (e.g., only one polarisation among HH, HV and VV).

In this scope, [12] proposed to extend some GLRT approaches to Low-Rank (LR) structured covariance matrices.

To enjoy the best of both worlds, this paper proposes a new detector based on both robust and LR model: we derive a GLRT for Compound Gaussian distributed observations that have a LR structured covariance matrix. This proposed detector is then applied for CD on a SAR-ITS UAVSAR dataset and exhibits promising results.

2. GENERIC FRAMEWORK

2.1. Data: We consider a multidimensional ITS, that means each pixel at a given date corresponds to a vector of data of dimension p . These p channels can correspond to a polarimetric diversity ($p = 3$), or to another kind of diversity such as a spectro-angular one, obtained through wavelet transforms [5]. The CD is applied using a local window around the pixel of interest. Locally, the data set is denoted

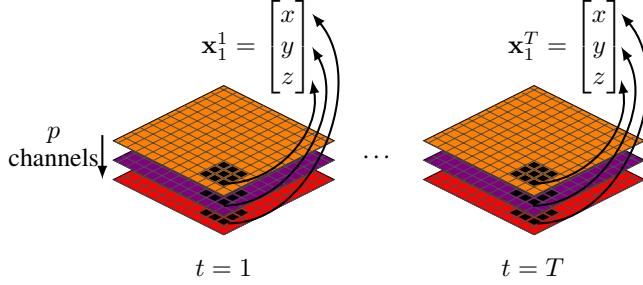


Fig. 1. Illustration of data for $p = 3, N = 9$. The pixels highlighted in black correspond to the local observations.

$\{\mathbf{x}_k^t\}_{(k,t) \in \llbracket 1, K \rrbracket \times \llbracket 1, T \rrbracket}$, corresponds to the concatenation of all channels for pixel at date t and spatial index k , as described in Figure 1.

2.2 GLRT for CD: For a given time t , the data is assumed to follow a given distribution of parameter θ_t , leading to the likelihood denoted $\mathcal{L}(\{\mathbf{x}_k^t\}_{k \in \llbracket 1, K \rrbracket} | \theta_t)$. The parameters θ_t characterise the local data at each date. Hence, if there is a change, the parameter is expected to vary.

For the sake of clear exposition, we will focus on $T = 2$ (CD between two acquisitions), which can be straightforwardly extended to $T > 2$. The CD problem can be formulated as a binary hypothesis test:

$$\begin{cases} H_0 : \theta_1 = \theta_2 & (\text{no change}), \\ H_1 : \theta_1 \neq \theta_2 & (\text{change}). \end{cases} \quad (1)$$

In order to derive a metric of decision, the Generalized Likelihood Ratio Test (GLRT), is considered. This test consists in computing the following quantity:

$$\hat{\Lambda} = \frac{\max_{\theta_1, \theta_2} \prod_{t=1}^2 \mathcal{L}(\{\mathbf{x}_k^t\}_{k \in \llbracket 1, K \rrbracket} | H_1; \{\theta_1, \theta_2\})}{\max_{\theta_1} \prod_{t=1}^2 \mathcal{L}(\{\mathbf{x}_k^t\}_{k \in \llbracket 1, K \rrbracket} | H_0; \theta_1)}, \quad (2)$$

where $\{\theta_1, \theta_2\}$ (resp. θ_1) corresponds to the parameters of the distribution of the observations under H_1 (resp. H_0). Hence, to develop efficient detectors, the problem remains to select an assumed distribution (and corresponding parameters) that accurately reflects the behavior of the data. Additionally, depending on the assumptions, the evaluation of the GLRT may lead to complex optimization problems.

3. GLRTS ON COVARIANCE MATRICES

3.1. Gaussian CD [6]: Assuming Gaussian distributed samples, the CD can be performed by testing a change in the covariance matrix. The corresponding GLRT, denoted $\hat{\Lambda}_G$ corresponds to (2) with the following distribution/parameters:

$$\mathbf{x}_k^t \sim \mathcal{CN}(\mathbf{0}, \Sigma^t) \quad \text{and} \quad \theta_G^t = \{\Sigma^t\} \quad (3)$$

this test has a closed-form expression and is well studied in the statistical literature [8].

3.2. LR-Gaussian CD [12]: Radar signals usually lie in lower dimensional subspaces, leading to a LR structured covariance matrix. The Gaussian GLRT that accounts for this prior knowledge, denoted $\hat{\Lambda}_{LRG}$, can be formulated as (2) with distribution/parameters:

$$\mathbf{x}_k^t \sim \mathcal{CN}(\mathbf{0}, \Sigma_R^t + \sigma^2 \mathbf{I}) \quad \text{and} \quad \theta_{LRG}^t = \{\Sigma_R^t\} \quad (4)$$

where Σ_R^t is the rank R signal covariance matrix and $\sigma^2 \mathbf{I}$ is the covariance matrix of the thermal noise. More details about this GLRT (setting R, σ^2 , computation...) can be found in [12] and section 5.2 of this paper.

3.3. Compound Gaussian CD [11]: For heterogeneous images, the Gaussian assumption may be a poor approximation of the underlying physics. In order to be robust to local power disparities, we can rely on the Compound Gaussian (CG) model (also referred to as a mixture of scaled Gaussian), which can accurately fit the empirical distribution of high-resolution data [10]. This model corresponds to a Gaussian one, where each realization is scaled by a local power factor τ referred to as texture (assumed unknown deterministic in this work). Hence, a corresponding GLRT for change detection, denoted $\hat{\Lambda}_{CG}$, can be formulated as (2) with distribution/parameters:

$$\mathbf{x}_k^t \sim \mathcal{CN}(\mathbf{0}, \tau_k^t \Sigma^t) \quad \text{and} \quad \theta_{CG}^t = \{\Sigma^t, \{\tau_k^t\}_{k \in \llbracket 1, K \rrbracket}\} \quad (5)$$

i.e., we test if both the covariance matrix Σ^t and the textures $\{\tau_k^t\}$ change between acquisitions. The computation of this quantity involves fixed-point equations that can be computed numerically. A study of this approach can be found in [11].

4. PROPOSED DETECTOR

In order to enjoy the improvement brought by both non-Gaussian and structure assumptions, we propose the following detector:

4.1. LR-Compound Gaussian CD: Assuming samples distributed as CG with a LR structured covariance matrix, the proposed GLRT, denoted $\hat{\Lambda}_{LRCG}$, corresponds to (2) with distribution/parameters:

$$\begin{aligned} \mathbf{x}_k^t &\sim \mathcal{CN}(\mathbf{0}, \tau_k^t (\Sigma_R^t + \sigma^2 \mathbf{I})) \\ \theta_{LRCG}^t &= \{\Sigma_R^t, \{\tau_k^t\}_{k \in \llbracket 1, K \rrbracket}\} \end{aligned} \quad (6)$$

where Σ_R^t is the rank R signal covariance matrix and $\sigma^2 \mathbf{I}$ is the covariance matrix of the thermal noise. Again, we test if both the covariance matrix and the textures change between acquisitions. The computation of this quantity involves optimisation techniques similar to ones used in [11] and [12]. The

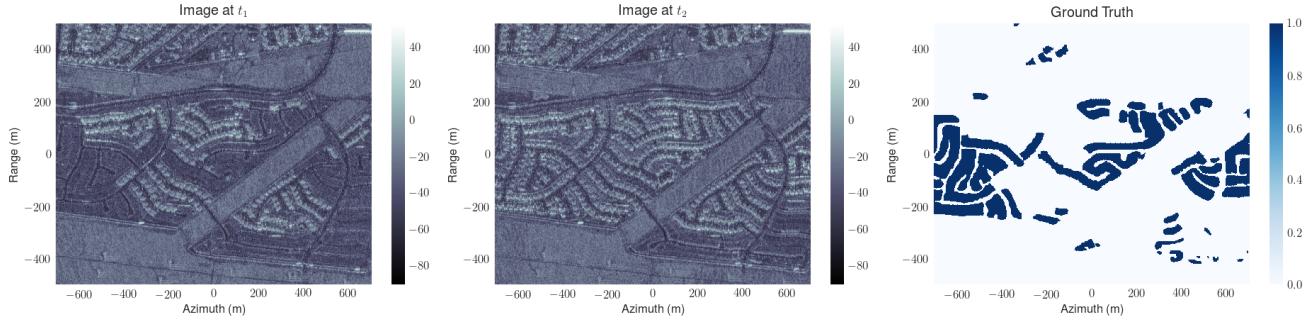


Fig. 2. UAVSAR Dataset used in this study. Left: April 23, 2009. Middle: May 15, 2011. Right: Ground Truth.

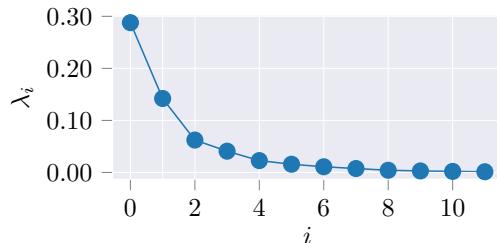


Fig. 3. Repartition of eigenvalues mean over the ITS.

cumbersome and technical calculus is left for a forthcoming paper¹. The following section will present an application of this detector for CD in SAR-ITS.

5. STUDY ON REAL UAVSAR DATASET

5.1. Description of data

To assess the performance of the proposed method, a pair of two images from UAVSAR SanAnd_26524_03 Segment 4 dataset² has been chosen since a ground truth has been established in [7] by using comparison with optical data. The images presented in Figure 2, correspond to full-polarisation data with a resolution of 1.67m in range and 0.6m in azimuth. Since the scatterers present in this scene exhibit an interesting spectro-angular behaviour, each polarisation of these images has been subjected to the wavelet transform presented in [5], allowing to obtain images of dimension $p = 12$.

5.2. Selection of rank and noise level

To compute the proposed detector, the rank R must be estimated beforehand. Several approaches exist in the literature [13] for its estimation. In this paper, we consider a simple approach by considering the distribution of the mean of eigenvalues over the ITS plotted in Figure 3. For this dataset, $R = 3$ appears to be an interesting value to separate signal

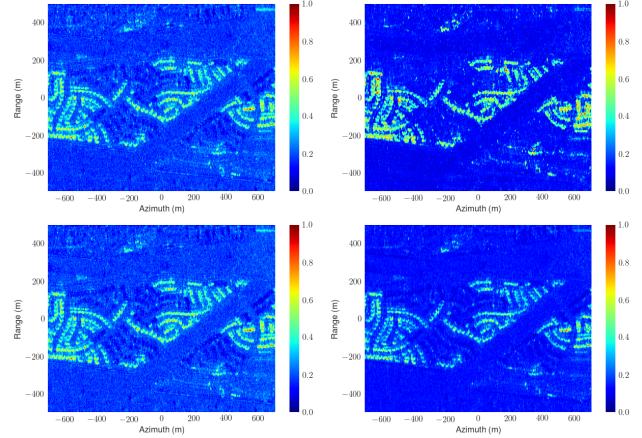


Fig. 4. Output of detectors (normalised) with $(p = 12, N = 25)$. Top-Left: $\hat{\Lambda}_G$. Top-Right: $\hat{\Lambda}_{LRG}$. Bottom Left: $\hat{\Lambda}_{CG}$. Bottom Right: $\hat{\Lambda}_{LRCG}$ (proposed).

from noise components. Notably, this rank gathers 81% of the total variance. The noise variance σ^2 is estimated locally with the mean of the $(p - R)$ lowest eigenvalues computed with an SVD of the SCM of all samples $\{\mathbf{x}_k^t\}$ in the patch.

5.3. Results

As a mean to assess the effectiveness of combining LR structure with a robust model, it is compared to the following detectors: *i*) the classic Gaussian statistic proposed in [6] (Section 3.1); *ii*) the LR Gaussian statistic of [12] (Section 3.2); *iii*) the CG statistic proposed in [11] (Section 3.3). Figure 4 presents the outputs of each detector for a window size of 5×5 . It appears that the LR detectors outputs (right column) contain less visual false alarms compared to the non-LR ones, which is expected since the most relevant channels are used to compute the CD, making it less sensitive to noise.

Figure 5 shows the Receiver Operator Curve (ROC) for the results of Figure 4. The proposed method allows obtaining the best performance of detection for any given false

¹A python code for the proposed methodology is available at <https://github.com/AmmarMian/Robust-Low-Rank-CD>

²Available at <https://uavstar.jpl.nasa.gov/>.

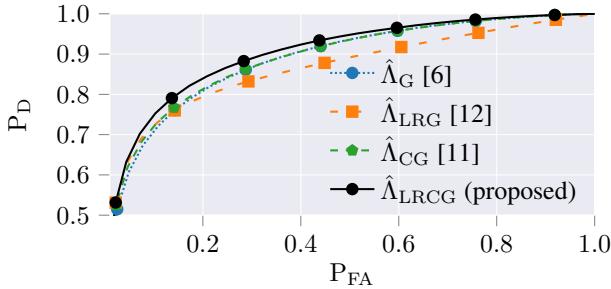


Fig. 5. Probability of detection P_D versus the the probability of false alarm P_{FA} with ($p = 12, N = 25$)

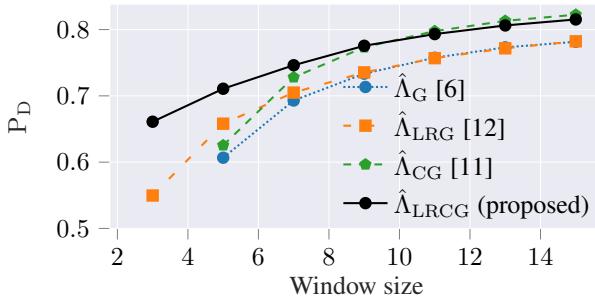


Fig. 6. P_D versus the size of window at $P_{FA} = 5\%$ with ($p = 12, R = 3$)

alarm rate, which is an interesting result. The gain is most apparent with regards to the LR Gaussian detector which performs poorly for false alarms rate greater than 10%. This is explained by the fact that an LR structure in Gaussian context results in the loss of some signal power while in our model, the texture parameters account for the entirety of this power.

Finally, Figure 6 shows the evolution of the performance in terms of P_D at $P_{FA} = 5\%$ when increasing the size of the window used to compute the detectors. Increasing the window improves the results at the cost of a resolution loss. The interest of LR methods is well demonstrated here: they allow to obtain good detection results with a lower size of window compared to their non-LR counterparts. The proposed method exhibits the best performance, which was to be expected given that it has been derived using a model more appropriated to the data.

6. 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.
- [2] M. Preiss, D. A. Gray, and N. J. S. Stacy, “Detecting scene changes using synthetic aperture radar interferometry,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2041–2054, Aug 2006.
- [3] Y. Bazi, L. Bruzzone, and F. Melgani, “An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 874–887, April 2005.
- [4] 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 *2017 25th European Signal Processing Conference (EUSIPCO)*, Aug 2017, pp. 1942–1946.
- [5] A. Mian, J-P. Ovarlez, A. M. Atto, and G. Ginolhac, “Design of new wavelet packets adapted to high-resolution SAR images with an application to target detection,” *IEEE Transactions on Geoscience and Remote Sensing*, 2019.
- [6] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, “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*, vol. 41, no. 1, pp. 4–19, Jan 2003.
- [7] 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. to appear, pp. 1–13, 2019.
- [8] D. Ciuonzo, V. Carotenuto, and A. De 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.
- [9] T. W. Anderson, *An introduction to multivariate statistical analysis*, vol. 2, Wiley New York, 1958.
- [10] M. 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, 2007.
- [11] 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.
- [12] R. Ben Abdallah, A. Mian, A. Breloy, M. N. El Korso, and D. Lautru, “Detection methods based on structured covariance matrices for multivariate SAR images processing,” *IEEE Geoscience and Remote Sensing Letters*, 2019.
- [13] P. Stoica and Y. Selen, “Model-order selection: a review of information criterion rules,” *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, July 2004.

CHAPTER 4

CHANGE-POINT ESTIMATION BY EXPLOITING ITERATIVE HYPOTHESIS TESTING

Contents:

4.1	Derivation of statistics under robust model adapted to this strategy	140
4.2	Simulation on synthetic dataset	141
4.2.1	Results on synthetic time-series	141
4.2.2	Results on synthetic images	144
4.2.3	Application to SAR image time series	149
4.3	A Bayesian MSE Lower-Bound for the tuning	152
4.3.1	Data model	153
4.3.2	Covariance inequality	154
4.3.3	Hybrid bound for the change-point model	155
4.3.4	Derivation of $\mathbf{F}(\Sigma)$, $\rho(s)$ and $\phi_{\sigma_j, \ell}(s)$	155
4.3.5	Computation of the tightest bound	156
4.3.6	Simulations	157
4.4	Conclusions	159
4.A	Appendix	161
4.A.1	Proof of proposition 4.1.1 at p. 140	161
4.A.2	Proof of proposition 4.1.2 at p. 141	164
4.A.3	Proof of proposition 4.3.4 at p. 156	166

In this chapter, we consider the utilization of the robust covariance equality testing schemes developed in Chapter 3 in order to tackle the problem of change-point estimation for SAR images. We have seen in Chapter 1 that several approaches can be considered for change-point estimation in time-series of multivariate datasets which could be adapted in the robust data model. However, as discussed we will consider here only an extension of the methodology presented in [Conradsen et al., 2016] since it is an approach based on covariance equality testing for which the extension to non-Gaussian models is immediate.

We will thus adapt the statistics of chapter 3 to work in this setup. We will consider only the deterministic compound-Gaussian model since we have seen that the elliptical one is impractical in real data applications (due to the fact that g is unknown in most real images application and that depending on the changes, the density generator most fitting the data could change). Moreover, we will not derive the scale only statistic since we have shown it does not possess the matrix CFAR property making it unable to guarantee a false

alarm rate. Simulations will be done to demonstrate the utility of this extension to robust models.

Finally, we will consider a side problem which consists in tuning the size of the window needed for the change-point estimation. To that end, we will derive a lower-bound on the mean square error of estimation for any Bayesian estimator of the change-point. Although the setup considered in this case will be in Gaussian context and assume a Bayesian estimation strategy, the insights offered by this study are of interest since they allow to compare the performance of estimation of our methodology to more standard Bayesian estimators.

4.1 DERIVATION OF STATISTICS UNDER ROBUST MODEL ADAPTED TO THIS STRATEGY

In order to use the algorithm 2 in the robust context, we will need statistics of decision for the following marginal scheme:

Consider $(t_1, t_2) \in \llbracket 1, T \rrbracket^2$, so that $t_2 > t_1$, decide between:

$$\begin{cases} H_{0,\text{marg}}^{t_1, t_2} : \boldsymbol{\theta}_{t_1} = \dots = \boldsymbol{\theta}_{t_{2-1}} = \boldsymbol{\theta}_{t_1, t_2-1} \text{ and } \boldsymbol{\theta}_{t_{2-1}} = \boldsymbol{\theta}_{t_2} \\ H_{1,\text{marg}}^{t_1, t_2} : \boldsymbol{\theta}_{t_1} = \dots = \boldsymbol{\theta}_{t_{2-1}} = \boldsymbol{\theta}_{t_1, t_2-1} \text{ and } \boldsymbol{\theta}_{t_{2-1}} \neq \boldsymbol{\theta}_{t_2} \end{cases} \quad (4.1)$$

This section is dedicated to the derivation of such statistics, under the deterministic compound-Gaussian model, for testing equality of either shape or shape and scale. We will keep the notations of Chapter 3 regarding the problems of detection and statistic names and refer to the new statistics as marginal statistics.

We have the following results:

Proposition 4.1.1. *The GLRT ratio under hypotheses of problem (3.9) (Scale and shape) for marginal scheme (4.1) is the following:*

$$\hat{\Lambda}_{\text{MT}}^{\text{marg}, t_1, t_2} = \frac{\left| \hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{MT}} \right|^{(t_2 - t_1)N}}{\left| \hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{MT}} \right|^{(t_2 - t_1 - 1)N} \left| \hat{\boldsymbol{\xi}}_T^{\text{TE}} \right|^N} \frac{((t_2 - t_1 - 1)p)^{(t_2 - t_1 - 1)Np} p^{Np}}{((t_2 - t_1)p)^{(t_2 - t_1)Np}} \times \prod_{k=1}^N \frac{\left(\sum_{t=t_1}^{t_2} q(\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{MT}}, \mathbf{x}_k^t) \right)^{(t_2 - t_1)p}}{\left(\sum_{t=t_1}^{t_2-1} q(\hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{MT}}, \mathbf{x}_k^t) \right)^{(t_2 - t_1 - 1)p} \left(q(\hat{\boldsymbol{\xi}}_T^{\text{TE}}, \mathbf{x}_k^T) \right)^p} \stackrel{H_1}{\gtrless} \stackrel{H_0}{\lesssim} \lambda, \quad (4.2)$$

where

$$\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{MT}} = \frac{p}{N} \sum_{k=1}^N \frac{\sum_{t=t_1}^{t_2} \mathbf{s}_k^t}{\sum_{t=t_1}^{t_2} q(\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{Mat}}, \mathbf{x}_k^t)}. \quad (4.3)$$

Proof. See Appendix 4.A.1. □

Proposition 4.1.2. *The GLRT ratio under hypotheses of problem (3.10) (Shape only) for marginal scheme (4.1) is the following:*

$$\hat{\Lambda}_{\text{Mat}}^{\text{marg}, t_1, t_2} = \frac{\left| \hat{\xi}_{t_1, t_2}^{\text{Mat}} \right|^{(t_2 - t_1)N}}{\left| \hat{\xi}_{t_1, t_2-1}^{\text{Mat}} \right|^{(t_2 - t_1 - 1)N} \left| \hat{\xi}_T^{\text{TE}} \right|^N} \times \prod_{k=1}^N \frac{\prod_{t=t_1}^{t_2} \left(q \left(\hat{\xi}_{t_1, t_2}^{\text{Mat}}, \mathbf{x}_k^t \right) \right)^p}{\left(\prod_{t=t_1}^{t_2-1} \left(q \left(\hat{\xi}_{t_1, t_2-1}^{\text{Mat}}, \mathbf{x}_k^t \right) \right)^p \right) \left(q \left(\hat{\xi}_T^{\text{TE}}, \mathbf{x}_k^T \right) \right)^p} \stackrel{H_1}{\gtrless} \lambda, \quad (4.4)$$

where

$$\hat{\xi}_{t_1, t_2}^{\text{Mat}} = \frac{p}{N} \sum_{t=t_1}^{t_2} \sum_{k=1}^N \frac{\mathbf{S}_k^t}{q \left(\hat{\xi}_{t_1, t_2}^{\text{MT}}, \mathbf{x}_k^t \right)}. \quad (4.5)$$

Proof. See Appendix 4.A.2. □

Notes:

- Since the optimizations have been done similarly to the one in chapter 3, the convergence analysis still stands, meaning that the detectors can be implemented in practice for any dataset given that the number of linearly independent samples is sufficient.
- At this point, no theoretical approximation of the distribution under the null hypothesis of those statistics has been obtained. The CFAR property is nonetheless obtained for those statistics (in the same way it has been shown for the omnibus ones). Thus we shall consider Monte-Carlo simulations for the choice of the threshold of decision.

4.2 SIMULATION ON SYNTHETIC DATASET

4.2.1 Results on synthetic time-series

Description of the simulation

A simulation to assess the performance of the algorithm with robust statistics has been undertaken. A synthetic time series has been generated where the observations are of the form $\mathbf{x}_k^t = \sqrt{\tau_k^t} \tilde{\mathbf{x}}$ for $(k, t) \in [\![1, N]\!] \times [\![1, T]\!]$, where $\tilde{\mathbf{x}} \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, \Sigma)$ and τ_k^t is a realization of a random variable with values on \mathbb{R}^+ .

Three possible situations have been considered:

- The Gaussian model $\sqrt{\tau_k^t} = 1$.
- An elliptical model obtained with $\tau_k^t \sim \Gamma(\alpha, \beta)$, where $\Gamma(\alpha, \beta)$ denotes the Gamma distribution with shape parameter α and scale parameter β .
- A model in which $\tau_k^t \sim \Gamma(\alpha, \beta)$ except that $\forall (t_1, t_2) \in [\![1, T]\!]^2$, $\tau_k^{t_1} = \tau_k^{t_2}$. This allows to have a texture equality constraint which was the model for detector $\hat{\Lambda}_{\text{MT}}$.

The covariance matrices are chosen to be Toeplitz of the form $(\Sigma_t)_{m,n} = \rho_t^{|m-n|}$. We consider a situation of a single change-point to be detected/estimated similarly as what was considered in subsection 3.2.6 of Chapter 3.

As for the previous chapter, we consider the Bartlett distance to measure the amplitude of change:

$$d_B(\Sigma_1, \Sigma_2) = \log \left(\frac{|\Sigma_1 + \Sigma_2|^2}{|\Sigma_1| |\Sigma_2|} \right) - 2p \log(2). \quad (4.6)$$

This allows to have an equivalent SNR quantity. In order for this quantity to be defined for all problems, we only consider a change in the shape matrix so the SNR is defined as $d_B(\Sigma_0, \Sigma_1)$ where Σ_0 is the shape matrix before the change-point and Σ_1 is the shape matrix after the change-point.

Finally, the thresholds of detection are computed numerically through Monte Carlo trials using the H_0 regime of the problem considered.

To assess the performance of change-point detection, we consider the following probability of detection:

$$P_D = \sum_{\hat{t}_C \in \hat{\mathbb{T}}_C(\hat{\Lambda})} P(\hat{t}_C = t_C), \quad (4.7)$$

where $\hat{\mathbb{T}}_C(\hat{\Lambda}) = \{ \hat{t}_1, \dots, \hat{t}_K \}$ is the set of detected change-points using the iterative algorithm associated with statistic $\hat{\Lambda}$ and t_C is the true value of change-point. This probability of detection differs from the one in Chapter 3 for which any detection resulted in a positive outcome while in this case, the detection has to done at the true change-point. This allows to take into account the estimation problem as well.

Results

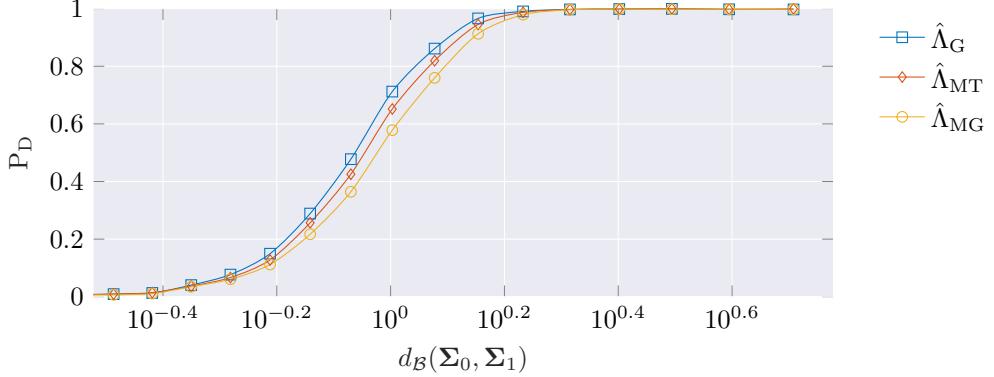


Figure 4.1: $P_D = f(\text{SNR})$ for a Gaussian model. $p = 10$, $N = 25$, $T = 10$, $t_C = 5$, $\rho_{t < t_C} = 0.01$. The false alarm rate is fixed at $P_{FA} = 10^{-3}$ and each point of the curve is plotted using 4800 Monte-Carlo trials.

First, we consider a Gaussian context in Figure 4.1. In this case, the algorithm associated with Gaussian-derived statistics yields the best results while the performance is slightly degraded for robust statistics. This result is coherent with the change detection results of Chapter 3.

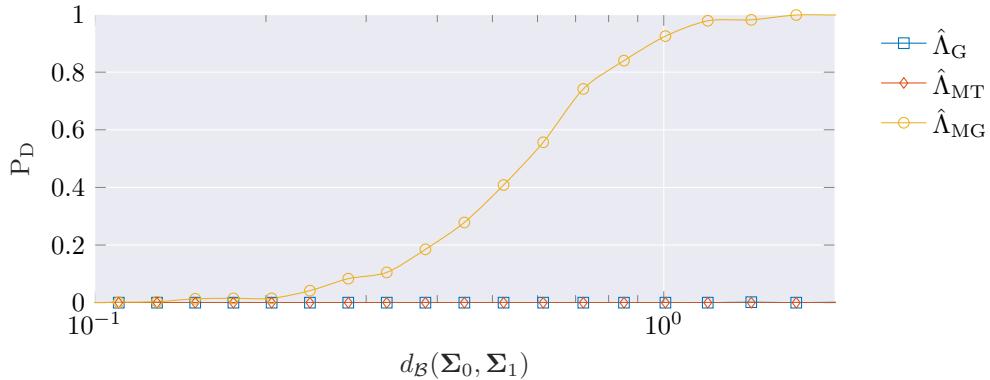


Figure 4.2: $P_D = f(\text{SNR})$ for an Elliptical model. $p = 10$, $N = 25$, $T = 10$, $t_C = 5$, $\rho_{t < t_C} = 0.01$, $\alpha = 0.1$, $\beta = 0.3$. The false alarm rate is fixed at $P_{FA} = 10^{-3}$ and each point of the curve is plotted using 4800 Monte-Carlo trials.

Next, in the elliptical model plotted in Figure 4.2, only the statistic associated with a change in shape matrix yields a detection. Indeed, the almost null detection for other statistics is explained by the fact that the detectors do not regulate well the false alarm rate in this model. Indeed, the detectors are not adapted for the model of the change in this simulation which means that even though there is no change according to this model, the test statistic will have a high value. Thus, the experimental thresholds obtained are much higher meaning that detection at the exact true change-point is a rare occurrence.

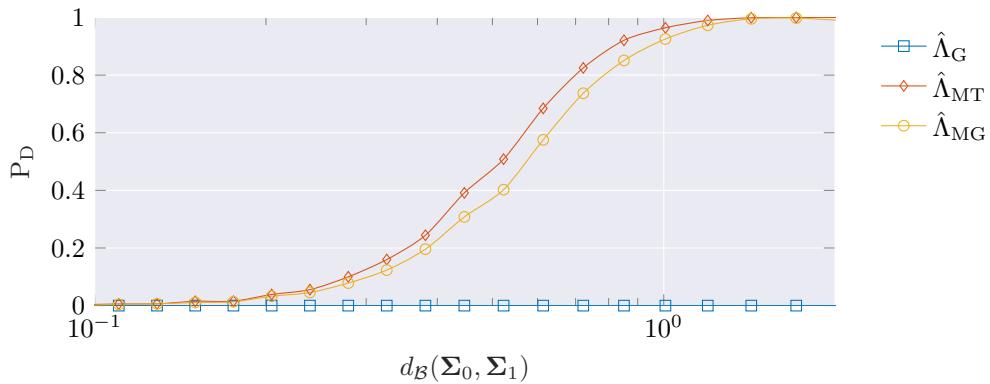


Figure 4.3: $P_D = f(\text{SNR})$ for an elliptical model with texture equality constraint. $p = 10$, $N = 25$, $T = 10$, $t_C = 5$, $\rho_{t < t_C} = 0.01$, $\alpha = 0.1$, $\beta = 0.3$. The false alarm rate is fixed at $P_{FA} = 10^{-3}$ and each point of the curve is plotted using 4800 Monte-Carlo trials.

Finally, in Figure 4.3, a texture equality constraint has been imposed between the dates to match the model associated with statistic $\hat{\Lambda}_{MT}$ (meaning that $\forall t \in [1, t_C] \tau_k^t = \tau_k$ and $t \in [t_C + 1, T], \tau_k^t = \tau'_k$). In this case, only the Gaussian-derived statistic does not allow a detection.

From these simulations, we can conclude that the observations of Chapter 3 with regards to the statistics still stand: $\hat{\Lambda}_{Mat}$ statistic allows to have the most robust behavior since its

performances does not vary for the three models considered. However, it is expected that on real data, the performances are poorer since the scale is an important parameter of the change.

4.2.2 Results on synthetic images

To briefly illustrate the usability of the algorithms on real images, we first generate synthetic images. To generate the data we consider the three models of the previous simulation: Gaussian, elliptical and elliptical with a texture equality constraint. We generate a noisy background and two types of changes: a cross-shaped pattern that appears and disappears and a growing circle. The data has been generated with parameters described in Figure 4.4. The circle corresponds to a moderate change in scale and an important one in shape while the cross-shaped pattern corresponds to an important change in scale and a small one in shape.

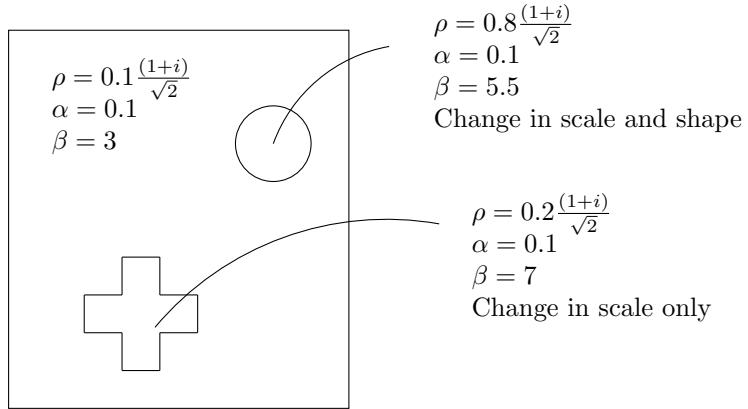


Figure 4.4: Description of synthetic images parameters. Parameters α and β only apply for elliptical model.

In order to visualize the results, we present the results as series of images on which a zero value (corresponding to black color) at pixel (x, y) for a time t means that no change was detected and estimated. A non-zero value (corresponding to white color) correspond to change detected at pixel (x, y) and time t .

Figure 4.5 presents the results associated with a Gaussian model. It appears that all statistics detect and estimate correctly the growing circle while $\hat{\Lambda}_{\text{Mat}}$ does not allow to discern the cross-shaped pattern. This result is sound since it is mainly consist in a change of scale.

Next, Figure 4.6 presents the results associated with the elliptical model and no texture equality constraint. In this case, only $\hat{\Lambda}_{\text{Mat}}$ has usable output: the growing circle is still well detected and estimated. However, the cross is not detected since it mostly corresponds to a change in scale. For the other two statistics, almost every pixel of the second and fourth images are marked as a change. This is explained by the fact that they are not able to regulate false alarms in this model. The statistic grows higher than the threshold for almost every pixel even though there is no apparent change. For the few pixels that were not mapped as change, the statistic's value grows significantly higher at next iteration resulting in a map

at the next timestamp as a change.

Finally, Figure 4.7 presents the results associated with the elliptical model and texture equality constraint. In this model, the Gaussian statistic's behavior does not change while $\hat{\Lambda}_{\text{Mat}}$ statistic allows detecting both change patterns. This is explained by the fact that in this model the false alarm is well regulated for this statistic. $\hat{\Lambda}_{\text{Mat}}$ statistic's output doesn't change much from Figure 4.6 since from its perspective both models allow to regulate false alarms.

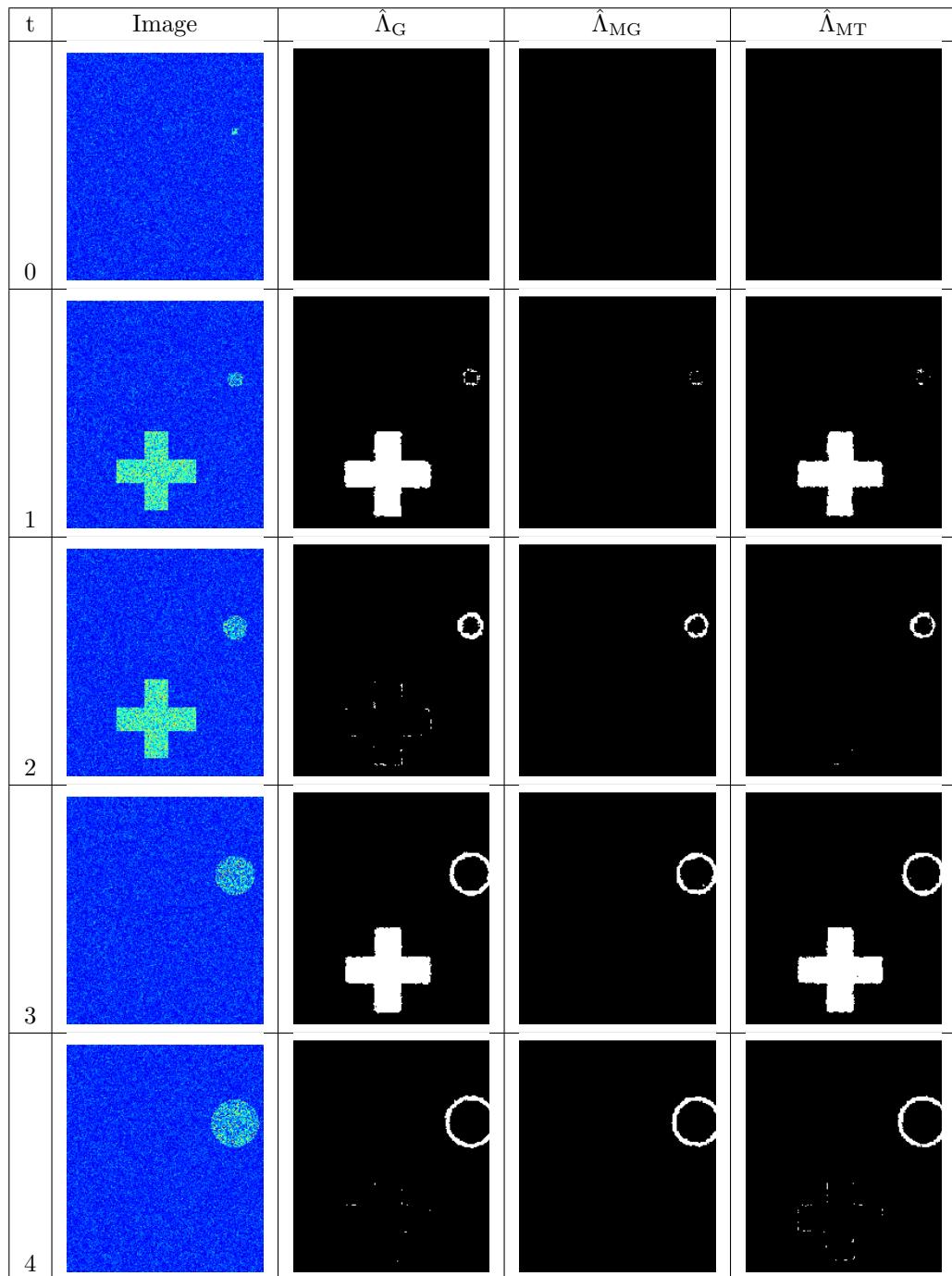


Figure 4.5: Estimation/Detection results at $P_{FA} = 10^{-3}$ for synthetic Gaussian data ($p = 5$, $N = 25$, $T = 5$). A change-point detected and estimated corresponds to non-zero value on the resulting image.

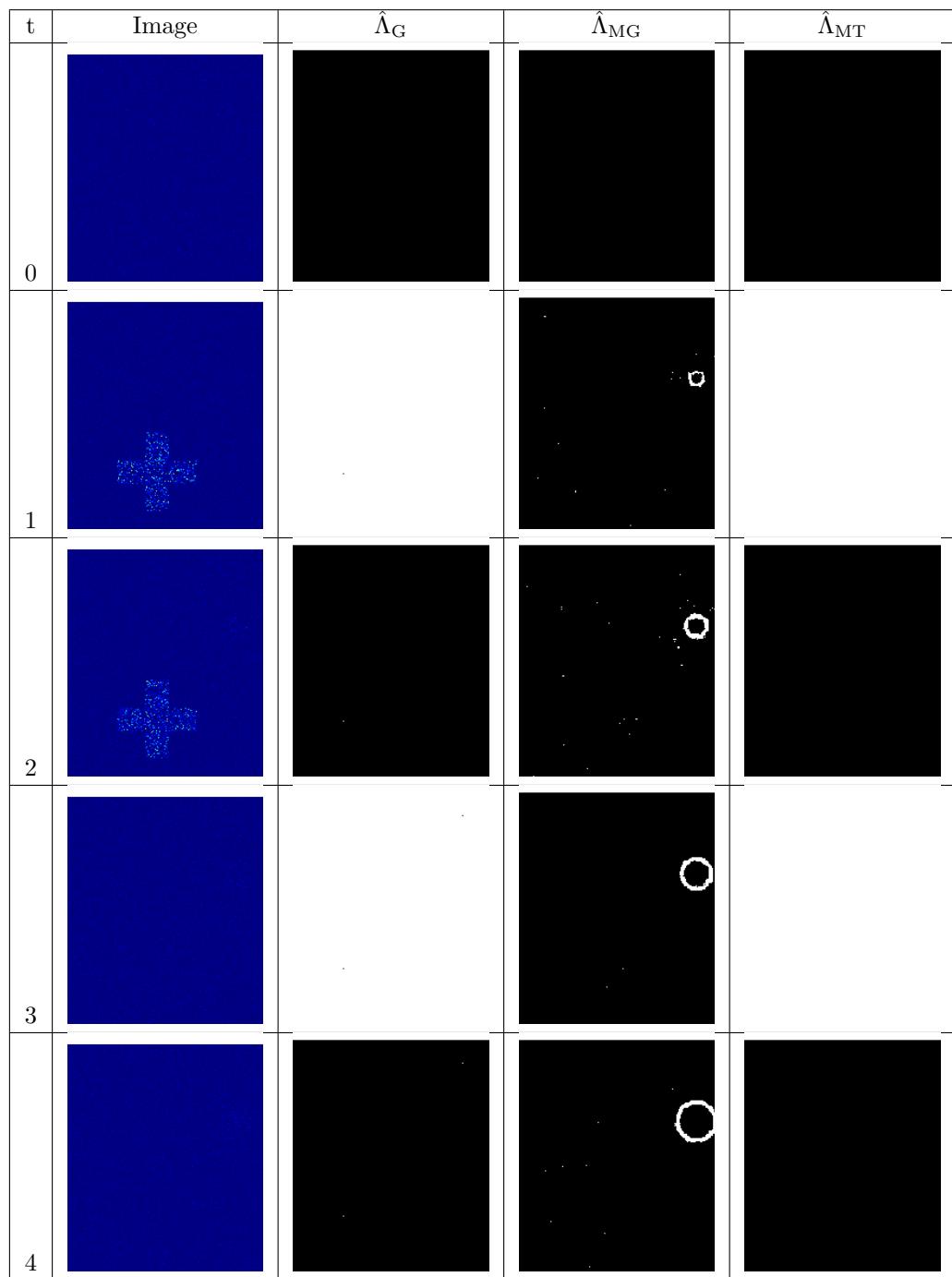


Figure 4.6: Estimation/Detection results at $P_{FA} = 10^{-3}$ for synthetic K-distributed data ($p = 5, N = 25, T = 5$). A change-point detected and estimated corresponds to non-zero value on the resulting image.

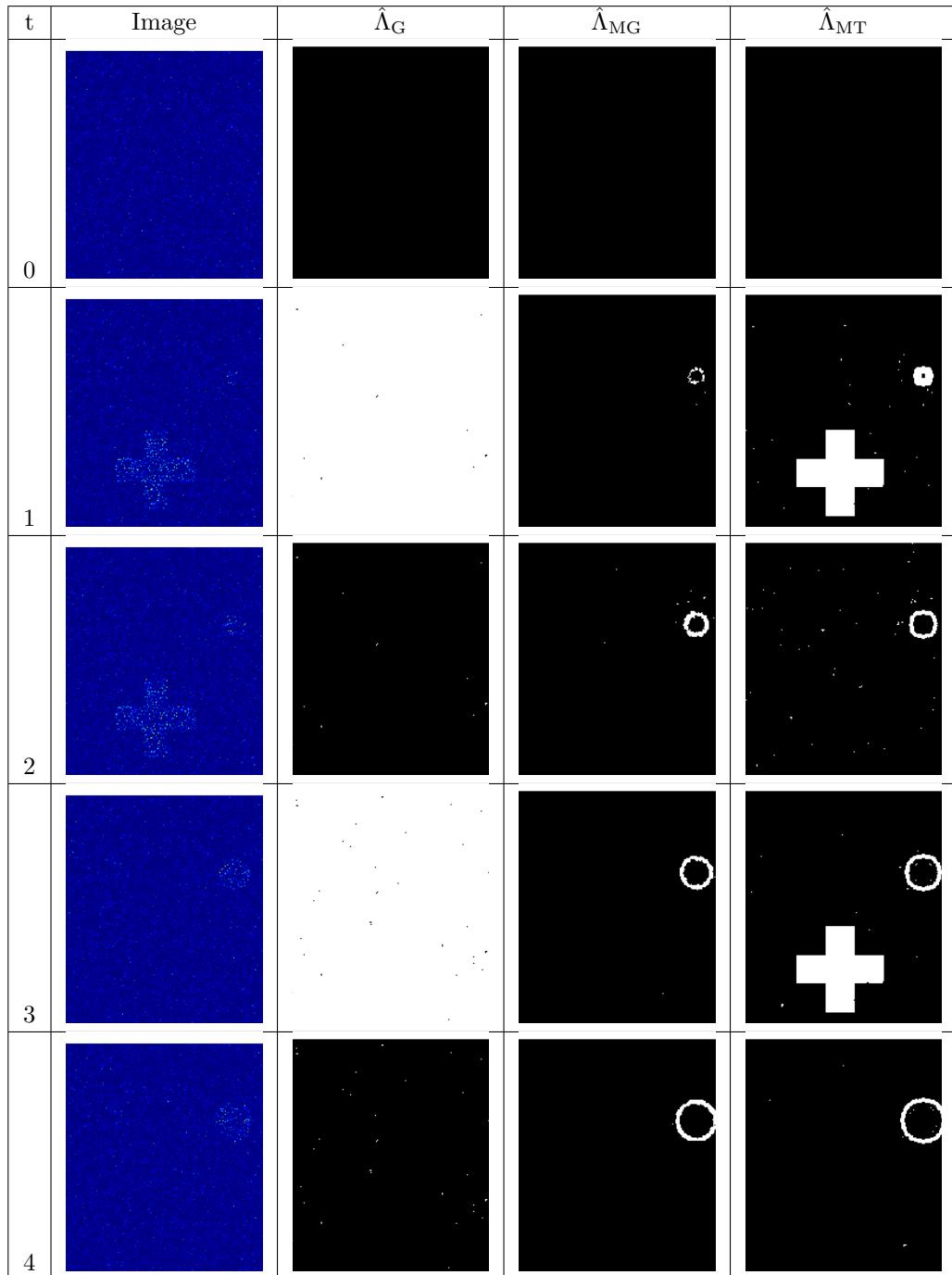


Figure 4.7: Estimation/Detection results at $P_{FA} = 10^{-3}$ for synthetic K-distributed data with texture equality constraint between dates ($p = 5, N = 25, T = 5$). A change-point detected and estimated corresponds to non-zero value on the resulting image.

4.2.3 Application to SAR image time series

The change point detection/estimation strategy has been assessed on real data coming from the SDMS CCD Challenge dataset. A subset (2 and 3 images) dataset was used in chapter 3 to compare change detection performances.

In this case, we consider the whole time series consisting of 10 images taken during the same day. The change-points mainly consist in vehicles appearing or disappearing in the several parking lots of the area. However, no ground truth is available for this time series. Figures 4.8 and 4.9 show the result of detection and estimation. In order to keep information about the amplitude of the change, the pixels detected as change-point have been mapped to a value in $(0, 1]$ according to the value $\hat{\Lambda}/\text{threhsold}$ which led to the detection.

Discussion: A first analysis of the results shows a high number of false alarms, especially at dates $t = 6$ and $t = 7$. Two main reasons explain this output: the linear patterns correspond to artifact already present in some images which can appear during the processing of the SAR images. Moreover, the great number of false detections at dates $t = 6$ and $t = 7$ is due to the co-registration which is off by several pixels. This leads to a situation where from a pixel point of view, the objects have shifted which correspond to a change. This highlights that the algorithm is not robust to co-registration issues and must only be used on well registered time series.

Concerning the Gaussian GLRT statistic, the non-Gaussian nature of the data can also explain that the threshold of detection is estimated to be too low to actually guarantee the false alarm rate. For $\hat{\Lambda}_{\text{Mat}}$, the detection is done solely on the shape matrix which can be in a sense linked to the coherence between the images. The zones appearing as changes are mostly corresponding to trees or other similar natural objects which are known for their non-coherent behavior between acquisitions.

Comparison between the different statistics shows that $\hat{\Lambda}_{\text{MT}}$ yields an overall better detection by reduction a significant amount of false alarms. This can be explained by the fact that, even though it can detect a change in both shape and scale, the value of the statistic is more sensitive to the scale. Indeed, while a change in shape is limited in amplitude due to the normalization, this is not the case for a change in scale which is not bounded. Thus, the detection of a change in shape is not as frequent as $\hat{\Lambda}_{\text{Mat}}$.

In conclusion, these results reveal the fact that the change detection problem is an ill-defined one. Indeed, while we wanted to detect changes corresponding to vehicles, other types of changes have also been included: some corresponding to the co-registration, others due to the non-coherent behavior of some objects and finally artifacts due to the SAR images construction. In absolute, these detections do not correspond to false alarms since they effectively impacted the covariance over time however they can be undesirable depending on the definition of change considered. From a practical standpoint, it appears that covariance based methods are useful for discovering any type of change in the series but not for specifically oriented applications.

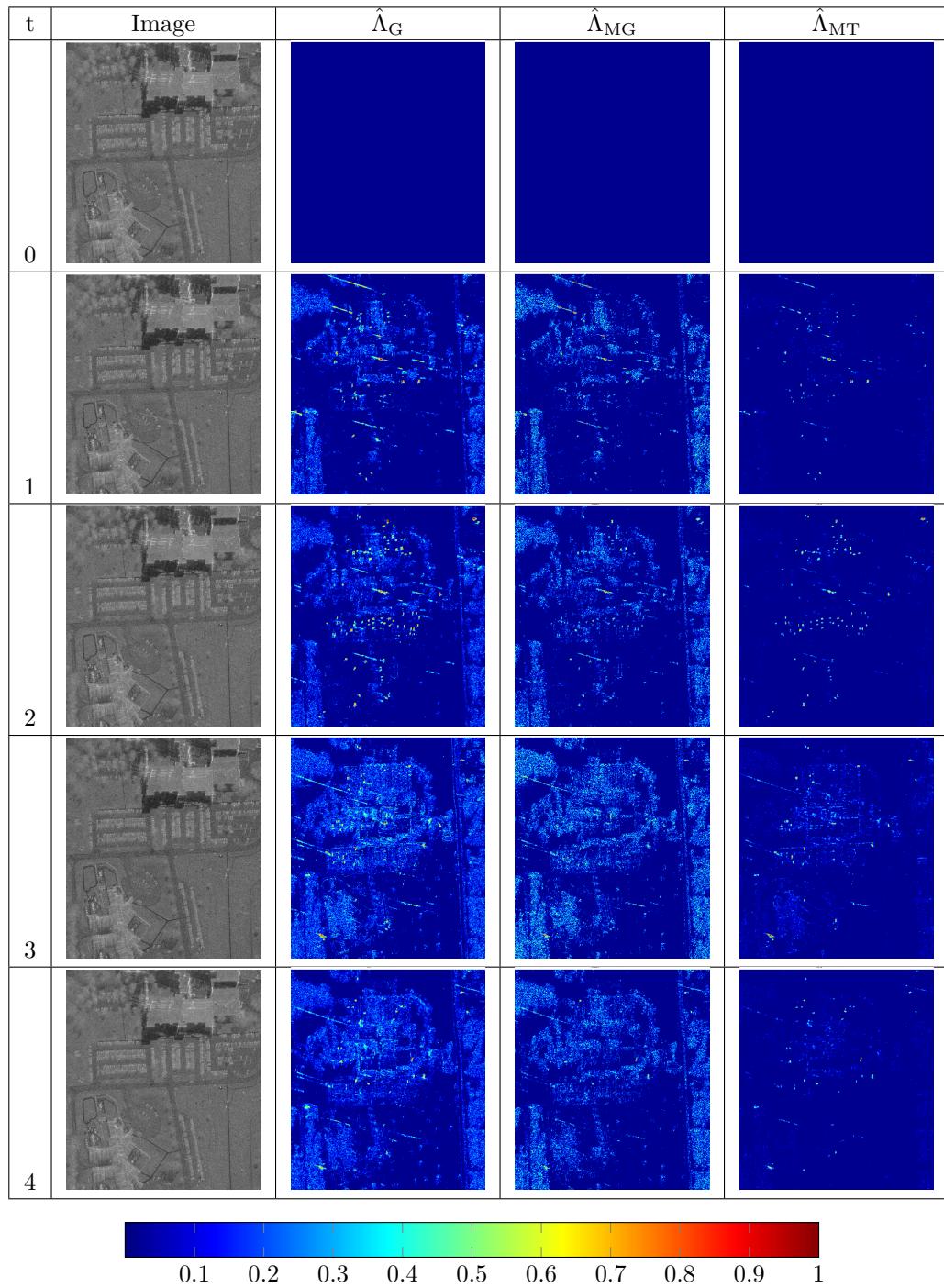


Figure 4.8: Estimation/Detection on SDMS dataset results at $P_{FA} = 10^{-4}$

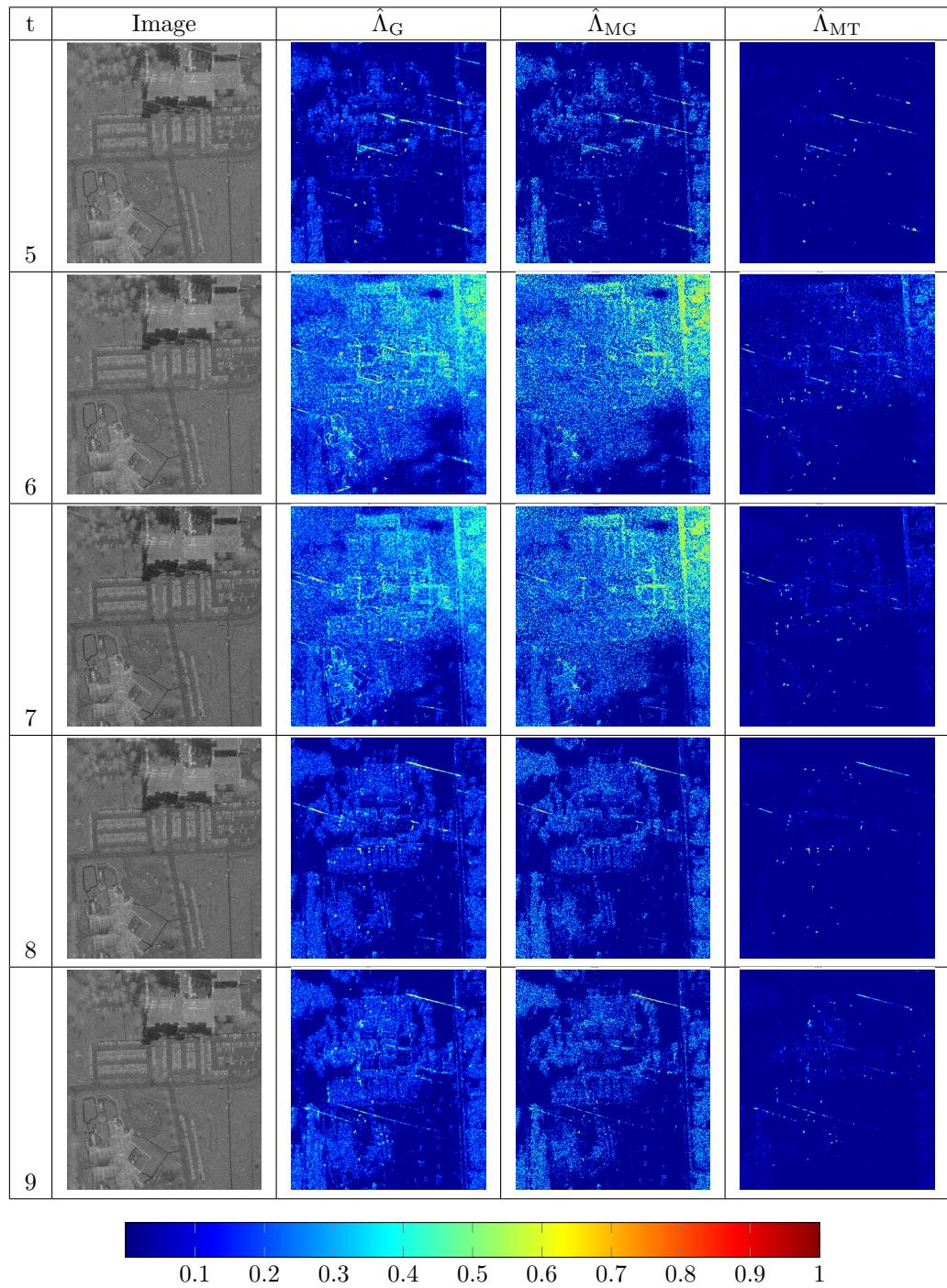


Figure 4.9: Estimation/Detection on SDMS dataset results at $P_{FA} = 10^{-4}$

4.3 A BAYESIAN MSE LOWER-BOUND FOR THE TUNING

In this section, we will consider the derivation of a lower bound on the MSE of any Bayesian estimator of change-point in Gaussian context. This study distances itself from the previous scenario in which there was no knowledge of the number of change-points as well as any prior about their distribution. As such, the input obtained from the lower bound is not applicable to the unknown number of change-points context. However, the study aims at being able to compare the performance of the estimation strategy with regards to more standard Bayesian change-point estimation schemes. The study presented here has been done in collaboration with Lucien Bacharach and Alexandre Renaux of *Laboratoire des Signaux et Systèmes, CentraleSupélec* who have considered the derivation of the bound in a general case without specifying a model.

Another aim of this study is to consider the problem of parameter tuning. Indeed, in our model, the dimension of the data, as well as the number of samples N for each date, are linked to the performance of detection and estimation. Concerning the detection performance, the number of samples impact has been studied experimentally on a real dataset in Section 3.3.

This study is aimed at tackling the problem of tuning the size of the windows with regards to the estimation performance. Indeed, intuitively having more samples is better for both estimation and detection performance but from a practical standpoint, the size of the window is linked to the resolution of detection. A window too large will yields detection at a poor spatial resolution while also increasing the heterogeneity of the data considered. Moreover, this estimation step must be done by taking into account the dimension of the dataset p which can be controlled thanks to the wavelet decomposition presented in Chapter 2. Thus one would want to tune this parameter in order to obtain a compromise between estimation performance and spatial resolution of detection.

Here, we consider this problem by using the Mean Square Error (MSE) of the estimated change-points as a measure of performance and propose a lower bound on its expected value. Indeed, the only available theoretical results on change-point estimation concern specific problems such as monovariate Gaussian time series with an asymptotic assumption of a large number of data before and after the change [Fotopoulos and Jandhyala, 2001, Fotopoulos et al., 2010]. In order to overcome this difficulty, the signal processing community generally focusses on MSE analysis with Monte Carlo simulations, which are computationally expensive, or the Cramer-Rao Bound (CRB). However, in the context of change-point estimation, a part of the unknown parameter vector lies on a discrete space violating the regularity condition of the CRB.

Lower-bound for change-point estimation has been firstly considered in [Rosa et al., 2010, Ferrari and Tourneret, 2003] where only the Barankin Bound (BB) was considered. Then, the extension to a tighter Bayesian bound (i.e. the WWB) has been proposed in [Bacharach et al., 2017] where a general semi-closed-form expression focussing on change-points without specifying the distribution of the data, is given. This work has been extended by including the possibility of unknown additional parameters in [Bacharach et al., 2019].

As a first step, we will only consider the Gaussian context, since the derivation of the bound becomes more intricate in the elliptical and compound-Gaussian context. More specifically, we will consider the distribution of the SCM over the observations which in the Gaussian model is known to be the Wishart distribution. We will thus, adapt hereafter this semi-closed expression to the case of Wishart distributed data.

4.3.1 Data model

We consider a time series of T independent, Wishart-distributed matrices $\{\mathbf{S}_t \in \mathbb{S}_{\mathbb{H}}^p | t \in [1, T]\}$, where $\mathbb{S}_{\mathbb{H}}^p$ is the set of positive definite Hermitian matrices, subjected to a single change-point t_C . This scenario corresponds to a case where the user has the knowledge that there is an abrupt change and wants to know its precise localization. This kind of problem can arise, for example, when looking for the time at which some flood has occurred or a region where a forest has been cut. The case of multiple change-points has also been considered but requires heavier derivations. Thus for the sake of clear exposition, we will focus on the case of one change-point. The studied scenario is then written as follows:

$$\begin{cases} \mathbf{S}_t \sim \mathbb{C}\mathcal{W}(p, N, \Sigma_0) & \text{for } t = 1, \dots, t_C \\ \mathbf{S}_t \sim \mathbb{C}\mathcal{W}(p, N, \Sigma_1) & \text{for } t = t_C + 1, \dots, T \end{cases}, \quad (4.8)$$

where $\mathbb{C}\mathcal{W}$ is the central complex Wishart distribution with Probability Distribution Function (PDF):

$$p_{\mathbf{S}_t; \Sigma}(\mathbf{S}_t; \Sigma) = \frac{|\mathbf{S}_t|^{N-p}}{\Gamma_p(N) |\Sigma|^N} \text{etr}(\Sigma^{-1} \mathbf{S}_t), \quad (4.9)$$

where, $\Gamma_p(N) = \pi^{p(p-1)/2} \prod_{j=1}^p \Gamma(N - j + 1)$, $\Gamma(\cdot)$ is the Gamma function and $\text{etr}(\cdot)$ is the exponential trace function.

In order to deal with complex parameters, we define the mapping from complex to real-case as a function $[\bullet]_{\mathbb{CR}} : \mathbb{C}^m \rightarrow \mathbb{R}^{2m}$, whose definition is: $[\mathbf{z}]_{\mathbb{CR}} = [\Re(\mathbf{z})^T, \Im(\mathbf{z})^T]^T$.

We also define:

$$\begin{aligned} \mathbf{x} &= [[\text{vech}(\mathbf{S}_1)]_{\mathbb{CR}}^T, \dots, [\text{vech}(\mathbf{S}_T)]_{\mathbb{CR}}^T]^T \in \mathbb{R}^{Tp^2}, \\ \boldsymbol{\sigma} &= [[\text{vech}(\Sigma_0)]_{\mathbb{CR}}^T, [\text{vech}(\Sigma_1)]_{\mathbb{CR}}^T]^T \in \mathbb{R}^{2p^2}, \end{aligned}$$

where $\text{vech}(\bullet)$ is the vectorisation operator with the upper triangular portion excluded.

The problem considered in this work is about the estimation of the unknown parameter $\boldsymbol{\theta} = [\boldsymbol{\sigma}^T, t_C]^T$, comprised of the covariance parameters as well as the change-point. We define $M = 2p^2 + 1$, the number of total unknown parameters.

Note that, in this work, the change-point t_C is assumed to be random and the covariance matrices are assumed to be deterministic. This is due to the fact that the tightest lower bound (with good regularity conditions) is the deterministic BB which has already been shown to exhibit too optimistic results in the context of change-point estimation [Ferrari and Tournieret, 2003, Rosa et al., 2010] contrary to the WWB. Concerning the covariance matrices, it is well known that the CRB is an efficient approximation [Kay, 2010]. This is why we consider such a hybrid bound which allows us to manage a trade-off between tightness and regularity conditions.

Finally, the distribution of the observations in this model denoted $p_{\mathbf{x}, t_C; \boldsymbol{\sigma}}(\mathbf{x}, t_C; \boldsymbol{\sigma}) = p_{\mathbf{x}|t_C; \boldsymbol{\sigma}}(\mathbf{x}|t_C; \boldsymbol{\sigma})p_{t_C}(t_C)$ where:

$$p_{\mathbf{x}|t_C; \boldsymbol{\sigma}}(\mathbf{x}|t_C; \boldsymbol{\sigma}) = \prod_{t=1}^{t_C} p_{\mathbf{S}_t; \Sigma_0}(\mathbf{S}_t; \Sigma_0) \prod_{t=t_C+1}^T p_{\mathbf{S}_t; \Sigma_1}(\mathbf{S}_t; \Sigma_1). \quad (4.10)$$

The aim is to obtain a lower bound on the performances of any estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta} \in \Theta = \mathbb{R}^{2p^2} \times \mathbb{N}$ using the hybrid CR/WWB. Note that, this does not consist only in the

concatenation of the CRB of the unknown covariances $\boldsymbol{\sigma}$ and the WWB of the unknown change-point t_C but in obtaining a matrix that considers also the coupling between the parameters. To this end, we consider the global MSE defined as follows:

$$\text{MSE}(\hat{\boldsymbol{\theta}}) = \mathbb{E}_{\mathbf{x}, t_C; \boldsymbol{\sigma}} \left\{ (\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})^T \right\}, \quad (4.11)$$

in which, $\mathbb{E}_{\mathbf{x}, t_C; \boldsymbol{\sigma}}$ denotes the expectation w.r.t the distribution $p_{\mathbf{x}, t_C; \boldsymbol{\sigma}}(\mathbf{x}, t_C; \boldsymbol{\sigma})$.

4.3.2 Covariance inequality

It has recently been shown in the literature [Ren et al., 2015] that eq. (4.11) can be bounded using the covariance inequality. In the context of our problem, we have the following proposition:

Proposition 4.3.1. Covariance Inequality

Let $\{\Psi_k(\mathbf{x}, \boldsymbol{\theta}) | k \in \llbracket 1, M \rrbracket\}$ be real-valued function set defined on $\mathbb{R}^{Tp^2} \times \Theta$ such that the following integral exists and satisfies for almost every (a.e.) $\mathbf{x} \in \mathbb{R}^{Tp^2}$, $\forall k \in \llbracket 1, M \rrbracket$, $\int_{\Theta} \Psi_k(\mathbf{x}, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta} = 0$. Then, using the definition of MSE at eq. (4.11), the following inequality holds:

$$\text{MSE}(\hat{\boldsymbol{\theta}}) \succeq \mathbf{V} \mathbf{P}^{-1} \mathbf{V}^T, \quad (4.12)$$

where $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semi-definite, \mathbf{V} is a $M \times M$ matrix whose elements are given by

$$(\mathbf{V})_{k,l} = \mathbb{E}_{\mathbf{x}, t_C; \boldsymbol{\sigma}} \left\{ \left((\hat{\boldsymbol{\theta}}(\mathbf{x}))_k - (\boldsymbol{\theta})_k \right) \Psi_l(\mathbf{x}, \boldsymbol{\theta}) \right\}, \quad (4.13)$$

$(\bullet)_{k,l}$ is the k -th line and l -th column of a matrix, and \mathbf{P} is a $M \times M$ matrix whose elements are given by

$$(\mathbf{P})_{k,l} = \mathbb{E}_{\mathbf{x}, t_C; \boldsymbol{\sigma}} \{ \Psi_k(\mathbf{x}, \boldsymbol{\theta}) \Psi_l(\mathbf{x}, \boldsymbol{\theta}) \}. \quad (4.14)$$

To obtain a hybrid CR/WWB, Ψ_k are chosen as follows:

$$\Psi_k(\mathbf{x}, \boldsymbol{\theta}) = \begin{cases} \Psi_k^{\text{CRB}}(\mathbf{x}, \boldsymbol{\theta}) & \text{for } k = 1, \dots, M-1 \\ \Psi_k^{\text{WWB}}(\mathbf{x}, \boldsymbol{\theta}) & \text{for } k = M \end{cases}, \quad (4.15)$$

where

$$\Psi_k^{\text{CRB}}(\mathbf{x}, \boldsymbol{\theta}) = \begin{cases} \frac{\partial \ln p_{\mathbf{x}, t_C; \boldsymbol{\sigma}}(\mathbf{x}, t_C; \boldsymbol{\sigma})}{\partial (\boldsymbol{\sigma})_k} & \text{if } \boldsymbol{\theta} \in \Theta' \\ 0 & \text{if } \boldsymbol{\theta} \notin \Theta' \end{cases} \quad (4.16)$$

and

$$\Psi_k^{\text{WWB}}(\mathbf{x}, \boldsymbol{\theta}) = \begin{cases} \frac{p_{\mathbf{x}, t_C; \boldsymbol{\sigma}}^s(\mathbf{x}, t_C + h; \boldsymbol{\sigma})}{p_{\mathbf{x}, t_C; \boldsymbol{\sigma}}^s(\mathbf{x}, t_C; \boldsymbol{\sigma})} - \frac{p_{\mathbf{x}, t_C; \boldsymbol{\sigma}}^{1-s}(\mathbf{x}, t_C - h; \boldsymbol{\sigma})}{p_{\mathbf{x}, t_C; \boldsymbol{\sigma}}^{1-s}(\mathbf{x}, t_C; \boldsymbol{\sigma})} & \text{if } \boldsymbol{\theta} \in \Theta' \\ 0 & \text{if } \boldsymbol{\theta} \notin \Theta' \end{cases} \quad (4.17)$$

for $s \in]0, 1[$, h is such that $t_C + h \in \llbracket 1, T-1 \rrbracket$ and $\Theta' = \{\boldsymbol{\theta} \in \Theta | f(\mathbf{x}, \boldsymbol{\theta}) > 0 \text{ a.e. } \mathbf{x} \in \mathbb{R}^{Tp^2}\}$.

Note that, any value for the terms h and s will lead to a lower bound on the MSE. However, they must be chosen cautiously in order to obtain the tightest bound. To obtain the tightest bound on the change-point, we have to compute:

$$\text{HCRWWB} = \sup_{h,s} \mathbf{V} \mathbf{P}^{-1} \mathbf{V}^T. \quad (4.18)$$

4.3.3 Hybrid bound for the change-point model

In the context of change-point estimation, the right-hand side of the inequality at eq. (4.12) can be obtained by using the semi closed-form expression provided in [Bacharach et al., 2019]:

$$\mathbf{V} = \begin{bmatrix} -\mathbf{I}_{2p^2} & \mathbf{0}_{2p^2, 1} \\ \mathbf{0}_{1, 2p^2} & v_{22} \end{bmatrix} \text{ and } \mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{12}^T & P_{22} \end{bmatrix}, \quad (4.19)$$

where the block-matrices are defined as follows:

- $\mathbf{P}_{11} = T/2 \operatorname{diag}(\mathbf{F}(\boldsymbol{\Sigma}_0), \mathbf{F}(\boldsymbol{\Sigma}_1))$, where $\mathbf{F}(\boldsymbol{\Sigma}_0)$ (resp. $\mathbf{F}(\boldsymbol{\Sigma}_1)$) is the Fisher information matrix with regards to $\boldsymbol{\Sigma}_0$ (resp. $\boldsymbol{\Sigma}_1$).

- $P_{22} = u(h) (\rho^{|h|} (\epsilon_h(2s)) + \rho^{|h|} (\epsilon_h(2s-1))) - 2u(2h)\rho^{|h|} (\epsilon_h(s))$, where

$$u(h) \triangleq \begin{cases} (T-1-|h|)/(T-1) & \text{if } |h| < T-1 \\ 0 & \text{otherwise} \end{cases},$$

$$\epsilon_h(s) = \begin{cases} s & \text{if } h > 0 \\ 1-s & \text{if } h < 0 \end{cases} \text{ and}$$

$$\rho(s) \triangleq \int_{\mathbb{S}_{\mathbb{H}}^p} p_{\mathbf{S}_t; \boldsymbol{\Sigma}_0}^s(\mathbf{S}_t; \boldsymbol{\Sigma}_0) p_{\mathbf{S}_t; \boldsymbol{\Sigma}_1}^{1-s}(\mathbf{S}_t; \boldsymbol{\Sigma}_1) d\mathbf{S}_t. \quad (4.20)$$

- $\mathbf{P}_{12} = [\mathbf{p}^T, \mathbf{q}^T]^T$, where the elements of vectors \mathbf{p} and \mathbf{q} are given by:

$$(\mathbf{p})_\ell = -hu(h)\rho^{|h|-1}(\epsilon_h(s))\phi_{\sigma_0, \ell}(\epsilon_h(s)),$$

$$(\mathbf{q})_\ell = hu(h)\rho^{|h|-1}(\epsilon_h(s))\phi_{\sigma_1, \ell}(\epsilon_h(s)),$$

and given $j \in \{0, 1\}$, $\ell \in \llbracket 1, p^2 \rrbracket$, $s \in]0, 1[$:

$$\phi_{\sigma_j, \ell}(s) \triangleq \int_{\mathbb{S}_{\mathbb{H}}^p} \frac{\partial \ln p_{\mathbf{S}_t; \boldsymbol{\Sigma}}(\mathbf{S}_t; \boldsymbol{\Sigma})}{\partial ([\operatorname{vech}(\boldsymbol{\Sigma})]_{\mathbb{C}\mathbb{R}})_\ell} \Big|_{\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_j} \times$$

$$p_{\mathbf{S}_t; \boldsymbol{\Sigma}_0}^s(\mathbf{S}_t; \boldsymbol{\Sigma}_0) p_{\mathbf{S}_t; \boldsymbol{\Sigma}_1}^{1-s}(\mathbf{S}_t; \boldsymbol{\Sigma}_1) d\mathbf{S}_t. \quad (4.21)$$

- $v_{22} = hu(h)\rho^{|h|}(\epsilon_h(s))$.

4.3.4 Derivation of $\mathbf{F}(\boldsymbol{\Sigma})$, $\rho(s)$ and $\phi_{\sigma_j, \ell}(s)$

In order to compute the bound, we finally need the closed-form expressions of $\mathbf{F}(\boldsymbol{\Sigma})$, $\rho(s)$ and $\phi_{\sigma_j, \ell}(s)$. We have:

Proposition 4.3.2. *The closed-form expression of $\rho(s)$ is given by:*

$$\rho(s) = \frac{|s\boldsymbol{\Sigma}_0^{-1} + (1-s)\boldsymbol{\Sigma}_1^{-1}|^{-N}}{|\boldsymbol{\Sigma}_0|^{sN} |\boldsymbol{\Sigma}_1|^{(1-s)N}}. \quad (4.22)$$

Proof. Derived in [Frery et al., 2014]. The result can be obtained by doing the substitution $\mathbf{Y} = \mathbf{ASA}$, with $\mathbf{A} = (s\boldsymbol{\Sigma}_0^{-1} + (1-s)\boldsymbol{\Sigma}_1^{-1})^{\frac{1}{2}}$, in the integral. \square

Proposition 4.3.3 (FIM of the Covariance for a \mathbb{CW} distribution). *For $\boldsymbol{\Sigma} \in \mathbb{S}_{\mathbb{H}}^p$, we have:*

$$\mathbf{F}(\boldsymbol{\Sigma}) = f_{\mathbb{CR}} \left(N \mathbf{D}_p^T (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{D}_p \right), \quad (4.23)$$

where \mathbf{D}_p is the duplication matrix defined for any matrix $\mathbf{S} \in \mathbb{C}^{p \times p}$, by

$$\mathbf{D}_p \text{vech}(\mathbf{S}) = \text{vec}(\mathbf{S}),$$

and $p_{\mathbb{CR}} : \mathbb{S}_{\mathbb{H}}^{p^2/2} \rightarrow \mathbb{S}_{++}^{p^2}$, where $\mathbb{S}_{++}^{p^2}$ is the set of real symmetric matrices, is defined as

$$f_{\mathbb{CR}}(\boldsymbol{\Sigma}) = \frac{1}{2} \begin{bmatrix} \Re(\boldsymbol{\Sigma}) & -\Im(\boldsymbol{\Sigma}) \\ \Im(\boldsymbol{\Sigma}) & \Re(\boldsymbol{\Sigma}) \end{bmatrix}.$$

Proof. The result comes by noticing that the score function w.r.t. the covariance matrix of a Wishart distribution is the same as the score function of a Gaussian distribution, and by an appropriate change of variable ($\mathbf{S}_t = \sum_{i=1}^N \mathbf{z}_i \mathbf{z}_i^H$ where $\{\mathbf{z}_i\}_{i=1,\dots,N} \sim \mathbb{CN}(\mathbf{0}_p, \boldsymbol{\Sigma})$, and \mathbb{CN} is the complex Normal distribution). The mapping $p_{\mathbb{CR}}$ allows obtaining its form in the real parametrisation. \square

Proposition 4.3.4. *The different terms of $\phi_{\sigma_j, \ell}(s)$ for $\ell \in \llbracket 1, p^2 \rrbracket$, $j \in \{0, 1\}$ are given by $\phi_{\sigma_j, \ell}(s) = ([\text{vech}(\boldsymbol{\Phi}_j(s))]_{\mathbb{CR}})_\ell$, where $\boldsymbol{\Phi}_j(s)$ is a $p \times p$ matrix given by:*

$$\boldsymbol{\Phi}_j(s) = N\rho(s)\boldsymbol{\Sigma}_j^{-1} (s\boldsymbol{\Sigma}_0^{-1} + (1-s)\boldsymbol{\Sigma}_1^{-1})^{-1} \boldsymbol{\Sigma}_j^{-1} - N\rho(s)\boldsymbol{\Sigma}_j^{-1}. \quad (4.24)$$

Proof. See Appendix 4.A.3. \square

Note: The extension of these results to non-Gaussian model becomes intricate due to the fact that $\rho(s)$, also known as the Rényi distance of order s , has yet been derived. Indeed, since in elliptical context, the distribution relies on the function g , it is not clear how a closed-form expression of the distance can be obtained.

4.3.5 Computation of the tightest bound

Let $\mathbf{h} = [h_1, \dots, h_K]^T \in \llbracket -T, T \rrbracket^K$ and $\mathbf{s} \in]0, 1[^K$. we want to compute the tightest bound which means that we need to compute the following quantity:

$$\mathbf{HCRWWB} = \sup_{\mathbf{s}, \mathbf{h}} W(\mathbf{h}, \mathbf{s}) = \mathbf{V} \mathbf{P}^{-1} \mathbf{V}^H. \quad (4.25)$$

Since it can be computationally expensive, we can choose $\mathbf{s} = 0.5 \times \mathbf{1}_{K,1}$. Indeed, this value seems in many cases to deliver the tightest bound [Weinstein and Weiss, 1988, Van Trees and Bell, 2007, Wen, 2001]. The maximum is obtained the Löwner-John ellipsoid as illustrated in Figure 4.10. This convex problem can be done by using standard toolbox such as cvx [Boyd and Vandenberghe, 2004] or more efficiently using Frank-Wolfe optimization methodology [Jambawalikar and Kumar, 2008]. We will, however, use a simple trace norm since these methodologies require heavy implementation procedure in order to be time-efficient.

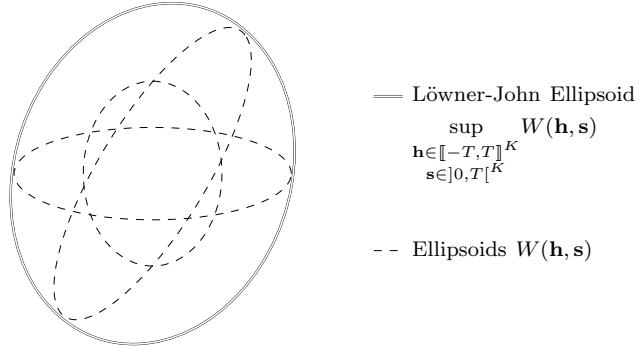


Figure 4.10: Illustration of the problem

4.3.6 Simulations

In order to validate the bound derived, Wishart time series subjected to a change-point as described in eq. (4.8) have been generated. t_C is generated using a uniform random prior and the covariance matrices have been chosen as Toeplitz matrices of the form: $(\Sigma_{k=0,1})_{i,j} = \alpha_k^{|i-j|}$.

Three estimators of the change-point have been considered:

- The Maximum A Posteriori (MAP) estimator which has the knowledge of the covariance matrices before and after the change:

$$\hat{t}_C = \underset{t_C \in [1, T-1]}{\operatorname{argmax}} p_{\mathbf{x}, t_C}(\mathbf{x}, t_C). \quad (4.26)$$

- The following hybrid estimator, derived from the Maximum A Posteriori/Maximum Likelihood [Yeredor, 2000], which estimates the covariance matrices before and after the change as well as the change point:

$$\hat{t}_C = \underset{t_C \in [1, T-1]}{\operatorname{argmax}} p_{\mathbf{x}, t_C; \hat{\sigma}}(\mathbf{x}, t_C; \hat{\sigma}), \quad (4.27)$$

where $\hat{\sigma} = \left[[\operatorname{vech}(\hat{\Sigma}_0)]_{\mathbb{CR}}^T, [\operatorname{vec}(\hat{\Sigma}_1)]_{\mathbb{CR}}^T \right]^T$ with:

$$\hat{\Sigma}_0 = \frac{1}{t_C N} \sum_{t=1}^{t_C} \mathbf{S}_t \text{ and } \hat{\Sigma}_1 = \frac{1}{(T-t_C)N} \sum_{t=t_C+1}^T \mathbf{S}_t.$$

- The iterative algorithm described in the first part of this chapter using Gaussian-derived statistics. Even though the time series is subjected to a change-point, since the algorithm relies on multiple hypothesis testing, there is a non null probability (although small) that all the tests would yield a rejection of the change hypothesis. When we do extensive Monte Carlo trials, this possibility arises. For the same reasons, when doing hypothesis tests, there is a probability that several hypotheses testing will yield a positive detection while we know that there is only one. Thus, the algorithm has been adapted to deliver always one change-point as follows:

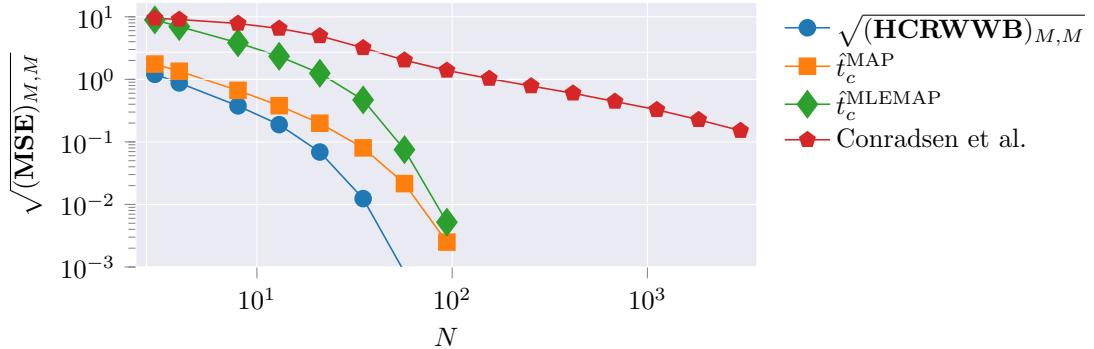


Figure 4.11: MSE on the change-point for $p = 3$, $T = 20$, $\alpha_0 = 0.1$, $\alpha_1 = 0.5$. The estimators curves have been computed with 4800 Monte Carlo trials.

- If no change is detected, a random value using the uniform prior is outputed.
- If there is more than one change point, the one having the highest ratio $\hat{\Lambda}$ /threshold is selected.

The false alarm rate has been fixed to 1%. It is expected, that this algorithm will have poor results but it is interesting to compare the performance with regards to the Bayesian schemes presented earlier.

First, let us consider Figure 4.11: the MSE of the estimated change-point for all estimators as well as the proposed bound has been plotted for several values of N . Without surprise, the MAP estimator has the best performance since it assumes the true knowledge of true change parameters. The derived bound behaves similarly to the MLEMAP estimator: the drop in the error for high N is well described by the bound. Although the gap between the bound and the hybrid estimator is rather high at low N , it decreases significantly for higher values. Since our aim is to design algorithms having good estimation performance, the region of interest is the one where the performance of the hybrid estimator is lower than those for 1 sample. In this region, the bound accurately describes the expected performance and can thus be used as an approximation for this peculiar estimator. Finally, the iterative algorithm has the poorest performance since it has to simultaneously detect and estimate the change-point. The gap between this estimator and the Bayesian ones increases significantly for high values of N , meaning that if a prior knowledge about the number of change-points is known, it might be better to consider the MLEMAP for the estimation problem.

The benefit of the bound is well described by Table 4.1 where the time-consumption¹ needed to compute the points of Figure 4.11 are given. Since the bounds need at least a thousand times less computational cost, it is preferable in a design context where many values for the parameters have to be tested.

In order to illustrate the usability of the bound in a design context, the bound has been computed for an extensive set of parameters (p, N) in Figure 4.12. Since for $p < N$, the empirical covariance matrices cannot be inverted, the bound is not usable that explains the lack of performance in this region. We can observe the drop in the error for any value of p at

¹The simulations were done on two Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30GHz processors.

Table 4.1: Time-consumption in seconds.

N	Bound	MAP	MLEMAP	Conradsen
3	0.524	468.349	469.125	1380.398
21	0.533	499.488	484.306	1467.237
57	0.521	499.442	499.466	1518.321
155	0.527	497.900	492.285	1620.447
254	0.518	494.653	494.678	1646.803
416	0.524	488.274	497.050	1638.840
682	0.530	528.987	511.411	1779.224
1118	0.526	500.287	581.645	1656.353
1831	0.526	654.252	674.176	3229.751
3000	0.526	755.564	743.614	3327.090

some point when increasing N . The dashed line which corresponds to the region at which the bound is accurate enough, allows to limit a region where, for a given situation with fixed p , an appropriate value for N can be chosen to guarantee good estimation performance.

An interesting observation arises through this simulation: when the dimension of vectors increase (assuming that all dimensions provide useful information), the number of samples required to obtain a given estimation performance decreases. For example, for $p = 3$ (polarimetric diversity), a window of 11×11 appear to be a good compromise while for $p = 10$, a window of 9×9 would suffice. This observation which could appear counter-intuitive is explained by the fact that as the dimension grows, the data before and after the change are more distant making it easier to discriminate.

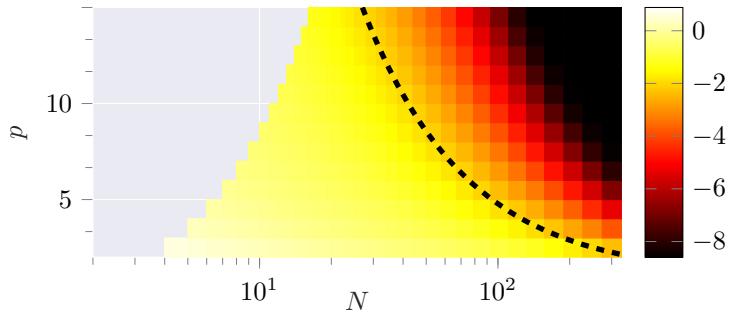


Figure 4.12: Evolution of $\log_{10} \sqrt{(\mathbf{HCRWWB})_{M,M}}$ for several parameters p and N , $T = 100$, $\alpha_0 = 0.1$ and $\alpha_1 = 0.3$. The dashed line corresponds to the region where $\sqrt{(\mathbf{HCRWWB})_{M,M}} = 10^{-2}$.

4.4 CONCLUSIONS

We have derived robust statistics that can be used in the algorithm designed by [Conradsen et al., 2016] and showed that they improved robustness of detection to some point. Study on real SDMS dataset showed better results for the algorithm associated with statistic $\hat{\Lambda}_{MT}$ (scale and shape) compared to the Gaussian and $\hat{\Lambda}_{Mat}$ statistics. This study also highlighted

that this approach is not robust to both co-registration and calibration problems of the time series. Moreover, depending on the coherence between the acquisitions testing the shape might not be an interesting approach since it results in too many false alarms.

We also considered the design of a lower bound in a Bayesian context, which allowed us to compare the estimation performance of the algorithm, in Gaussian case, with regards to more standard Bayesian estimations schemes. The Bayesian schemes perform better since they possess the knowledge of the number of change-points but they do not possess the properties of the recursive hypothesis algorithm, namely that we do not need to know the number of changes and can guarantee a false alarm rate if the distribution of data is adequate.

The lower bound also allowed us to consider the problem of tuning the number of samples needed to ensure good estimation performance. While not accurate when the number of samples is low, the distance between the bound and MSE decreases significantly at a high number of samples, which is the regime where a good estimation performance is possible. Since the bound is very time-efficient, it allowed to consider an extensive set of parameters (p, N) and showed that if p increases and assuming all dimensions provide useful information, the number of samples needed decreases.

A problem not considered in this chapter concerns the complexity of the proposed methodology relying on robust statistics as well as its online implementation capabilities. This point will be considered in the next chapter, where a novel approach based on Riemannian geometry will be developed to develop a recursive detection scheme.

4.A APPENDIX

4.A.1 Proof of proposition 4.1.1 at p. 140

Proposition. *The GLRT ratio under hypotheses of problem (3.9) for marginal scheme (4.1) is the following:*

$$\hat{\Lambda}_{\text{MT}}^{\text{marg}, t_1, t_2} = \frac{\left| \hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{MT}} \right|^{(t_2 - t_1)N}}{\left| \hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{MT}} \right|^{(t_2 - t_1-1)N} \left| \hat{\boldsymbol{\xi}}_T^{\text{TE}} \right|^N} \frac{((t_2 - t_1 - 1)p)^{(t_2 - t_1 - 1)Np} p^{Np}}{((t_2 - t_1)p)^{(t_2 - t_1)Np}} \times \\ \prod_{k=1}^N \frac{\left(\sum_{t=t_1}^{t_2} q(\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{MT}}, \mathbf{x}_k^t) \right)^{(t_2 - t_1)p}}{\left(\sum_{t=t_1}^{t_2-1} q(\hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{MT}}, \mathbf{x}_k^t) \right)^{(t_2 - t_1-1)p}} \left(q(\hat{\boldsymbol{\xi}}_T^{\text{TE}}, \mathbf{x}_k^T) \right)^p \stackrel{H_1}{\gtrless} \stackrel{H_0}{\lesssim} \lambda, \quad (4.28)$$

where

$$\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{MT}} = \frac{p}{N} \sum_{k=1}^N \frac{\sum_{t=t_1}^{t_2} \mathbf{S}_k^t}{\sum_{t=t_1}^{t_2} q(\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{Mat}}, \mathbf{x}_k^t)}. \quad (4.29)$$

Proof. For the marginal scheme, we have to compute the following GLRT:

$$\hat{\Lambda} = \frac{\max_{\boldsymbol{\theta}_{t_1, t_2-1}, \boldsymbol{\theta}_{t_2}} p_{\mathbb{W}_{t_1, t_2}}(\mathbb{W}_{t_1, t_2}; \boldsymbol{\theta}_{t_1, t_2-1}, \boldsymbol{\theta}_{t_2})}{\max_{\boldsymbol{\theta}_{t_1, t_2}} p_{\mathbb{W}_{t_1, t_2}}(\mathbb{W}_{t_1, t_2}; \boldsymbol{\theta}_{t_1, t_2})} \quad (4.30)$$

where $\boldsymbol{\theta}_{t_1, t_2} = \{\tau_1^{t_1, t_2}, \dots, \tau_N^{t_1, t_2}, \boldsymbol{\xi}_{t_1, t_2}\}$, $\boldsymbol{\theta}_{t_1, t_2-1} = \{\tau_1^{t_1, t_2-1}, \dots, \tau_N^{t_1, t_2-1}, \boldsymbol{\xi}_{t_1, t_2-1}\}$ and $\boldsymbol{\theta}_{t_2} = \{\tau_1^{t_2}, \dots, \tau_N^{t_2}, \boldsymbol{\xi}_{t_2}\}$.

Using the assumption that all observations are independent, we can rewrite:

$$\hat{\Lambda} = \frac{\max_{\boldsymbol{\theta}_{t_1, t_2-1}, \boldsymbol{\theta}_{t_2}} \prod_{k=1}^{k=N} \left(\prod_{t=t_1}^{t=t_2-1} p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^t; \boldsymbol{\theta}_{t_1, t_2-1}) \right) p_{\mathbf{x}_k^{t_2}}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^{t_2}; \boldsymbol{\theta}_{t_2})}{\max_{\boldsymbol{\theta}_{t_1, t_2}} \prod_{k=1}^{k=N} \prod_{t=t_1}^{t=t_2} p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^t; \boldsymbol{\theta}_{t_1, t_2})}.$$

Just as for the omnibus problem in 3.A.3, we optimise the numerator and denominator separately by plugging estimates in the likelihood functions:

$$\hat{\Lambda} = \frac{\mathcal{L}_1(\hat{\boldsymbol{\theta}}_{t_1, t_2-1}, \hat{\boldsymbol{\theta}}_{t_2})}{\mathcal{L}_0(\hat{\boldsymbol{\theta}}_{t_1, t_2})}, \quad (4.31)$$

where

$$\begin{aligned}\mathcal{L}_1(\boldsymbol{\theta}_{t_1,t_2-1}, \boldsymbol{\theta}_{t_2}) &= \prod_{k=1}^{k=N} \left(\prod_{t=t_1}^{t=t_2-1} p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^t; \boldsymbol{\theta}_{t_1,t_2-1}) \right) p_{\mathbf{x}_k^{t_2}}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^{t_2}; \boldsymbol{\theta}_{t_2}), \\ \mathcal{L}_0(\boldsymbol{\theta}_{t_1,t_2}) &= \prod_{\substack{k=1 \\ t=t_2}}^{k=N} p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^t; \boldsymbol{\theta}_{t_1,t_2}), \\ \hat{\boldsymbol{\theta}}_0 &= \operatorname{argmax}_{\boldsymbol{\theta}_{t_1,t_2}} \mathcal{L}_0(\boldsymbol{\theta}_{t_1,t_2}), \\ \hat{\boldsymbol{\theta}}_{t_1,t_2-1} &= \operatorname{argmax}_{\boldsymbol{\theta}_{t_1,t_2-1}} \mathcal{L}_1(\boldsymbol{\theta}_{t_1,t_2-1}, \boldsymbol{\theta}_{t_2}), \\ \hat{\boldsymbol{\theta}}_{t_2} &= \operatorname{argmax}_{\boldsymbol{\theta}_{t_2}} \mathcal{L}_1(\boldsymbol{\theta}_{t_1,t_2-1}, \boldsymbol{\theta}_{t_2}).\end{aligned}$$

We consider optimising \mathcal{L}_0 and \mathcal{L}_1 separately:

- For \mathcal{L}_0 , the problem is exactly the same as for the omnibus scheme presented in 3.A.3, if we remap the indices 1 to t_1 and T to t_2 .
- For \mathcal{L}_1 , we optimize alternatively for $\tau_k^{t_2}$, $\tau_k^{t_1,t_2-1}$, $\boldsymbol{\xi}_{t_2}$ and $\boldsymbol{\xi}_{t_1,t_2-1}$. We have:

$$\begin{aligned}\log \mathcal{L}_1 &= -\pi^{(t_2-t_1)Np} - (t_2-t_1-1)N \log |\boldsymbol{\xi}_{t_1,t_2-1}| - N \log |\boldsymbol{\xi}_{t_2}| \\ &\quad - (t_2-t_1-1)p \sum_{\substack{k=1 \\ t=t_1}}^{k=N} \log (\tau_k^t) - p \log (\tau_k^{t_2}) - \sum_{\substack{t=t_1 \\ k=1}}^{t=t_2-1} \frac{q(\boldsymbol{\xi}_{t_1,t_2-1}, \mathbf{x}_k^{t_1,t_2-1})}{\tau_k^{t_1,t_2-1}} \\ &\quad - \sum_{k=1}^{k=N} \frac{q(\boldsymbol{\xi}_{t_2}, \mathbf{x}_k^t)}{\tau_k^{t_2}}.\end{aligned}$$

Using the same optimisation procedure as omnibus scheme (taking the derivative and equalling it to 0), we obtain:

$$\begin{aligned}\hat{\tau}_k^{t_2} &= \frac{1}{p} q(\boldsymbol{\xi}_t, \mathbf{x}_k^{t_2}), \\ \hat{\tau}_k^{t_1,t_2-1} &= \frac{1}{(t_2-t_1-1)p} \sum_{t=t_1}^{t=t_2-1} q(\boldsymbol{\xi}_{t_1,t_2-1}, \mathbf{x}_k^t), \\ \hat{\boldsymbol{\xi}}_{t_2} &= \frac{1}{N} \sum_{k=1}^{k=N} \frac{\mathbf{S}_k^{t_2}}{\tau_k^{t_2}}, \\ \hat{\boldsymbol{\xi}}_{t_1,t_2-1} &= \frac{1}{(t_2-t_1-1)N} \sum_{\substack{k=1 \\ t=t_1}}^{k=N} \frac{\mathbf{S}_k^t}{\tau_k^{t_1,t_2-1}}, \text{ that we denote } \hat{\boldsymbol{\xi}}_{t_1,t_2-1}^{\text{MT}}.\end{aligned}\tag{4.32}$$

Finally, we have:

$$\begin{aligned}
\hat{\Lambda} &= \frac{\prod_{k=1}^{k=N} \left(\prod_{t=t_1}^{t=t_2-1} \frac{1}{\pi^p \left| \hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{MT}} \right| \left(\hat{\tau}_k^{t_1, t_2-1} \right)^p \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{MT}}, \mathbf{x}_k^t)}{\hat{\tau}_k^{t_1, t_2-1}} \right\}} \right)}{\prod_{k=1}^{k=N} \prod_{t=t_1}^{t=t_2} \frac{1}{\pi^p \left| \hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{MT}} \right| \left(\hat{\tau}_k^{t_1, t_2} \right)^p \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{MT}}, \mathbf{x}_k^t)}{\hat{\tau}_k^{t_1, t_2}} \right\}}} \times \\
&\quad \prod_{k=1}^{k=N} \frac{1}{\pi^p \left| \hat{\boldsymbol{\xi}}_{t_2}^{\text{TE}} \right| \left(\hat{\tau}_k^{t_2} \right)^p \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_{t_2}^{\text{TE}}, \mathbf{x}_k^{t_2})}{\hat{\tau}_k^{t_2}} \right\}} \\
&= \frac{\left| \hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{MT}} \right|^{(t_2-t_1)N}}{\left| \hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{MT}} \right|^{(t_2-t_1-1)N} \left| \hat{\boldsymbol{\xi}}_{t_2}^{\text{TE}} \right|^N} \prod_{k=1}^{k=N} \frac{\left(\hat{\tau}_k^{t_1, t_2} \right)^{(t_2-t_1)p}}{\left(\hat{\tau}_k^{t_1, t_2-1} \right)^{(t_2-t_1-1)p} \left(\hat{\tau}_k^{t_2} \right)^p} \times \\
&\quad \frac{\exp \left\{ - \sum_{k=1}^{t=t_2-1} \frac{q(\hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{MT}}, \mathbf{x}_k^t)}{\hat{\tau}_k^{t_1, t_2-1}} - \sum_{k=1}^N \frac{q(\hat{\boldsymbol{\xi}}_{t_2}^{\text{TE}}, \mathbf{x}_k^{t_2})}{\hat{\tau}_k^{t_2}} \right\}}{\exp \left\{ - \sum_{k=1}^{t=t_2} \frac{q(\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{MT}}, \mathbf{x}_k^t)}{\hat{\tau}_k^{t_1, t_2}} \right\}}
\end{aligned}$$

Now, if we replace the texture estimates by their expression, we have:

$$\hat{\Lambda}_{\text{MT}}^{\text{marg},t_1,t_2} = \frac{\left| \hat{\xi}_{t_1,t_2}^{\text{MT}} \right|^{(t_2-t_1)N}}{\left| \hat{\xi}_{t_1,t_2-1}^{\text{MT}} \right|^{(t_2-t_1-1)N} \left| \hat{\xi}_T^{\text{TE}} \right|^N} \frac{((t_2-t_1-1)p)^{(t_2-t_1-1)Np} p^{Np}}{((t_2-t_1)p)^{(t_2-t_1)Np}} \times \\ \prod_{k=1}^N \frac{\left(\sum_{t=t_1}^{t_2} q(\hat{\xi}_{t_1,t_2}^{\text{MT}}, \mathbf{x}_k^t) \right)^{(t_2-t_1)p}}{\left(\sum_{t=t_1}^{t_2-1} q(\hat{\xi}_{t_1,t_2-1}^{\text{MT}}, \mathbf{x}_k^t) \right)^{(t_2-t_1-1)p} \left(q(\hat{\xi}_T^{\text{TE}}, \mathbf{x}_k^T) \right)^p} \stackrel{H_1}{\gtrless} \stackrel{H_0}{\lesssim} \lambda,$$

□

Since the covariance estimate are solution to fixed-point equations, we do not replace them and have the final form of the statistic.

4.A.2 Proof of proposition 4.1.2 at p. 141

Proposition. *The GLRT ratio under hypotheses of problem (3.10) for marginal scheme (4.1) is the following:*

$$\hat{\Lambda}_{\text{Mat}}^{\text{marg},t_1,t_2} = \frac{\left| \hat{\xi}_{t_1,t_2}^{\text{Mat}} \right|^{(t_2-t_1)N}}{\left| \hat{\xi}_{t_1,t_2-1}^{\text{Mat}} \right|^{(t_2-t_1-1)N} \left| \hat{\xi}_T^{\text{TE}} \right|^N} \times \\ \prod_{k=1}^N \frac{\prod_{t=t_1}^{t_2} \left(q(\hat{\xi}_{t_1,t_2}^{\text{Mat}}, \mathbf{x}_k^t) \right)^p}{\left(\prod_{t=t_1}^{t_2-1} \left(q(\hat{\xi}_{t_1,t_2-1}^{\text{Mat}}, \mathbf{x}_k^t) \right)^p \right) \left(q(\hat{\xi}_T^{\text{TE}}, \mathbf{x}_k^T) \right)^p} \stackrel{H_1}{\gtrless} \stackrel{H_0}{\lesssim} \lambda, \quad (4.33)$$

where

$$\hat{\xi}_{t_1,t_2}^{\text{Mat}} = \frac{p}{N} \sum_{t=t_1}^{t_2} \sum_{k=1}^N \frac{\mathbf{S}_k^t}{q(\hat{\xi}_{t_1,t_2}^{\text{MT}}, \mathbf{x}_k^t)}. \quad (4.34)$$

Proof. For the marginal scheme, we have to compute the following GLRT:

$$\hat{\Lambda} = \frac{\max_{\boldsymbol{\theta}_{t_1,t_2-1}, \boldsymbol{\theta}_{t_2}, \boldsymbol{\Phi}_{t_1}, \dots, \boldsymbol{\Phi}_{t_2}} p_{\mathbb{W}_{t_1,t_2}}(\mathbb{W}_{t_1,t_2}; \boldsymbol{\theta}_{t_1,t_2-1}, \boldsymbol{\theta}_{t_2}, \boldsymbol{\Phi}_{t_1}, \dots, \boldsymbol{\Phi}_{t_2})}{\max_{\boldsymbol{\theta}_{t_1,t_2}, \boldsymbol{\Phi}_{t_1}, \dots, \boldsymbol{\Phi}_{t_2}} p_{\mathbb{W}_{t_1,t_2}}(\mathbb{W}_{t_1,t_2}; \boldsymbol{\theta}_{t_1,t_2}, \boldsymbol{\Phi}_{t_1}, \dots, \boldsymbol{\Phi}_{t_2})} \quad (4.35)$$

where $\boldsymbol{\theta}_{t_1,t_2} = \{\xi_{t_1,t_2}\}$, $\boldsymbol{\theta}_{t_1,t_2-1} = \{\xi_{t_1,t_2-1}\}$, $\boldsymbol{\theta}_{t_2} = \{\xi_{t_2}\}$ and $\forall t \in [t_1, t_2]$, $\boldsymbol{\Phi}_t = \{\tau_1^t, \dots, \tau_N^t\}$.

Using the assumption that all observations are independent, we can rewrite:

$$\hat{\Lambda} = \frac{\max_{\theta_{t_1,t_2-1}, \theta_{t_2}, \Phi_{t_1}, \dots, \Phi_{t_2}} \prod_{k=1}^{k=N} \left(\prod_{t=t_1}^{t=t_2-1} p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^t; \theta_{t_1,t_2-1}, \Phi_t) \right) p_{\mathbf{x}_k^{t_2}}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^{t_2}; \theta_{t_2}, \Phi_{t_2})}{\max_{\theta_{t_1,t_2}, \Phi_{t_1}, \dots, \Phi_{t_2}} \prod_{k=1}^{k=N} \prod_{t=t_1}^{t=t_2} p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^t; \theta_{t_1,t_2}, \Phi_t)}.$$

This expression can be computed by optimising the numerator and denominator separately just as done in all previous derivations:

$$\hat{\Lambda} = \frac{\mathcal{L}_1(\hat{\theta}_{t_1,t_2-1}, \hat{\theta}_{t_2}, \hat{\Phi}_{t_1}^1, \dots, \hat{\Phi}_{t_2}^1)}{\mathcal{L}_0(\hat{\theta}_0, \hat{\Phi}_{t_1}^0, \dots, \hat{\Phi}_{t_2}^0)}, \quad (4.36)$$

where

$$\begin{aligned} \mathcal{L}_1(\theta_{t_1,t_2-1}, \theta_{t_2}, \Phi_{t_1}, \dots, \Phi_{t_2}) &= \prod_{k=1}^{k=N} \left(\prod_{t=t_1}^{t=t_2-1} p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^t; \theta_{t_1,t_2-1}, \Phi_t) \right) p_{\mathbf{x}_k^{t_2}}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^{t_2}; \theta_{t_2}, \Phi_{t_2}), \\ \mathcal{L}_0(\theta_0, \Phi_{t_1}, \dots, \Phi_{t_2}) &= \prod_{k=1}^{k=N} \prod_{t=t_1}^{t=t_2} p_{\mathbf{x}_k^t}^{\mathbb{C}\mathcal{N}}(\mathbf{x}_k^t; \theta_{t_1,t_2}, \Phi_t), \\ \hat{\theta}_{t_1,t_2} &= \underset{\theta_{t_1,t_2}}{\text{argmax}} \mathcal{L}_0(\theta_{t_1,t_2}, \Phi_{t_1}, \dots, \Phi_{t_2}), \\ \forall t \in [t_1, t_2], \hat{\Phi}_t^0 &= \underset{\Phi_t}{\text{argmax}} \mathcal{L}_0(\theta_{t_1,t_2}, \Phi_{t_1}, \dots, \Phi_{t_2}), \\ \hat{\theta}_{t_1,t_2-1} &= \underset{\theta_{t_1,t_2-1}}{\text{argmax}} \mathcal{L}_1(\theta_{t_1,t_2-1}, \theta_{t_2}, \Phi_{t_1}, \dots, \Phi_{t_2}), \\ \hat{\theta}_{t_2} &= \underset{\theta_{t_2}}{\text{argmax}} \mathcal{L}_1(\theta_{t_1,t_2-1}, \theta_{t_2}, \Phi_{t_1}, \dots, \Phi_{t_2}), \\ \forall t \in [t_1, t_2], \hat{\Phi}_t^1 &= \underset{\Phi_t}{\text{argmax}} \mathcal{L}_1(\theta_{t_1}, \dots, \theta_{t_2}, \Phi_{t_1}, \dots, \Phi_{t_2}). \end{aligned}$$

Concerning the derivation of $\hat{\theta}_{t_1,t_2}$ and $\hat{\Phi}_{t_1,t_2}^0$, it has already been done in 3.A.4 by remapping date 1 to t_1 and date T to t_2 . We will refer to these estimates when putting a superscript 0 on the parameters.

We consider here the case for $\log \mathcal{L}_1$:

$$\begin{aligned} \log \mathcal{L}_1 &= -\pi^{(t_2-t_1)Np} - (t_2-t_1-1)N \log |\xi_{t_1,t_2-1}| - N \log |\xi_{t_2}| - p \sum_{\substack{k=1 \\ t=t_1}}^{k=N} \log (\tau_k^t) \\ &\quad - \sum_{\substack{k=N \\ t=t_1}}^{t=t_2-1} \frac{q(\xi_{t_1,t_2-1}, \mathbf{x}_k^{(t_1,t_2-1)})}{\tau_k^t} - \sum_{k=1}^{k=N} \frac{q(\xi_{t_2}, \mathbf{x}_k^t)}{\tau_k^{t_2}}. \end{aligned}$$

Optimizing using the same methodology as before, leads to:

$$\begin{aligned} \forall t \in \llbracket t_1, t_2 - 1 \rrbracket, \hat{\tau}_k^t &= \frac{1}{p} q(\boldsymbol{\xi}_{t_1, t_2-1}, \mathbf{x}_k^t), \text{ that we denote } \hat{\tau}_k^{t,1}, \\ \hat{\tau}_k^{t_2} &= \frac{1}{p} q(\boldsymbol{\xi}_{t_1, t_2-1}, \mathbf{x}_k^{t_2}), \text{ that we denote } \hat{\tau}_k^{t_2,1}, \\ \hat{\boldsymbol{\xi}}_{t_2} &= \frac{1}{N} \sum_{k=1}^{k=N} \frac{\mathbf{S}_k^{t_2}}{\tau_k^{t_2}}, \\ \hat{\boldsymbol{\xi}}_{t_1, t_2-1} &= \frac{1}{(t_2 - t_1 - 1)N} \sum_{\substack{k=1 \\ t=t_1}}^{t=t_2-1} \frac{\mathbf{S}_k^t}{\tau_k^t}, \text{ that we denote } \hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{Mat}}. \end{aligned} \quad (4.37)$$

Finally, we have:

$$\hat{\Lambda} = \frac{\prod_{k=1}^{k=N} \left(\prod_{t=t_1}^{t=t_2-1} \frac{1}{\pi^p |\hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{Mat}}| \left(\hat{\tau}_k^{t,1} \right)^p \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{Mat}}, \mathbf{x}_k^t)}{\hat{\tau}_k^{t,1}} \right\}} \right)}{\prod_{k=1}^{k=t_2} \frac{1}{\pi^p |\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{Mat}}| \left(\hat{\tau}_k^{t,0} \right)^p \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{Mat}}, \mathbf{x}_k^{t,0})}{\hat{\tau}_k^{t,0}} \right\}}} \times \frac{\prod_{k=1}^{k=N} \frac{1}{\pi^p |\hat{\boldsymbol{\xi}}_{t_2}^{\text{TE}}| \left(\hat{\tau}_k^{t_2,1} \right)^p \exp \left\{ -\frac{q(\hat{\boldsymbol{\xi}}_{t_2}^{\text{TE}}, \mathbf{x}_k^{t_2})}{\hat{\tau}_k^{t_2,1}} \right\}}}{}$$

Now, if we replace the texture estimates by their expression, we have:

$$\begin{aligned} \hat{\Lambda}_{\text{Mat}}^{\text{marg}, t_1, t_2} &= \frac{\left| \hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{Mat}} \right|^{(t_2 - t_1)N}}{\left| \hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{Mat}} \right|^{(t_2 - t_1 - 1)N} \left| \hat{\boldsymbol{\xi}}_T^{\text{TE}} \right|^N} \times \\ &\prod_{k=1}^N \frac{\prod_{t=t_1}^{t_2} \left(q(\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{Mat}}, \mathbf{x}_k^t) \right)^p}{\left(\prod_{t=t_1}^{t_2-1} \left(q(\hat{\boldsymbol{\xi}}_{t_1, t_2-1}^{\text{Mat}}, \mathbf{x}_k^t) \right)^p \right) \left(q(\hat{\boldsymbol{\xi}}_T^{\text{TE}}, \mathbf{x}_k^T) \right)^p} \stackrel{H_1}{\gtrless} \lambda, \end{aligned}$$

□

4.A.3 Proof of proposition 4.3.4 at p. 156

Proposition 4.A.1. *The different terms of $\phi_{\sigma_j, \ell}(s)$ for $\ell \in \llbracket 1, p^2 \rrbracket$, $j \in \{0, 1\}$ are given by $\phi_{\sigma_j, \ell}(s) = ([\text{vech}(\boldsymbol{\Phi}_j(s))]_{\mathbb{C}\mathbb{R}})_\ell$, where $\boldsymbol{\Phi}_j(s)$ is a $p \times p$ matrix given by:*

$$\boldsymbol{\Phi}_j(s) = N\rho(s)\boldsymbol{\Sigma}_j^{-1} \left(s\boldsymbol{\Sigma}_0^{-1} + (1-s)\boldsymbol{\Sigma}_1^{-1} \right)^{-1} \boldsymbol{\Sigma}_j^{-1} - N\rho(s)\boldsymbol{\Sigma}_j^{-1}. \quad (4.38)$$

Proof. Considering the complex nature of the matrices, we use the Wirtinger derivative [Wirtinger, 1927] in our derivations. The result is then obtained using the complex to real mapping. Then we remark that since the derivative in eq. (4.21) does not involve the variable of integration and the fact that the right-hand side of the integral is made of scalar terms, it is possible to compute the following expression:

$$\Phi_j(s) = \int_{\mathbb{S}_{\mathbb{H}}^p} \frac{\partial \ln p_{\mathbf{S}_t; \Sigma}(\mathbf{S}_t; \Sigma)}{\partial \Sigma} \Big|_{\Sigma=\Sigma_j} \times \\ p_{\mathbf{S}_t; \Sigma_0}^s(\mathbf{S}_t; \Sigma_0) p_{\mathbf{S}_t; \Sigma_1}^{1-s}(\mathbf{S}_t; \Sigma_1) d\mathbf{S}_t .$$

and we obtain our result by taking $\phi_{\sigma_j, \ell}(s) = ([\text{vech}(\Phi_j(s))]_{\mathbb{C}\mathbb{R}})_\ell$.

The Wirtinger derivative leads us to:

$$\frac{\partial \ln p_{\mathbf{S}_t; \Sigma'}(\mathbf{S}_t; \Sigma')}{\partial \Sigma'} = -N\Sigma'^{-1} + (\Sigma'^{-1} \mathbf{S}_t \Sigma'^{-1})^H .$$

Thus,

$$\begin{aligned} \Phi_j(s) &= -N\Sigma_j^{-1}\rho(s) + \Sigma_j^{-1} \left(\int_{\mathbb{S}_{\mathbb{H}}^p} \mathbf{S}_t \frac{|\mathbf{S}_t|^{N-p}}{\Gamma_p(N)|\Sigma_0|^{sN}|\Sigma_1|^{(1-s)N}} \right. \\ &\quad \times \text{etr}(-(s\Sigma_0^{-1} + (1-s)\Sigma_1^{-1})\mathbf{S}_t) d\mathbf{S}_t \left. \right) \Sigma_j^{-1}, \\ &= -N\Sigma_j^{-1}\rho(s) + \Sigma_j^{-1}\rho(s) \times \\ &\quad \left(\int_{\mathbb{S}_{\mathbb{H}}^p} \mathbf{S}_t \frac{|\mathbf{S}_t|^{N-p}}{\Gamma_p(N) \left| (s\Sigma_0^{-1} + (1-s)\Sigma_1^{-1})^{-1} \right|^N} \times \right. \\ &\quad \left. \text{etr}(-(s\Sigma_0^{-1} + (1-s)\Sigma_1^{-1})\mathbf{S}_t) d\mathbf{S}_t \right) \Sigma_j^{-1}, \\ &= -N\Sigma_j^{-1}\rho(s) + \Sigma_j^{-1}\rho(s) \mathbb{E}_{\mathbf{S}; \mathbf{A}} \{ \mathbf{S}_t \} \Sigma_j^{-1}, \end{aligned}$$

where $\mathbf{A} = (s\Sigma_0^{-1} + (1-s)\Sigma_1^{-1})^{-1}$. Finally using the expectation of a Wishart distribution, we obtain the result. \square

CHAPTER 5

RIEMANNIAN GEOMETRY AND ITS APPLICATION TO SAR PROBLEMS

Contents:

5.1	Some elements of Riemannian geometry	171
5.1.1	Riemannian geometry basics	171
5.1.2	Riemannian geometry over covariance matrices in Gaussian context .	172
5.1.3	The case of shape matrices	173
5.1.4	The case of texture parameters	174
5.2	Development of Riemannian geometry adapted to deterministic compound-Gaussian model	175
5.3	Recursive change-point detection	176
5.3.1	Online implementation in Gaussian context	176
5.3.2	Problems relative to an online implementation in robust case	177
5.3.3	Recursive Stochastic optimization in $\mathcal{M}_{p,N}$	178
5.3.4	Application to recursive change-point detection	179
5.3.5	Numerical illustrations	179
5.4	Clustering SAR image time series	182
5.4.1	Description of the problem	182
5.4.2	K-mean algorithm for clustering	183
5.4.3	Geometric mean on $\mathcal{M}_{p,N}$	183
5.4.4	Application to real SAR images	185
5.5	Conclusions	191
5.A	Proof of proposition 5.2.1 at p. 169	192
5.B	Proof of proposition 5.3.1 at p. 179	192

The work done in previous parts has shown that the deterministic compound-Gaussian model with its set of texture parameters has led to attractive results in change detection applications over Gaussian or \mathbb{CAE} models. This is due to the fact that the Gaussian model does not allow to consider a heterogeneity on the scattered power over the local observation window. On the other hand, while the \mathbb{CAE} brings welcome robustness to the detection, discarding the scale does not allow a good detection performance. As such the deterministic compound-Gaussian model has proven to be a good trade-off on real-time series change detection.

Despite those promising results, several issues can still be raised. On the one hand, the change-point detection algorithm presented in chapter 4 suffers from a computational high-complexity which becomes unmanageable when the number of images in the series is large. The complexity comes from the expression of the fixed point estimators used in the test:

$$\hat{\xi}_{t_1, t_2}^{\text{MT}} = \frac{p}{N} \sum_{k=1}^N \frac{\sum_{t=t_1}^{t_2} \mathbf{s}_k^t}{\sum_{t=t_1}^{t_2} q(\hat{\xi}_{t_1, t_2}^{\text{MT}}, \mathbf{x}_k^t)}. \quad (5.1)$$

When the number of images $t_2 - t_1 + 1$ grows, it requires that we keep in memory all the data corresponding to these dates and it requires that we compute the fixed-point from the beginning for each new image which grows in complexity since the sum goes from $t = t_1$ to $t = t_2$.

On the other hand, the binary nature of change detection can be a limit the interpretability of detected changes. Cyclicity means that a pixel could be subjected to seasonal changes such as a field which is always collected at the same time. Impulsivity would mean a pixel which change dramatically very fast such as the apparition of a building over a field. Nature of the objects subjected to change refers to the type of objects that the pixel represent.

It would be interesting to be able to cluster the time series while not putting the same class to a field, a car or a building while also taking into account the temporal evolution of those objects. That's what we are trying to do in this part. Indeed, from a practical standpoint, one might want to analyze the change patterns, such as cyclicity or impulsivity, or consider the nature of the objects subjected to change.

In order to consider those two aspects, we propose to introduce Riemannian geometry as a potential solution. Indeed, covariance matrices, which we considered up to this point, usually lie in a Riemannian manifold for which it is possible to consider the associated Riemannian geometry. In differential geometry as well as SAR litterature, this approach has been successfully integrated in online estimation and clustering problems. In our case, we will consider Riemannian geometry on the set of parameters (shape matrix and texture parameters) of the CCG distribution for which we will construct the associated geometry since it has not been done previously in the litterature.

Riemannian geometry has already been considered in subsection 3.2.4, when we treated the convergence properties of the novel fixed-point estimates through the concept of geodesic convexity. However, we only considered the manifold of PDH matrices. In the present chapter, we define a new manifold which considers the global structure of the CCG model parameters.

This work emerged as a collaboration with Florent Bouchard, a post-doc at LISTIC laboratory whose knowledge about Riemannian geometry has allowed us to consider the analysis of SAR time series through this novel point of view. The aim of this chapter is to present preliminary reflexions and results that have been considered in the last months of the thesis. As such, a large amount of work has yet to be done on this subject, which we let open to future extensions.

The chapter is organized as follows: We first recall classical results of Riemannian geometry over covariance matrices lying in $\mathbb{S}_{\mathbb{H}}^p$ and restrain it to the submanifold of unitary determinant matrices. We then consider as well the manifold of texture elements in $(\mathbb{R}^+)^N$. Then we define formally the new manifold which is obtained through the Cartesian product of these two manifolds and derive its Riemannian geometry. Finally, both clustering and

change-point detection problems are treated using the new tools developed and preliminary results are presented.

5.1 SOME ELEMENTS OF RIEMANNIAN GEOMETRY

5.1.1 Riemannian geometry basics

We introduce here briefly the diverse Riemannian geometry objects that will be considered in further developments. Far from being a formal introduction to the domain¹, our aim is to illustrate intuitively the concepts we will refer to from this point.

As described in [Absil et al., 2009] p.18, a d -dimensional manifold can be informally defined as a set \mathcal{M} which can be covered with a collection of coordinates patches that identify certain subsets of \mathcal{M} with open subsets of \mathbb{R}^d . In other terms, it is a topological space that locally resembles Euclidean space near each point. As an example, a Boy's surface has been plotted in Figure 5.1 for which a collection of patches in \mathbb{R}^3 has been plotted. We consider in this study, manifolds which lie in the ambient space \mathbb{M}_p , of matrices of size $p \times p$.

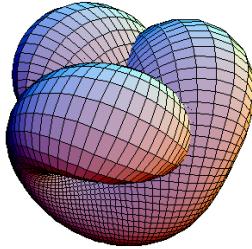


Figure 5.1: Example of manifold: Boy's surface in a 3-dimensional space. Credit: Maksim.

A Riemannian manifold is a manifold \mathcal{M} for which it is possible to define an inner product $\langle \bullet, \bullet \rangle$ on the tangent space $T_\Sigma \mathcal{M}$ at each point Σ that varies “smoothly” in the manifold. This property makes it possible to define several geometric notions such as:

- The orthogonal projection $P_\Sigma^\mathcal{M}$, which is an application from the ambient space \mathbb{M}_p , of matrices of size $p \times p$, into the tangent space $T_\Sigma \mathcal{M}$;
- The geodesic $\gamma^\mathcal{M} : \mathbb{R} \rightarrow \mathcal{M}$, which is a function which allows to move from one point $\gamma^\mathcal{M}(0) = \Sigma_0$ to another $\gamma^\mathcal{M}(1) = \Sigma_1$;
- The exponential mapping $\exp_\Sigma^\mathcal{M}$ which allows to obtain a point lying in the Manifold \mathcal{M} from a tangent vector in $T_\Sigma \mathcal{M}$;
- The logarithm mapping $\log_\mathcal{M}$, which is the inverse of $\exp_\Sigma^\mathcal{M}$, i.e. it is an application from \mathcal{M} onto $T_\Sigma \mathcal{M}$;
- The geodesic distance $\delta_\mathcal{M}$ between two points belonging to the manifold.

These various concepts are illustrated in Figure 5.2.

Riemannian geometry offers an alternative approach to more traditional Euclidean methodologies. One interesting point is that the metric in the tangent space of the manifold can be defined while taking into account the statistical properties of the objects considered. Say that $\boldsymbol{\theta} \in \mathcal{M}$ defines the parameters of a probability distribution $p_{\mathbf{x}}$, and suppose that \mathcal{M} defines a manifold. Then each point of the manifold can be mapped to a given distribution.

¹See [Absil et al., 2009] for an overview of Riemannian geometry over matrices.

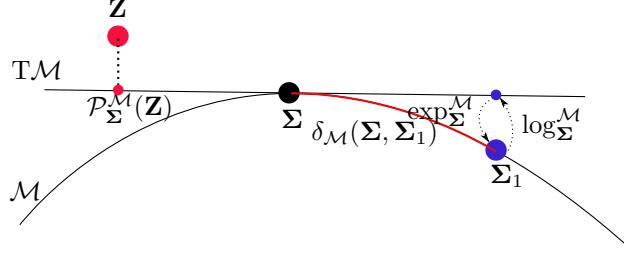


Figure 5.2: Illustration of Riemannian geometry concepts.

It is then possible to define the Fisher information metric [Smith, 2005], which is a metric that allows computing the informational difference between these distributions. For two close points in the manifold, the metric can be related to the Kullback-Leibler divergence between the distributions engendered by the two parameters. This metric can be equipped as a Riemannian metric in the tangent space of the manifold. Hence, this approach allows to consider the natural geometry associated with the manifold while also considering the probability model, which makes it attractive.

In the context of this thesis, our analysis is based on covariance matrices which lie in a Riemannian manifold. We will give hereafter examples of Riemannian geometry tools obtained using the Fisher information metric in a Gaussian context.

5.1.2 Riemannian geometry over covariance matrices in Gaussian context

Let us consider a set $\mathbf{X} = \{\mathbf{x}_k \in \mathbb{C}^p | k \in [1, N]\}$ i.i.d realizations of a Gaussian-distributed random variable \mathbf{x} with covariance parameter Σ . The log-likelihood of this set of observations is given by (up to constants):

$$\log \mathcal{L}(\mathbf{X}; \Sigma) = -N \text{Tr}\{\hat{\Sigma}^{\text{SCM}} \Sigma^{-1}\} + N \log |\Sigma|, \quad (5.2)$$

where $\hat{\Sigma}^{\text{SCM}} = \frac{1}{N} \sum_{k=1}^N \mathbf{x}_k \mathbf{x}_k^H$.

Given a positive definite Hermitian matrix $\Sigma \in \mathbb{S}_{\mathbb{H}}^p$ and the model presented in eq. (5.2), it is possible to define the Fisher information metric for two points $\zeta, \eta \in T_\Sigma \mathbb{S}_{\mathbb{H}}^p$ as [Skovgaard, 1984, Smith, 2005]:

$$\langle \zeta, \eta \rangle_{\Sigma}^{\mathbb{S}_{\mathbb{H}}^p} = \text{Tr}(\Sigma^{-1} \zeta \Sigma^{-1} \eta). \quad (5.3)$$

The Riemannian geometry associated with this metric is well-known². We have:

- The orthogonal projection $\mathcal{P}_{\Sigma}^{\mathbb{S}_{\mathbb{H}}^p}$ at $\Sigma \in \mathbb{S}_{\mathbb{H}}^p$ from $\mathbb{C}^{p \times p}$ to the tangent space $T_\Sigma \mathbb{S}_{\mathbb{H}}^p$:

$$\mathcal{P}_{\Sigma}^{\mathbb{S}_{\mathbb{H}}^p}(\mathbf{Z}) = \text{herm}(\mathbf{Z}), \quad (5.4)$$

where $\text{herm}(\mathbf{Z}) = \frac{1}{2}(\mathbf{Z} + \mathbf{Z}^H)$.

²See [Bhatia, 2009] for an extensive description.

- The geodesic between two points Σ_0 and Σ_1 , also denoted $\Sigma_0 \#_t^{\mathbb{S}_{\mathbb{H}}^p} \Sigma_1$, is given by:

$$\gamma^{\mathbb{S}_{\mathbb{H}}^p}(t) = \Sigma_0^{\frac{1}{2}} \left(\Sigma_0^{-\frac{1}{2}} \Sigma_1 \Sigma_0^{-\frac{1}{2}} \right)^t \Sigma_0^{\frac{1}{2}}, \quad (5.5)$$

where $(\bullet)^t$ corresponds to the power function over matrices defined through the exponential and logarithm matrix.

- The exponential mapping of $\eta \in T_{\Sigma} \mathbb{S}_{\mathbb{H}}^p$ is given by:

$$\exp_{\Sigma}^{\mathbb{S}_{\mathbb{H}}^p}(\mathbf{P}) = \Sigma^{\frac{1}{2}} \exp(\Sigma^{-\frac{1}{2}} \eta \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}}, \quad (5.6)$$

where $\exp(\bullet)$ is the matrix exponential.

- The logarithm mapping at $\Sigma_0 \in \mathbb{S}_{\mathbb{H}}^p$ of $\Sigma_1 \in \mathbb{S}_{\mathbb{H}}^p$ is given by:

$$\log_{\Sigma}^{\mathbb{S}_{\mathbb{H}}^p}(\Sigma_1) = \Sigma^{\frac{1}{2}} \log(\Sigma^{-\frac{1}{2}} \Sigma_1 \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}}, \quad (5.7)$$

where $\log(\bullet)$ is the matrix logarithm.

- The geodesic distance is given by:

$$\delta_{\mathbb{S}_{\mathbb{H}}^p}^2(\Sigma_0, \Sigma_1) = \|\log(\Sigma_0^{-\frac{1}{2}} \Sigma_1 \Sigma_0^{-\frac{1}{2}})\|_2^2. \quad (5.8)$$

Up to this point we considered the covariance matrix as a whole without taking into account the scale/shape parametrization which is primordial in the definition of the compound-Gaussian model. We recall briefly that this model is parametrized by a set of two parameters: the shape matrix ξ which is a PDH matrix normalized by either the trace or the determinant and a texture vector $\tau = [\tau_1, \dots, \tau_N]^T \in (\mathbb{R}^+)^N$. A set of observations $\mathbf{X} = \{\mathbf{x}_k \in \mathbb{C}^p | k \in [1, N]\}$ is said to follow a deterministic compound-Gaussian model if every \mathbf{x}_k are independent and we have:

$$\mathbf{x}_k \sim \mathcal{CN}(\mathbf{0}_p, \tau_k \xi) \quad (5.9)$$

In order to consider Riemannian geometry adapted to the deterministic compound-Gaussian model, we will first consider separately the space of shape matrices $\mathbb{S}_{\mathbb{H}, |\bullet|}^p = \{\xi \in \mathbb{S}_{\mathbb{H}}^p | |\xi| = 1\}$ and the space of texture parameters $(\mathbb{R}^+)^N$.

5.1.3 The case of shape matrices

Although we considered the normalization by the trace through this document as discussed in 1.3, we choose to consider the determinant one for this development. The reason for this change in normalization comes from the fact that, as discussed in [Hallin and Paindaveine, 2007], the Fisher information matrix towards the parameters of an elliptical distribution becomes block-diagonal, with no correlations between shape matrix and texture parameters, for a normalization by the determinant. Although this phenomenon was not of interest until now, it implies tremendous simplifications when it comes to the derivation of the Fisher information metric in the deterministic compound-Gaussian case. Thus, we will consider the manifold defined by: $\mathbb{S}_{\mathbb{H}, |\bullet|}^p$ defined as $\{\xi \in \mathbb{S}_{\mathbb{H}}^p : |\xi| = 1\}$.

Since in the Gaussian case, there is no change in the Fisher information metric if the covariance is normalized, we can equip $\mathbb{S}_{\mathbb{H}, |\bullet|}^p$ with the metric defined at eq. (5.3). Thus, we can define:

- The tangent space at a point ξ is the space:

$$T_\xi \mathbb{S}_{\mathbb{H}, |\bullet|}^p = \{\eta \in \mathbb{S}_{\mathbb{H}}^p : \text{Tr}(\xi^{-1}\eta) = 0\}. \quad (5.10)$$

- The orthogonal projection $\mathcal{P}_\xi^{\mathbb{S}_{\mathbb{H}, |\bullet|}^p}$ from $\mathbb{C}^{p \times p}$ to $T_\xi \mathbb{S}_{\mathbb{H}, |\bullet|}^p$ as:

$$\mathcal{P}_\xi^{\mathbb{S}_{\mathbb{H}, |\bullet|}^p}(\mathbf{Z}) = \text{herm}(\mathbf{Z}) - \frac{1}{p} \text{Tr}(\xi^{-1}\mathbf{Z})\xi. \quad (5.11)$$

Concerning the other quantities, $\mathbb{S}_{\mathbb{H}, |\bullet|}^p$ is a closed submanifold of $\mathbb{S}_{\mathbb{H}}^p$, i.e the geodesics are the same as for $\mathbb{S}_{\mathbb{H}}^p$. As a consequence, the distance, as well as the exponential and logarithm mappings, are the same as the previous case.

5.1.4 The case of texture parameters

We consider here the geometry of $(\mathbb{R}^+)^N$ which is the manifold of the vector of texture parameters τ of the deterministic compound-Gaussian model.

Let us first define the Hadamard product operator \odot on \mathbb{R}^N which is the elementwise product of two vectors. We define the notations $\odot_{k=1}^N \tau_i = \tau_1 \odot \dots \odot \tau_N$. We also define the Hadamard power of a vector $\tau^M = e^{M \log(\tau)}$, where \exp and \log are the elementwise exponential and logarithm operators. A useful property for further derivations is:

$$\langle \tau_0 \odot \tau_1, \tau_2 \rangle^{\mathbb{R}^N} = \langle \tau_0, \tau_2 \odot \tau_1 \rangle^{\mathbb{R}^N}, \quad (5.12)$$

where $\langle \bullet, \bullet \rangle^{\mathbb{R}^N}$ is the usual Euclidean inner product on \mathbb{R}^N .

It is interesting to notice that space \mathbb{R}^N , equipped with the operator \odot , is isomorphic to the space of diagonal matrices \mathbb{D}^N of size $N \times N$, equipped with the matricial product. Analogously, $(\mathbb{R}^+)^N$ can be linked to the space of diagonal matrices with positive elements \mathbb{D}_{++}^N , which is a closed sub-manifold of $\mathbb{S}_{\mathbb{H}}^p$. This allows to keep the Riemannian geometry described previously in subsection 5.1.2 for the case of texture parameters by implicitly taking the transformation $\tau \rightarrow \text{diag}(\tau)$.

As such, we will refer to the Riemannian geometry of $\mathbb{S}_{\mathbb{H}}^p$ when we use the metric $\langle \bullet, \bullet \rangle^{(\mathbb{R}^+)^N}$ as well as other Riemannian concepts on the space $(\mathbb{R}^+)^N$. We have following results:

- As $(\mathbb{R}^+)^N$ is an open of \mathbb{R}^N , its tangent space $T_\tau (\mathbb{R}^+)^N$ can be identified with \mathbb{R}^N .
- The natural Riemannian metric between $\zeta \in T_\tau (\mathbb{R}^+)^N$, $\eta \in T_\tau (\mathbb{R}^+)^N$ is:

$$\langle \zeta, \eta \rangle_\tau^{(\mathbb{R}^+)^N} = \langle \zeta \odot \tau^{-1}, \eta \odot \tau^{-1} \rangle^{\mathbb{R}^N}. \quad (5.13)$$

- The geodesic between $\tau_0, \tau_1 \in (\mathbb{R}^+)^N$ is:

$$\tau_0 \#_t^{(\mathbb{R}^+)^N} \tau_1 = \tau_0^{1-t} \odot \tau_1^t. \quad (5.14)$$

- The exponential mapping at point τ is given by:

$$\exp_\tau^{(\mathbb{R}^+)^N}(\zeta_\tau) = \tau \odot \exp(\tau^{-1} \odot \zeta_\tau), \quad (5.15)$$

where \exp is the pointwise exponential for vectors.

- The logarithm mapping is given by:

$$\log_{\tau_0}^{(\mathbb{R}^+)^N}(\tau_1) = \tau_0 \odot \log(\tau_0^{-1} \odot \tau_1), \quad (5.16)$$

where \log is the pointwise logarithm for vectors.

- The natural distance between two points τ_0 and τ_1 is given by:

$$\delta_{(\mathbb{R}^+)^N}^2(\tau_0, \tau_1) = \|\log(\tau_0^{-1} \odot \tau_1)\|_2^2. \quad (5.17)$$

5.2 DEVELOPMENT OF RIEMANNIAN GEOMETRY ADAPTED TO DETERMINISTIC COMPOUND-GAUSSIAN MODEL

Let us consider a set $\mathbf{X} = \{\mathbf{x}_k \in \mathbb{C}^p | k \in \llbracket 1, N \rrbracket\}$ independent realizations of a deterministic compound-Gaussian distribution with shape parameter ξ and texture $\tau = [\tau_1, \dots, \tau_N]^T$. The log-likelihood of this set of observations is given by (up to constants):

$$\begin{aligned} \log \mathcal{L}(\mathbf{X}; \theta) &= \sum_{k=1}^N \mathcal{L}_{\mathcal{G}}(\mathbf{x}_k; \tau_k \xi), \text{ where:} \\ \mathcal{L}_{\mathcal{G}}(\mathbf{x}_k; \Sigma) &= -\text{Tr}\{\mathbf{x}_k \mathbf{x}_k^H \Sigma^{-1}\} - \log |\Sigma|, \end{aligned} \quad (5.18)$$

where parameter $\theta = (\xi, \tau)$ of the deterministic compound-Gaussian model lies in the space $\mathcal{M}_{p,N} = \mathbb{S}_{\mathbb{H},|\bullet|}^p \times (\mathbb{R}^+)^N$ which defines a manifold. Indeed, a product of two usual manifolds defines itself a manifold as explained in [Absil et al., 2009]. Its tangent space $T_{\theta} \mathcal{M}_{p,N}$ at $\theta = (\xi, \tau)$ is simply given by:

$$T_{\theta} \mathcal{M}_{p,N} = T_{\xi} \mathbb{S}_{\mathbb{H},|\bullet|}^p \times T_{\tau} (\mathbb{R}^+)^p. \quad (5.19)$$

We have the following result concerning the Fisher information metric:

Proposition 5.2.1. *The Fisher information metric of the deterministic compound-Gaussian model over $\mathcal{M}_{p,N}$ is defined for $\theta = (\xi, \tau) \in \mathcal{M}_{p,N}$, $\zeta = (\zeta_{\xi}, \zeta_{\tau}) \in T_{\theta} \mathcal{M}_{p,N}$ and $\eta = (\eta_{\xi}, \eta_{\tau}) \in T_{\theta} \mathcal{M}_{p,N}$ by:*

$$\langle \zeta, \eta \rangle_{\theta}^{\mathcal{M}_{p,N}} = \langle \zeta_{\xi}, \eta_{\xi} \rangle_{\xi}^{\mathbb{S}_{\mathbb{H},|\bullet|}^p} + \langle \zeta_{\tau}, \eta_{\tau} \rangle_{\tau}^{(\mathbb{R}^+)^N}. \quad (5.20)$$

Proof. See appendix 5.A. □

It is peculiarly important to notice that the simple expression of the metric is obtained thanks to the normalization by the determinant constraint which allow discarding crossed terms between the two manifolds as in eq. (5.45). The metric obtained is attractive since the geometries of $\mathbb{S}_{\mathbb{H},|\bullet|}^p$ and $(\mathbb{R}^+)^N$ are both known.

From this result, we can derive the other Riemannian concepts by considering the geometry associated with $\mathbb{S}_{\mathbb{H},|\bullet|}^p$ for the shape matrix part and the geometry associated with $(\mathbb{R}^+)^N$ for the texture part of the parameters. We have:

- The projection on the tangent space is given by:

$$\mathcal{P}_{\theta}^{\mathcal{M}_{p,N}}(\mathbf{Z}_{\xi}, \mathbf{Z}_{\tau}) = \left(\mathcal{P}_{\xi}^{\mathbb{S}_{\mathbb{H},|\bullet|}^p}(\mathbf{Z}_{\xi}), \mathbf{Z}_{\tau} \right). \quad (5.21)$$

- The geodesic $\gamma^{\mathcal{M}_{p,N}} : \mathbb{R} \rightarrow \mathcal{M}_{p,N}$ between two points $\boldsymbol{\theta}_0 = (\boldsymbol{\xi}_0, \boldsymbol{\tau}_0) = \gamma^{\mathcal{M}_{p,N}}(0)$ and $\boldsymbol{\theta}_1 = (\boldsymbol{\xi}_1, \boldsymbol{\tau}_1) = \gamma^{\mathcal{M}_{p,N}}(1)$ belonging to $\mathcal{M}_{p,N}$ is given by:

$$\gamma^{\mathcal{M}_{p,N}}(t) = \boldsymbol{\theta}_0 \#_t^{\mathcal{M}_{p,N}} \boldsymbol{\theta}_1 = \left(\boldsymbol{\xi}_0 \#_t^{\mathbb{S}_{\mathbb{H},|\bullet|}^p} \boldsymbol{\xi}_1, \boldsymbol{\tau}_0 \#_t^{(\mathbb{R}^+)^N} \boldsymbol{\tau}_1 \right). \quad (5.22)$$

- The exponential mapping on $\mathcal{M}_{p,N}$ at a point $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\tau}) \in \mathcal{M}_{p,N}$ for a point $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_{\boldsymbol{\xi}}, \boldsymbol{\zeta}_{\boldsymbol{\tau}}) \in T_{\boldsymbol{\theta}}\mathcal{M}_{p,N}$:

$$\exp_{\boldsymbol{\theta}}^{\mathcal{M}_{p,N}}(\boldsymbol{\zeta}) = \left(\exp_{\boldsymbol{\xi}}^{\mathbb{S}_{\mathbb{H},|\bullet|}^p}(\boldsymbol{\zeta}_{\boldsymbol{\xi}}), \exp_{\boldsymbol{\tau}}^{(\mathbb{R}^+)^N}(\boldsymbol{\zeta}_{\boldsymbol{\tau}}) \right). \quad (5.23)$$

- The logarithm mapping on $\mathcal{M}_{p,N}$ at a point $\boldsymbol{\theta}_0 = (\boldsymbol{\xi}_0, \boldsymbol{\tau}_0) \in \mathcal{M}_{p,N}$ for a point $\boldsymbol{\theta}_1 = (\boldsymbol{\xi}_1, \boldsymbol{\tau}_1) \in \mathcal{M}_{p,N}$:

$$\log_{\boldsymbol{\theta}_0}^{\mathcal{M}_{p,N}}(\boldsymbol{\theta}_1) = \left(\log_{\boldsymbol{\xi}_0}^{\mathbb{S}_{\mathbb{H},|\bullet|}^p}(\boldsymbol{\xi}_1), \log_{\boldsymbol{\tau}_0}^{(\mathbb{R}^+)^N}(\boldsymbol{\tau}_1) \right). \quad (5.24)$$

- The natural distance between two points $\boldsymbol{\theta}_0 = (\boldsymbol{\xi}_0, \boldsymbol{\tau}_0)$ and $\boldsymbol{\theta}_1 = (\boldsymbol{\xi}_1, \boldsymbol{\tau}_1)$ belonging to $\mathcal{M}_{p,N}$ is given by:

$$\delta_{\mathcal{M}_{p,N}}^2 = \delta_{\mathbb{S}_{\mathbb{H},|\bullet|}^p}^2(\boldsymbol{\xi}_0, \boldsymbol{\xi}_1) + \delta_{(\mathbb{R}^+)^N}^2(\boldsymbol{\tau}_0, \boldsymbol{\tau}_1). \quad (5.25)$$

The Riemannian geometry developed here allows us to manipulate compound-Gaussian distributions for various purposes. For instance, this distance can be seen as an alternative to the statistic $\hat{\Lambda}_{MT}$ derived in chapter 3 for the case of two images. However, since there is no evident statistical property such as CFARness and given the fact it is limited to a pair of distributions, it is impractical for the sole purpose of change detection.

For others purposes such as online estimation and clustering, these tools may consist in a promising solution as we will see hereafter.

5.3 RECURSIVE CHANGE-POINT DETECTION

We consider here the problem of change-point detection/estimation as described in chapter 4. The algorithm developed to find the change-points relies on iterative hypothesis testing for which a statistic of decision has to be computed at each iteration. This statistic involves the data of the whole time series which makes it in principle highly-complex from an algorithmic point of view since at each iteration the statistic has to fetch all the data.

5.3.1 Online implementation in Gaussian context

In practice, depending on the statistic of decision, it can be factorized in functions of only one set of data at a time. Indeed, if consider the Gaussian GLRT omnibus and marginal statistics of decision defined as:

$$\hat{\Lambda}_{\mathcal{CN},\text{omni}}^{t_1, t_2} = \frac{\left| \hat{\Sigma}_{t_1, t_2}^{\text{SCM}} \right|^{(t_2 - t_1)N}}{\prod_{t=t_1}^{t_2} \left| \hat{\Sigma}_t^{\text{SCM}} \right|^N}, \text{ and}$$

$$\hat{\Lambda}_{\mathbb{C}\mathcal{N},\text{marg}}^{t_1,t_2} = \frac{\left| \hat{\Sigma}_{t_1,t_2}^{\text{SCM}} \right|^{(t_2-t_1)N}}{\left| \hat{\Sigma}_{t_2} \right|^N \left| \hat{\Sigma}_{t_1,t_2-1}^{\text{SCM}} \right|^{(t_2-t_1-1)N}},$$

where $\hat{\Sigma}_{t_1,t_2}^{\text{SCM}} = \frac{1}{t_2-t_1} \sum_{t=t_1}^{t_2} \hat{\Sigma}_t^{\text{SCM}}$ and $\hat{\Sigma}_t = \frac{1}{N} \sum_{k=1}^N \mathbf{x}_k^t \mathbf{x}_k^{t \text{ H}}$.

It can be noticed that both omnibus and marginal statistics can be implemented in an online fashion. Indeed, we have:

$$\hat{\Lambda}_{\mathbb{C}\mathcal{N},\text{omni}}^{t_1,t_2+1} = \hat{\Lambda}_{\mathbb{C}\mathcal{N},\text{omni}}^{t_1,t_2} \times \frac{\left| \left\{ (t_2-t_1) \hat{\Sigma}_{t_1,t_2}^{\text{SCM}} + \hat{\Sigma}_{t_2+1}^{\text{SCM}} \right\} / (t_2+1-t_1) \right|^{(t_2+1-t_1)N}}{\left| \hat{\Sigma}_{t_1,t_2}^{\text{SCM}} \right|^{(t_2-t_1)N} \left| \hat{\Sigma}_{t_2+1}^{\text{SCM}} \right|^N}, \text{ and } (5.26)$$

$$\hat{\Lambda}_{\mathbb{C}\mathcal{N},\text{marg}}^{t_1,t_2+1} = \hat{\Lambda}_{\mathbb{C}\mathcal{N},\text{marg}}^{t_1,t_2} \times \frac{\left| \hat{\Sigma}_{t_2} \right|^N \left| \hat{\Sigma}_{t_1,t_2-1}^{\text{SCM}} \right|^{(t_2-t_1-1)N}}{\left| \hat{\Sigma}_{t_1,t_2}^{\text{SCM}} \right|^{2(t_2-t_1)N} \left| \hat{\Sigma}_{t_2+1}^{\text{SCM}} \right|^N}. \quad (5.27)$$

$\hat{\Sigma}_{t_1,t_2}^{\text{SCM}}$ and $\hat{\Sigma}_{t_2}^{\text{SCM}}$ have already been computed for the previous iterations of the statistic so we only need to estimate $\hat{\Sigma}_{t_2+1}^{\text{SCM}}$ for the computation.

This means that in practice, say that we want to test a change-point of a new image in a time series of 1000 images, we do not have to iterate through the data of all the series in order to compute the statistics of decision.

5.3.2 Problems relative to an online implementation in robust case

While true for Gaussian statistics, the statistics derived under deterministic compound-Gaussian model $\hat{\Lambda}_{\text{MT}}$ cannot be implemented in an online fashion if the shape matrix estimates are computed using the MLE solution which requires the use of a fixed-point algorithm.

Indeed, unlike with SCM estimators, if $\hat{\xi}_{t_1,t_2}^{\text{MT}}$ is the solution of the fixed-point estimate between dates t_1 and t_2 , we do not have the strict equality:

$$\hat{\xi}_{t_1,t_2+1}^{\text{MT}} \neq \frac{(t_2-t_1)\hat{\xi}_{t_1,t_2}^{\text{MT}} + \hat{\xi}_{t_2+1}^{\text{MT}}}{t_2+1-t_1}. \quad (5.28)$$

Computing an iteration of the detector necessitate to re-compute the estimates using the fixed-point which means that the complexity grows with the number of dates ($t_2 - t_1$) and we must keep in memory all the data.

Although there is no strict equality, the arithmetic mean of the fixed-point estimates at each date is still expected to be somewhat close. The intuition comes from the fact that the asymptotic distribution of Tyler's estimate is known to be Gaussian with a fixed-variance. Thus taking an arithmetic mean of estimates with same asymptotic distributions should reduce considerably the variance if the number of dates is high. Thus in practice, it consists in a viable option.

On the other hand, another possibility to tackle this problem consists of stochastic optimization of the log-likelihood cost function. This approach aims at minimizing the cost-function by taking the gradient descent of the log-likelihood restricted to one subset of

Algorithm 4 Resursive estimation with decreasing step

```

1: Inputs:  $\{\mathbf{X}_{t_1}, \dots, \mathbf{X}_{t_2}\}$ , initialisation using MLE  $\theta_{t_1, t_1} = \{\hat{\boldsymbol{\xi}}_{t_1}^{\text{MT}}, \hat{\boldsymbol{\tau}}_{t_1}\} \in \mathcal{M}_{p,N}$ , initial
   step  $\alpha_0 > 0$ 
2: Outputs:  $(t_2 - t_1)$  iterates  $\{\hat{\theta}_{t_1, t_1+r} : r \in \llbracket 1, t_2 - t_1 \rrbracket\}$  in  $\mathcal{M}_{p,N}$ 
3: for  $r = 1, \dots, t_2 - t_1$  do
4:    $\hat{\theta}_{t_1, t_1+r} = \exp_{\hat{\theta}_{t_1, t_1+r-1}}^{\mathcal{M}_{p,N}} \left( \frac{\alpha_0}{r} \text{grad}_{\mathcal{M}_{p,N}} \log \mathcal{L}_{CG}(\mathbf{X}_t, \hat{\theta}_{t_1, t_1+r-1}) \right)$ 
5: end for

```

Algorithm 5 Resursive estimation with constant step

```

1: Inputs:  $\{\mathbf{X}_{t_1}, \dots, \mathbf{X}_{t_2}\}$ , initialisation using MLE  $\theta_{t_1, t_1} = \{\hat{\boldsymbol{\xi}}_{t_1}^{\text{MT}}, \hat{\boldsymbol{\tau}}_{t_1}\} \in \mathcal{M}_{p,N}$ , step
    $\gamma > 0$ 
2: Outputs:  $(t_2 - t_1)$  iterates  $\{\hat{\theta}_{t_1, t_1+r} : r \in \llbracket 1, t_2 - t_1 \rrbracket\}$  in  $\mathcal{M}_{p,N}$ 
3: for  $r = 1, \dots, t_2 - t_1$  do
4:    $\tilde{\theta}_{t_1, t_1+r} = \exp_{\tilde{\theta}_{t_1, t_1+r-1}}^{\mathcal{M}_{p,N}} \left( \gamma \text{grad}_{\mathcal{M}_{p,N}} \log \mathcal{L}_{CG}(\mathbf{X}_t, \hat{\theta}_{t_1, t_1+r-1}) \right)$ 
5:    $\hat{\theta}_{t_1, t_1+r} = \tilde{\theta}_{t_1, t_1+r-1} \#_{\frac{1}{r+1}}^{\mathcal{M}_{p,N}} \tilde{\theta}_{t_1, t_1+r}$ 
6: end for

```

the whole data and iterate those subsets. This approach is widely popular for distributed optimizations such as done in deep learning [Roy et al., 2018]. Recently, [Zhou and Said, 2018] has considered the stochastic optimization of parameters lying in Riemannian manifold and showed an asymptotic efficiency of the approach. Since, as we described in this chapter, the parameters of the log-likelihood of a deterministic compound-Gaussian model lie in a Riemannian manifold, it is interesting to study this methodology and compare it to the arithmetic mean presented before in eq. (5.28).

5.3.3 Recursive Stochastic optimization in $\mathcal{M}_{p,N}$

Let us consider $T = (t_2 - t_1)$ group of observations $\{\mathbf{X}_t | t \in \llbracket t_1, t_2 \rrbracket\}$ following the same deterministic compound-Gaussian model parametrized by $\boldsymbol{\theta}_{t_1, t_2}$ compromised of shape matrix $\boldsymbol{\xi}_{t_1, t_2}$ and texture $\boldsymbol{\tau}_{t_1, t_2} = [\tau_1, \dots, \tau_N]^T$:

$$\forall (k, t) \in \llbracket 1, N \rrbracket \times \llbracket t_1, t_2 \rrbracket, \mathbf{x}_k^t \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, \tau_k \boldsymbol{\xi}_{t_1, t_2}). \quad (5.29)$$

The global log-likelihood towards all the observations $\{\mathbf{x}_k^t | (k, t) \in \llbracket 1, N \rrbracket \times \llbracket t_1, t_2 \rrbracket\}$ is given by:

$$\begin{aligned} \log \mathcal{L}_{\text{glob}}(\mathbf{X}_{t_1}, \dots, \mathbf{X}_{t_2}; \boldsymbol{\theta}_{t_1, t_2}) &= \sum_{t=t_1}^{t_2} \log \mathcal{L}_{CG}(\mathbf{X}_t, \boldsymbol{\theta}_{t_1, t_2}), \text{ where:} \\ \log \mathcal{L}_{CG}(\mathbf{X}_t, \boldsymbol{\theta}_{t_1, t_2}) &= \sum_{k=1}^N \mathcal{L}_{\mathcal{G}}(\mathbf{x}_k; \tau_k \boldsymbol{\xi}_{t_1, t_2}). \end{aligned} \quad (5.30)$$

The MLE estimation of parameter $\boldsymbol{\theta}_{t_1, t_2}$ has been tackled in Chapter 3 for the computation of the GLRT. Considering stochastic optimization, following the methodology of [Zhou and Said, 2018], a recursive optimization of parameter $\boldsymbol{\theta}_{t_1, t_2}$ can be obtained through either algorithm 4 or 5, in which $\log \mathcal{L}_{CG}$ is defined at eq. (5.18).

In order to implement the algorithms, we need to compute the gradient of $\log \mathcal{L}_{\mathcal{CG}}$, the log-likelihood of one set of observations towards the parameter $\boldsymbol{\theta}$. We have the following result:

Proposition 5.3.1. *The Riemannian gradient of $\log \mathcal{L}_{\mathcal{CG}}(\mathbf{X}_t; \boldsymbol{\theta})$ at $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\tau})$ is:*

$$\text{grad}_{\mathcal{M}_{p,N}} \log \mathcal{L}_{\mathcal{CG}}(\mathbf{X}_t; \boldsymbol{\theta}) = \left(\mathcal{P}_{\mathbb{H}, |\bullet|}^{\mathbb{S}^p} \left(\sum_{k=1}^N \tau_k^{-1} \mathbf{x}_k^t \mathbf{x}_k^{t^H} \right), \mathbf{g}_{\boldsymbol{\tau}} \right), \quad (5.31)$$

where for $1 \leq k \leq K$, the k -th element of $\mathbf{g}_{\boldsymbol{\tau}}$ is $(\mathbf{g}_{\boldsymbol{\tau}})_k = \text{Tr}(\boldsymbol{\xi}^{-1} \mathbf{x}_k^t \mathbf{x}_k^{t^H}) - \tau_k p$.

Proof. See appendix 5.B. \square

5.3.4 Application to recursive change-point detection

Computing $\hat{\Lambda}_{\mathcal{CCG}, \text{omni}}^{t_1, t_2}$ and $\hat{\Lambda}_{\mathcal{CCG}, \text{marg}}^{t_1, t_2}$ in an online fashion can be done by considering the instantaneous likelihood-ratio as follows:

$$\hat{\Lambda}_{\mathcal{CCG}, \text{omni}}^{t_1, t_2+1} = \hat{\Lambda}_{\mathcal{CCG}, \text{omni}}^{t_1, t_2} \times \frac{\mathcal{L}(\mathbf{X}_{t_2+1}; \hat{\boldsymbol{\tau}}_{t_2+1}^{\text{mle}}, \hat{\boldsymbol{\xi}}_{t_1, t_2+1}^{\text{mle}})}{\mathcal{L}(\mathbf{X}_{t_2+1}; \hat{\boldsymbol{\tau}}_{t_1, t_2+1}^{\text{rec}}, \hat{\boldsymbol{\xi}}_{t_1, t_2+1}^{\text{rec}})} \quad (5.32)$$

$$\hat{\Lambda}_{\mathcal{CCG}, \text{marg}}^{t_1, t_2+1} = \hat{\Lambda}_{\mathcal{CCG}, \text{marg}}^{t_1, t_2} \times \frac{\mathcal{L}(\mathbf{X}_{t_2}; \hat{\boldsymbol{\tau}}_{t_1, t_2}^{\text{rec}}, \hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{rec}}) \mathcal{L}(\mathbf{X}_{t_2+1}; \hat{\boldsymbol{\tau}}_{t_2+1}^{\text{mle}}, \hat{\boldsymbol{\xi}}_{t_2+1}^{\text{mle}})}{\mathcal{L}(\mathbf{X}_{t_2}; \hat{\boldsymbol{\tau}}_{t_2}^{\text{mle}}, \hat{\boldsymbol{\xi}}_{t_2}^{\text{mle}}) \mathcal{L}(\mathbf{X}_{t_2+1}; \hat{\boldsymbol{\tau}}_{t_1, t_2+1}^{\text{rec}}, \hat{\boldsymbol{\xi}}_{t_1, t_2+1}^{\text{rec}})} \quad (5.33)$$

where the superscript mle indicates an estimation using MLE while rec indicates a recursive estimator which is either the Euclidian arithmetic mean or the recursive stochastic estimation.

5.3.5 Numerical illustrations

Methodologies considered

In order to compare the attractiveness of the Riemannian optimization strategy, we compare it to various methodologies:

- The MLE computed trough fixed-point equation:

$$\begin{aligned} \hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{mle}} &= \frac{p}{N} \sum_{k=1}^N \frac{\sum_{t=t_1}^{t_2} \mathbf{s}_k^t}{\sum_{t=t_1}^{t_2} q(\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{mle}}, \mathbf{x}_k^t)} \\ \hat{\boldsymbol{\tau}}_{t_1, t_2}^{\text{mle}} &= \frac{1}{p} \sum_{t=t_1}^{t_2} q(\hat{\boldsymbol{\xi}}_{t_1, t_2}^{\text{mle}}, \mathbf{x}_k^t) \end{aligned} \quad (5.34)$$

- A recursive Euclidean arithmetic mean of the MLE at each date:

$$\begin{aligned}\hat{\boldsymbol{\xi}}_{t_1}^{\text{rec,euc}} &= \hat{\boldsymbol{\xi}}_{t_1}^{\text{mle}} \\ \hat{\boldsymbol{\tau}}_{t_1}^{\text{rec,euc}} &= \hat{\boldsymbol{\tau}}_{t_1}^{\text{mle}} \\ \forall t > t_1, \hat{\boldsymbol{\xi}}_{t_1,t}^{\text{rec,euc}} &= \frac{(t-t_1)\hat{\boldsymbol{\xi}}_{t_1,t-1}^{\text{rec,euc}} + \hat{\boldsymbol{\xi}}_t^{\text{mle}}}{t-t_1+1} \\ \forall t > t_1, \hat{\boldsymbol{\tau}}_{t_1,t}^{\text{rec,scm}} &= \frac{(t-t_1)\hat{\boldsymbol{\tau}}_{t_1,t}^{\text{rec,euc}} + p^{-1}q\left(\hat{\boldsymbol{\xi}}_{t_1,t}^{\text{rec,euc}}, \mathbf{x}_k^t\right)}{t-t_1+1}\end{aligned}\quad (5.35)$$

- The recursive estimation thanks to algorithms 4 and 5 that we denotes $\hat{\boldsymbol{\xi}}^{\text{rec,rie}}, \hat{\boldsymbol{\tau}}^{\text{rec,rie}}$.

Data generation

A synthetic time series has been generated where the observations are of the form $\mathbf{x}_k^t = \sqrt{\tau_k^t} \tilde{\mathbf{x}}$ for $(k, t) \in [\![1, N]\!] \times [\![1, T]\!]$, where $\tilde{\mathbf{x}} \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, \boldsymbol{\Sigma})$. We impose $\forall k \in [\![1, N]\!] \forall (t, t') \in [\![1, T]\!], \tau_k^t = \tau_k^{t'}$ (texture equality constraint) and we generate τ_k^1 as a realization of $\Gamma(\alpha, \beta)$, where $\beta = 1/\alpha$.

The covariance matrices are chosen to be Toeplitz of the form $(\boldsymbol{\xi})_{m,n} = \rho^{|m-n|}$.

Results

Figure 5.3 shows the results of optimization error in natural distance. The estimation has been done between $t_1 = 1$ and $t_2 = t_1 + r$ for $r \in [\![1, T]\!]$. The time needed to compute the estimates has also been given. Although this measure is highly dependent of the machine considered, it gives a first insight about time consumption of the different methodologies.

Several observations can be made from this preliminary result:

- The MLE is the best estimator in all regards which is expected since it is the true maximum of the log-likelihood. However, the complexity grows with r making it unattractive for online schemes.
- The arithmetic mean of fixed points appears to be a good alternative since the performance of estimation are near the MLE even at a finite distance for both shape and texture estimation.
- The recursive Riemannian methodology with constant step appears to be unstable since the error on shape matrix grows at some point. It seems that the estimate distance itself from the optimum due to an inappropriate tuning of the step.
- The recursive Riemannian methodology with constant step appears to work better. However, the choice of the step has made a good estimation of the shape while it appears too small for the textures. Indeed, we can choose a greater step to improve texture estimation but this result in numerical instability concerning shape matrix. This means that it would be advantageous to choose the steps separately for shape and texture. This method appears to be somewhat faster than the arithmetic mean one which is advantageous in an online setup.
- As an alternative to choosing the steps separately, we considered a composite estimator where the Riemannian method has been used for shape estimation while the textures are computed through an arithmetic mean of quadratic forms as for the recursive Euclidean counter-part. This methodology seems to obtain asymptotically the same performance than the MLE just as the Euclidean arithmetic mean which is an

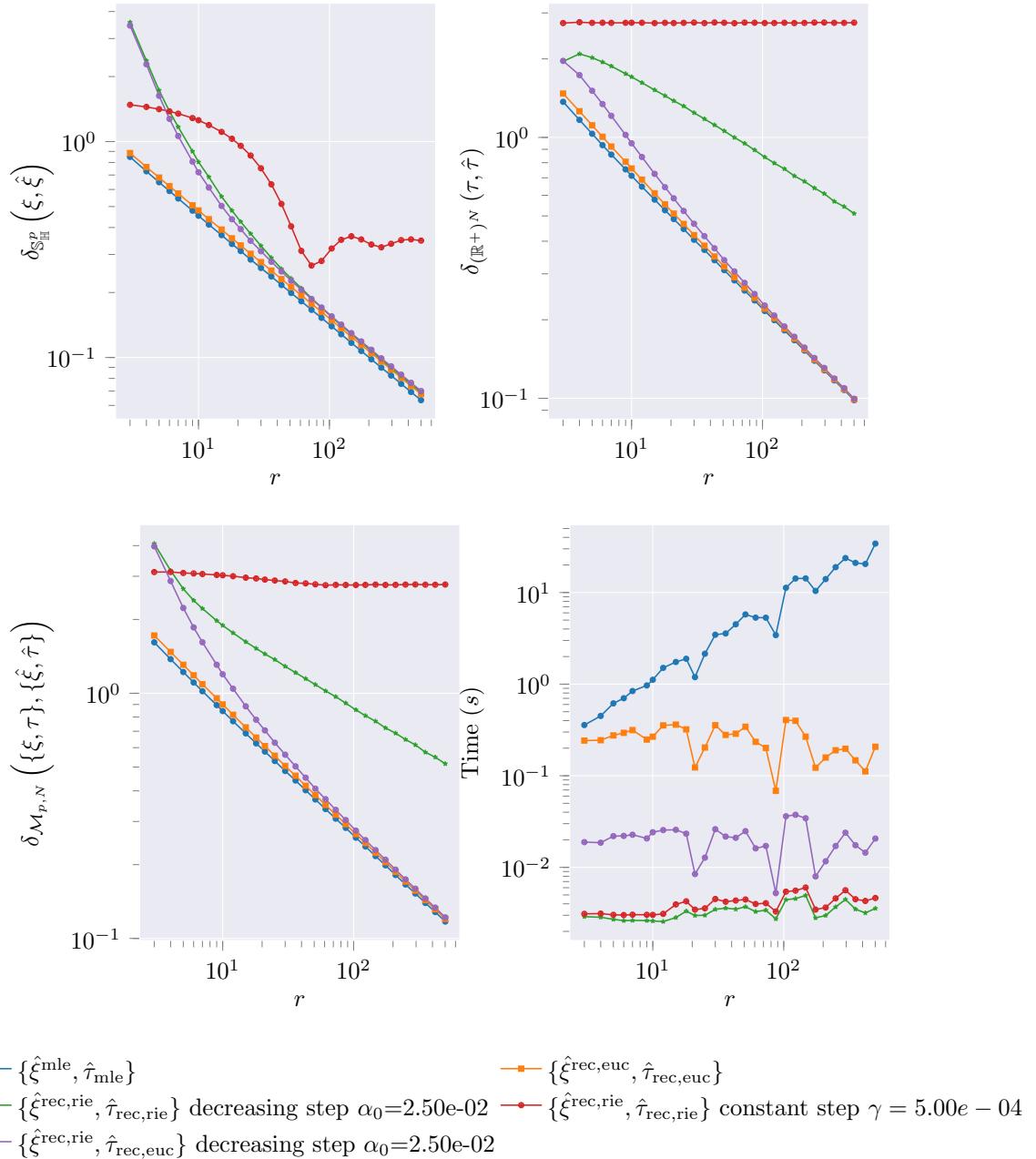


Figure 5.3: Top-left: Natural distance to shape matrix. Top-Right: Natural distance to texture parameters. Bottom-left: Natural compound-Gaussian distance. Bottom-Right: Mean time of computation over the Monte-Carlo trials. Parameters: $p = 10$, $N = 49$, $\rho = 0.71 + i * 0.71$, $\alpha = 0.01$, $\beta = 100.00$, $T = 400$, 4800 Monte-Carlo Trials.

improvement. The time needed to compute the estimate appears to slightly lower as well for this composite methodology.

From those preliminary results, it would appear that the recursive Euclidean methodology is a good candidate for online implementation of the robust change-point algorithm. The composite Riemannian methodology has good asymptotic performance in this simulation, making it a valid alternative for a large number of images.

These results were done without taking an appropriate time for the tuning of the step in Riemannian methodologies. Moreover, these results were done on a limited set of parameters due to time constraints and must thus be still validated.

Moreover, the estimation is not the sole criterion to assess the quality of the different methodologies. Indeed, from a detection perspective, the analysis of the online robust statistics with different estimation schemes has still to be done. It would be interesting to consider the CFARness property and consider the detection performance as well.

5.4 CLUSTERING SAR IMAGE TIME SERIES

In this last section, we consider a clustering problem for which Riemannian geometry has potential applications. We will describe shortly hereafter, the problem and its applications. Then we will describe the k-means algorithm, which has been considered for the clustering, and how Riemannian geometry can be plugged into the process. Finally, we will present preliminary results on real SAR data.

5.4.1 Description of the problem

Let us consider a time series of T SAR images for which we want to consider automatic clustering of objects according to their physical properties (polarimetric, or spectro-angular properties among others) or their temporal evolution.

When it concerns the clustering in such time series, two problems can be considered:

- Problem (1): Global clustering of the data without taking into account the time dimension. In this case, features are selected for each pixel at each spatial and time location, and a clustering algorithm allows to discriminate between classes. For SAR images, covariance matrices estimated using a local spatial neighborhood is often chosen as a feature for the clustering [Rignot et al., 1992, Formont et al., 2013, Reigber et al., 2010, Anfinsen et al., 2007].
- Problem (2): Time Series clustering in which, for each spatial location, the time variation of the vectors are taken into account for the clustering. For this case, distances adapted to time-series comparison such as dynamic time-warping [Bankó and Abonyi, 2012] have been adapted to work in time-series clustering. Otherwise, computing the sample covariance matrix of the data along the time dimension has also been considered to obtain a feature representing the whole time series [Hallac et al., 2017, Frambourg et al., 2013].

We will consider, in this preliminary work, these two problems by using the Riemannian geometry tools developed in the present chapter. The idea is to consider features coming from a robust estimation of covariance and textures in a compound-Gaussian model and consider the underlying Riemannian geometry in the clustering methodology. We have indeed seen that those parameters have been relevant to compare data across time for change detection. Thus, it would be expected that they would improve clustering with regards to covariance-only based methods.

Algorithm 6 General implementation of the algorithm with arbitrary distance d and mean function f_{μ}

```

Initialise Centers of classes  $\{\boldsymbol{\mu}_k^{(0)}\}_{1 \leq k \leq K}$ 
Compute features  $\hat{\mathbf{y}}_i$ 
while  $\max_k d(\boldsymbol{\mu}_k^{(iter)}, \boldsymbol{\mu}_k^{(iter-1)}) > \delta$  or  $iter \leq iter\_max$  do
    for  $i = 1 \dots N_{px}$  do
        for  $k = 1 \dots K$  do
            Compute  $d(\boldsymbol{\mu}_k^{(iter)}, \hat{\mathbf{y}}_i)$ 
        end for
        Assign sample  $i$  to class  $\operatorname{argmin}_k d(\boldsymbol{\mu}_k^{(iter)}, \hat{\mathbf{y}}_i)$ 
    end for
    for  $k = 1 \dots K$  do
        Update mean of class  $t$ :  $\boldsymbol{\mu}_k^{(iter)} = f_{\mu}(\{\hat{\mathbf{y}}_i : i \in \mathbf{C}_t\})$ 
    end for
    iter=iter+1
end while

```

5.4.2 K-mean algorithm for clustering

To cluster the images, we consider the use of the well-known k-means algorithm [MacQueen et al., 1967] which has been frequently considered in SAR clustering literature [Zhang et al., 2015, Doulgeris et al., 2008, Pallotta et al., 2019]. The methodology presented at Algorithm 6, relies on two functions: a distance and a mean function. The idea is to compute means of data belonging to the same class, then assign for each point the class corresponding to the closest (using the distance function) mean and iterate until stability of assigned classes.

Concerning problem (1), we compute covariance and textures parameters for each pixel using the spatial neighborhood. While for the problem (2), we compute for each pixel, the covariance and texture parameters along the time dimension. The distance used correspond to the natural distance on $\mathcal{M}_{p,N}$ given at eq. (5.25). While for the mean, we consider the geometric mean of points in $\mathcal{M}_{p,N}$ which will be detailed later in subsection 5.4.3.

This methodology assumes the knowledge of the number of classes, which can be a limiting factor since for an arbitrary image, the number of relevant classes to discriminate the objects present in the scene. To tackle this issue, we consider, for the problem (1), the $H - \alpha$ decomposition of polarimetric images [Cloude and Pottier, 1997]. This decomposition highlights physical scattering behaviors of objects present in the scene by defining 9 zones according to their values in the $H - \alpha$ plane as presented in Figure 5.4. This allows for any image to obtain an initial classification based on physical scattering behaviors. Then using this initialization, the k-means algorithm can be done.

Concerning problem (2), there is no reason to use this initial classification since the physical behavior of each object can change over time. It is thus difficult to choose the number of classes using such methodology. However, other approaches [Doulgeris et al., 2011, Formont et al., 2011] based on class merging schemes can be considered. Due to time constraints, we will, however, not be able to consider those algorithms.

5.4.3 Geometric mean on $\mathcal{M}_{p,N}$

In order to implement the k-means algorithm with Riemannian metrics, we need to be able to compute the geometric mean of M points $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M = (\boldsymbol{\xi}_1, \boldsymbol{\tau}_1), \dots, (\boldsymbol{\xi}_M, \boldsymbol{\tau}_M)$ lying in manifold $\mathcal{M}_{p,N}$. This geometric mean is formally defined as the solution to the following

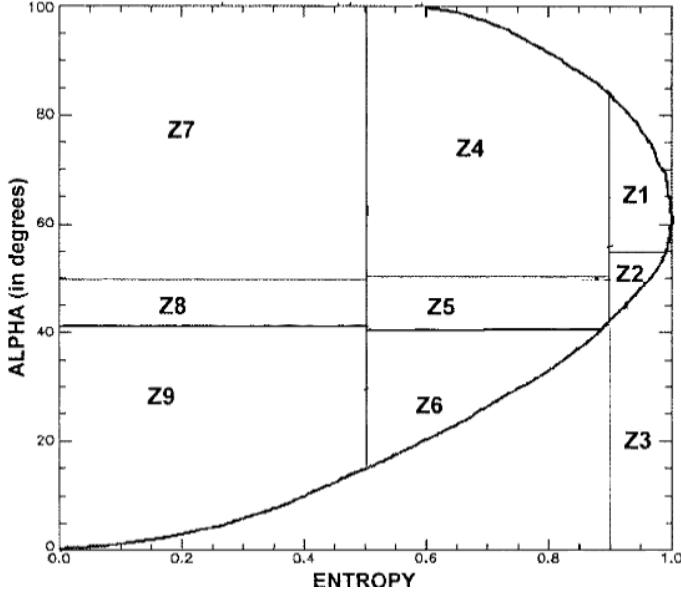


Figure 5.4: $H - \alpha$ clustering. Z1: High Entropy Multiple Scattering, Z2: High Entropy Vegetation Scattering, Z3: (Not a Feasible Region), Z4: Medium Entropy Multiple Scattering, Z5: Medium Entropy Vegetation Scattering, Z6: Medium Entropy Surface Scattering, Z7: Low Entropy Multiple Scattering, Z8: Low Entropy Dipole Scattering, Z9: Low Entropy Surface Scattering.

optimization problem:

$$\boldsymbol{\theta}_{\text{mean}} = \underset{\boldsymbol{\theta} \in \mathcal{M}_{p,N}}{\operatorname{argmin}} \sum_{m=1}^M \delta_{\mathcal{M}_{p,N}}^2(\boldsymbol{\theta}, \boldsymbol{\theta}_m) \quad (5.36)$$

Since $\delta_{\mathcal{M}_{p,N}}^2$ decomposes in two separate terms, $\delta_{\mathbb{S}_{\mathbb{H},|\bullet|}^p}^2$ for shape matrix and $\delta_{(\mathbb{R}^+)^N}^2$ for textures, the mean $\boldsymbol{\theta}_{\text{mean}}$ is equal to $(\boldsymbol{\xi}_{\text{mean}}, \boldsymbol{\tau}_{\text{mean}})$, where $\boldsymbol{\xi}_{\text{mean}}$ is geometric mean on $\mathbb{S}_{\mathbb{H},|\bullet|}^p$ and $\boldsymbol{\tau}_{\text{mean}}$ is the geometric mean on $(\mathbb{R}^+)^N$.

The geometric mean in $\mathbb{S}_{\mathbb{H},|\bullet|}^p$ is the unique element in $\mathbb{S}_{\mathbb{H},|\bullet|}^p$ such that [Moakher, 2005]:

$$\sum_{m=1}^M \log(\boldsymbol{\xi}_{\text{mean}}^{-1} \boldsymbol{\xi}_m) = 0 \quad (5.37)$$

For $M > 2$, there exists no analytical form of the solution but it can be obtained through an iterative algorithm minimizing the objective function³:

$$f(\boldsymbol{\xi}) = \sum_{m=1}^M \delta_{\mathbb{S}_{\mathbb{H},|\bullet|}^p}^2(\boldsymbol{\xi}, \boldsymbol{\xi}_m). \quad (5.38)$$

Concerning $(\mathbb{R}^+)^N$, the geometric mean is simply given by:

$$\boldsymbol{\tau}_{\text{mean}} = (\odot_{m=1}^M \boldsymbol{\tau}_m)^{1/M}. \quad (5.39)$$

³See [Fletcher and Joshi, 2004] or [Jeuris et al., 2012] for a detailed description of those algorithms.

5.4.4 Application to real SAR images

Classification under problem (1)

We consider the EMISAR image⁴, shown at Figure 5.5, which is a well-known image in the SAR clustering litterature [Feng et al., 2014, Doulgeris et al., 2019, Doulgeris et al., 2011]. Since we do not consider the temporal aspect for the clustering, we only consider a single image as a first step.



Figure 5.5: Left: Span of EMISAR SAR image (Span) over Foulom region in Denmark. Right: Optical image for comparison (@Google)

In order to apply our methodology, we compute covariance and textures parameters on a 7×7 spatial windows using the MLE of a deterministic compound-Gaussian distribution corresponding to the Tyler estimator for shape matrix and quadratic form for texture parameters. Then we cluster the data according to algorithm 6 using natural distance and geometric mean. The initialization is done using $H - \alpha$ decomposition.

In order to compare the results, we compare several other clustering methodologies:

- The classic Wishart classifier [Guo et al., 2015] which consider covariance matrices estimated rthrough SCM as features. The Wishart distance is used to compute distances between those covariances $\hat{\Sigma}$ and mean of class μ :

$$\delta_{CW} = \log |\mu| + \text{Tr}(\mu^{-1} \hat{\Sigma}). \quad (5.40)$$

The mean is computed by doing a simple arithmetic mean between elements of the same class.

- The Wishart classifier but with a geometric mean in the Riemannian sense on the manifold $\mathbb{S}_{\mathbb{H}}^p$ as described in [Ovarlez et al., 2011].
- The heterogeneous classification methodology of [Formont et al., 2011] which considers a modified distance on shape matrices estimated through Tyler's estimator $\hat{\xi}$:

$$\delta_{CW} = \log \frac{|\mu|}{|\hat{\xi}|} + \text{Tr}(\mu^{-1} \hat{\xi}). \quad (5.41)$$

The mean is computed through the geodesic mean on $\mathbb{S}_{\mathbb{H}}^p$.

- The Riemannian clustering based on shape matrix and texture that we developed here.

The $H - \alpha$ initialisation is presented at figure 5.6 and the results of the k-means algorithm are presented at Figure 5.7.

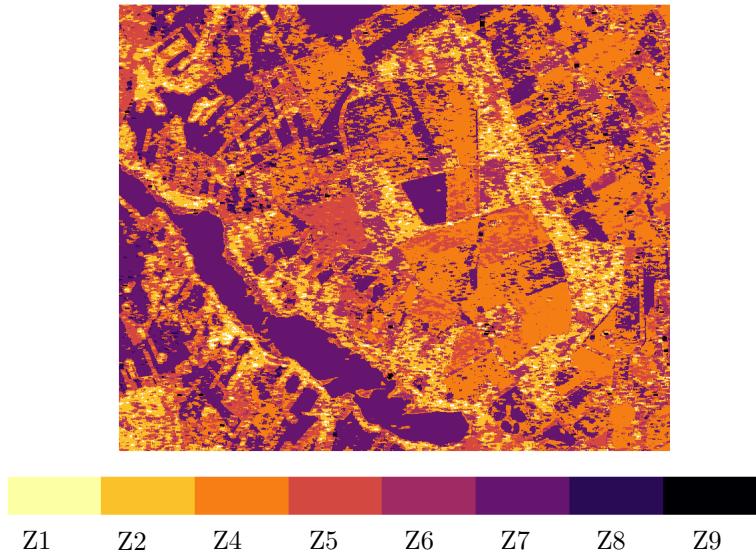


Figure 5.6: $H - \alpha$ initialisation

As always, with no ground truth available it is difficult to compare the outputs provided by the different methodologies. Moreover, the interpretation of the classification is subjective relative to the observer's zones of interest in the image.

However, as a first analysis we can have the following remarks:

- Concerning the Wishart classifier, some classes (for example Z9) have been assigned to very few pixels corresponding of outliers. Since there are fewer classes available, the overall classification of the scene is more homogeneous than the others.
- While using the Riemannian mean, in place of the arithmetic mean, improves the situation, it can be noticed that the lake is mapped to the same class as most of the fields which is surprising since they do not correspond to the same physical objects as was observed in the initialization.
- The use of the classification methodology based on Tyler's estimator allows differentiating the lake from the fields. However, the classification is much more heterogeneous: many pixels of the same field can be mapped to different classes.
- Finally, our Riemannian methodology appears to obtain a more homogeneous classification which is coherent with the $H - \alpha$ initialization. There is, however, a problem at the interface between fields: the upper border of a field is not necessarily matched with the lower border. The same observation can be done for left and right borders. This is due to the fact that the texture parameters are ordered in a vector. Thus depending on the border, the textures parameters corresponding to the field do not have necessarily the same order. As a consequence, the distance between pixels at two different borders can be high.

⁴Available at <https://earth.esa.int/web/polsarpro/data-sources/sample-datasets#EMISAR>.

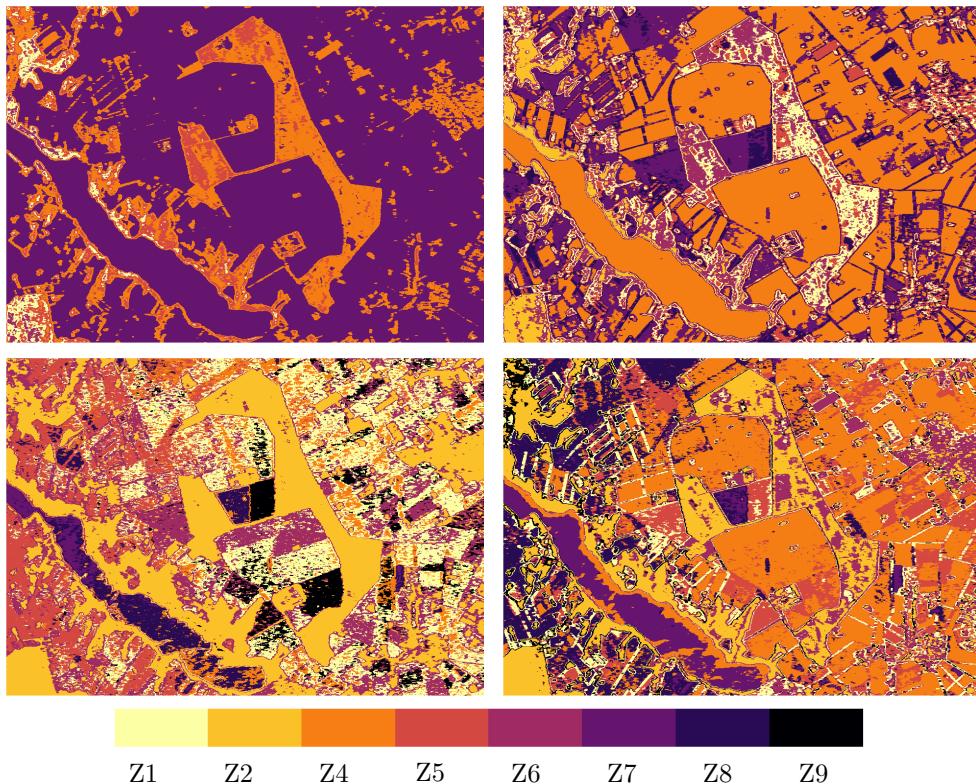


Figure 5.7: Results of clustering on EMISAR image. Top-left: Wishart classifier. Top-right: Wishart classifier with geodesic mean. Bottom-left: Robust shape clustering. Bottom-right: Shape and texture Riemannian clustering (proposed).

In conclusion, it appears that the methodology proposed has interesting results but suffers from a transition problem which could be solved by a similarity matching strategy between textures at the cost of some complexity.

Classification under problem (2)

We consider a time series of UAVSAR data referenced as Snjoaq dataset comprising of 17 co-registered time series, from which we selected a zone corresponding to crop fields. The data is presented in Figure 5.8.

Results of clustering for $K = 4$ classes is given in Figure 5.10. The initialization has been done using random samples as mean for each class. In the outputs presented, there is an impulsive behavior in the interior of the fields. This is due to the fact that the algorithm has been stopped before convergence to have a reasonable computation time with the limited resources available. The results can be made more homogeneous by having more iterations.

It is interesting to see that again for these images, the Wishart classifier with arithmetic mean does not fare well. Indeed, similarly as previously, classes C2 and C4 represent a few outlier pixels while the whole image is panned by two classes. It is difficult to interpret what those classes represent.

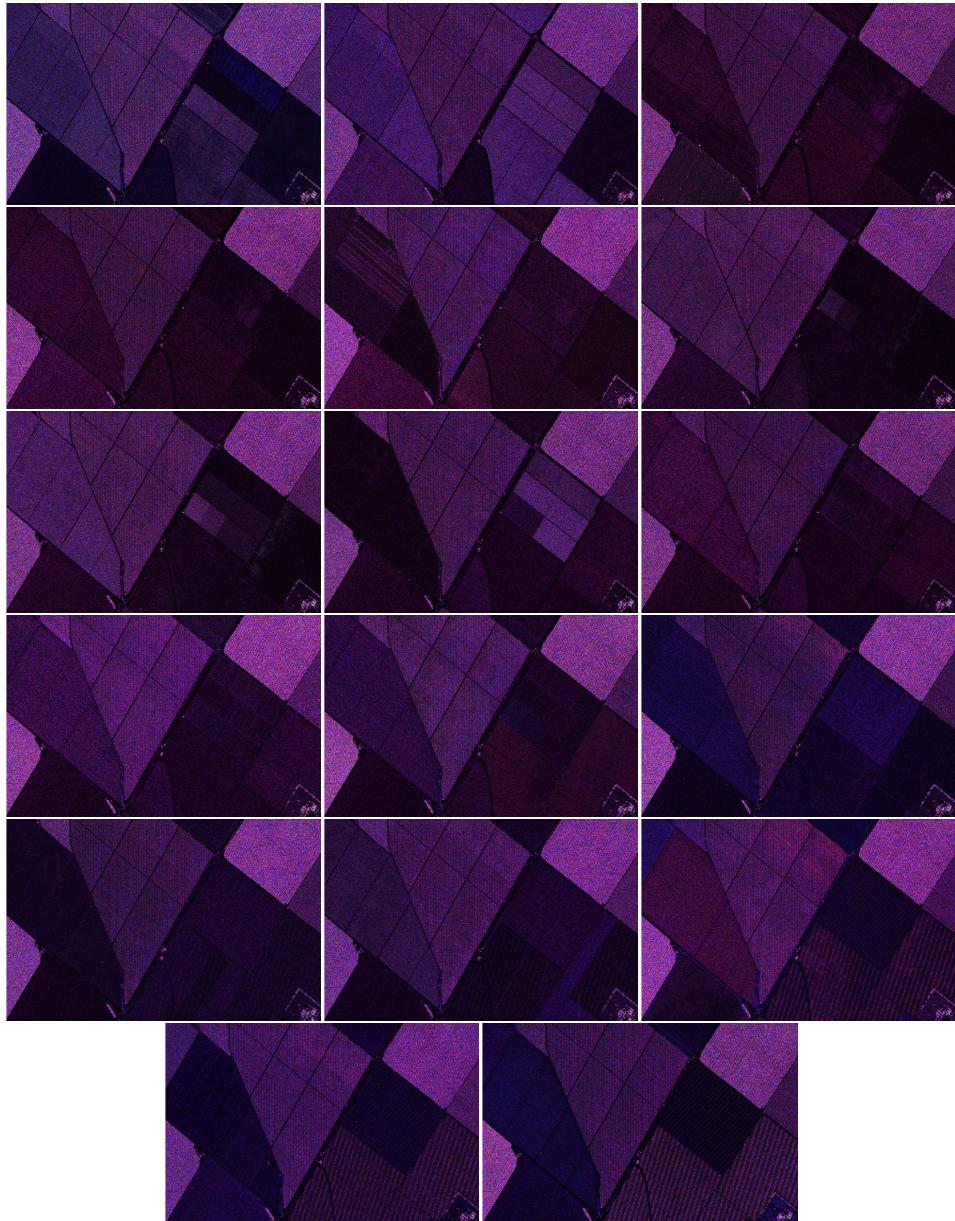


Figure 5.8: UAVSAR Snjoaq dataset in Pauli representation

The Riemannian mean allows being more robust to those outliers. Thus, the four classes can now span the whole image. It appears that in this case, class C1 corresponds to zones subjected to changes while the other 3 classes correspond to zones that do not vary significantly over time as can be seen in Figure 5.9, where the change detection statistic $\hat{\Lambda}_{MT}$ has been used to highlight changing zones.

Clustering using the robust distance associated with shape alone yields a very heterogeneous classification such as obtained for the problem (1). In this case, the output is not

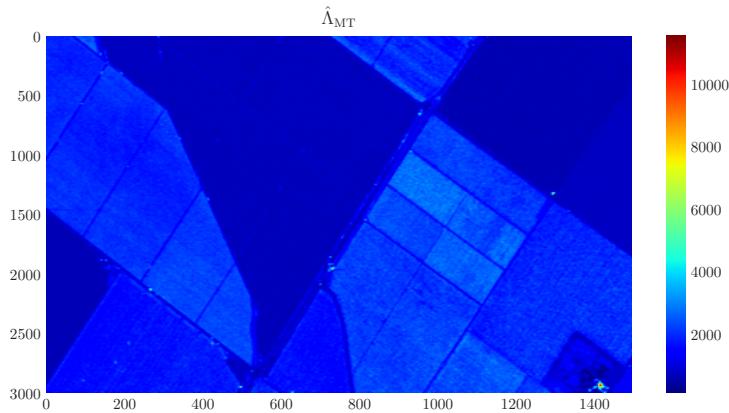


Figure 5.9: Output of $\hat{\Lambda}_{MT}$ statistic with a window size of 7×7 .

interpretable.

Finally, the proposed approach yields a single class (C4) for the nonchanging zones. The changing zones are classified into three different classes according to their pattern of evolution which was the purpose of the problem. As such the output obtained is interesting.

We must, however, note that the methodology used only works for objects that change exactly in the same patterns such as fields that are planted and harvested at the same time. Indeed, since the distance used does not consider time-shifts or warping, it does not allow to compare objects having similar patterns which are not synchronized in time.

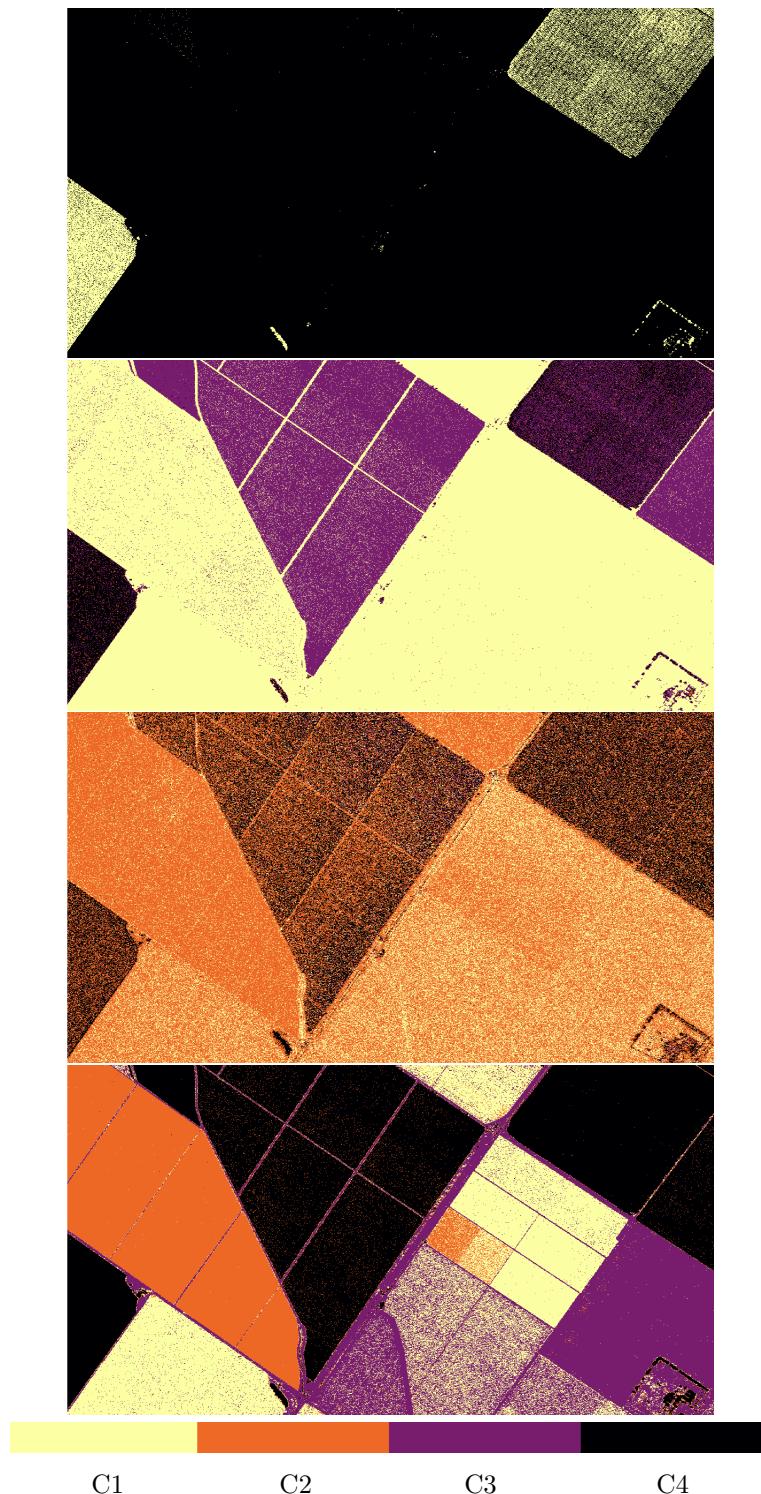


Figure 5.10: Results of Snjoaq temporal clustering with $K = 4$ classes. Top: Wishart classifier. Middle-Top: Wishart classifier with Riemannian mean. Middle-Bottom: Robust shape clustering. Bottom: Proposed approach.

5.5 CONCLUSIONS

We have developed in this chapter Riemannian geometry adapted to deterministic compound-Gaussian model as an opening to several subjects presented in this thesis.

First, the complexity of the change-point detection algorithm of chapter 4 has been considered. We presented a recursive stochastic gradient descent to obtain an online implementation of the algorithm. This work is still ongoing but as a preliminary result, we found that compared to recursive arithmetic mean, the gain in term of estimation performance is not apparent. However, the methodology seems to be more time-efficient.

It still remains to conduct an analysis of the online scheme in terms of false alarm regulation and detection performance.

We also considered a clustering problem in order to bring more information about changes than a simple detection: either a list of classes that each pixels transition to or a clustering according to the pattern of evolution. For the first case, we showed that the use of Riemannian mean over arithmetic, even in the Gaussian case, improves the robustness to outliers. The methodology proposed has interesting results but suffers from an ordering problem at the interface between objects. For the second clustering problem, the methodology proposed appears to allow to discriminate change patterns better than a simple covariance-based approach.

This work, however only considered non-elastic distances⁵, while the literature on time-series clustering mostly relies on such elastic distances. Thus an interesting extension of this work could be to adapt this formalism to the Riemannian tools we used here. Notably, as a first step, a composite distance on covariance and dynamic time warping on textures parameter over time can be of interest.

⁵Meaning that it does not account shift and distortions in time.

5.A PROOF OF PROPOSITION 5.2.1 AT P. 169

Proposition. *The Fisher information metric of the deterministic compound-Gaussian model over $\mathcal{M}_{p,N}$ is defined for $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\tau}) \in \mathcal{M}_{p,N}$, $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_\xi, \boldsymbol{\zeta}_\tau) \in T_{\boldsymbol{\theta}}\mathcal{M}_{p,N}$ and $\boldsymbol{\eta} = (\boldsymbol{\eta}_\xi, \boldsymbol{\eta}_\tau) \in T_{\boldsymbol{\theta}}\mathcal{M}_{p,N}$ by:*

$$\langle \boldsymbol{\zeta}, \boldsymbol{\eta} \rangle_{\boldsymbol{\theta}}^{\mathcal{M}_{p,N}} = \langle \boldsymbol{\zeta}_\xi, \boldsymbol{\eta}_\xi \rangle_\xi^{\mathbb{S}_{\mathbb{H},|\bullet|}^p} + \langle \boldsymbol{\zeta}_\tau, \boldsymbol{\eta}_\tau \rangle_\tau^{(\mathbb{R}^+)^N}. \quad (5.42)$$

Proof. Let us define for $k \in \llbracket 1, N \rrbracket$, the functions $\phi_k(\boldsymbol{\theta}) = \tau_k \boldsymbol{\xi}$. The directional derivative of ϕ_k in the direction $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_\xi, \boldsymbol{\zeta}_\tau)$ is given by:

$$D\phi_k(\boldsymbol{\theta})[\boldsymbol{\zeta}] = (\boldsymbol{\zeta}_\tau)_k \boldsymbol{\xi} + \tau_k \boldsymbol{\zeta}_\xi, \quad (5.43)$$

where $(\bullet)_k$ is the k -th element of vector \bullet .

By definition, the Fisher information metric is given by:

$$\langle \boldsymbol{\zeta}, \boldsymbol{\eta} \rangle_{\boldsymbol{\theta}}^{\mathcal{M}_{p,N}} = \mathbb{E}_{\mathbf{X}} \{ D \log \mathcal{L}(\mathbf{X}; \boldsymbol{\theta})[\boldsymbol{\zeta}] \times D \log \mathcal{L}(\mathbf{X}; \boldsymbol{\theta})[\boldsymbol{\eta}] \}, \quad (5.44)$$

which can be written as:

$$\begin{aligned} \langle \boldsymbol{\zeta}, \boldsymbol{\eta} \rangle_{\boldsymbol{\theta}}^{\mathcal{M}_{p,N}} &= \sum_{k=1}^N \mathbb{E}_{\mathbf{X}} \{ D(\mathcal{L}_{\mathcal{G}}(\mathbf{X}, \phi_k(\boldsymbol{\theta})))[\boldsymbol{\zeta}] \times D(\mathcal{L}_{\mathcal{G}}(\mathbf{X}, \phi_k(\boldsymbol{\theta})))[\boldsymbol{\eta}] \} \\ &= \sum_{k=1}^N \mathbb{E}_{\mathbf{X}} \{ D\mathcal{L}_{\mathcal{G}}(\mathbf{X}, \phi_k(\boldsymbol{\theta}))[D\phi_k(\boldsymbol{\theta})[\boldsymbol{\zeta}]] \times \\ &\quad D\mathcal{L}_{\mathcal{G}}(\mathbf{X}, \phi_k(\boldsymbol{\theta}))[D\phi_k(\boldsymbol{\theta})[\boldsymbol{\eta}]] \} \end{aligned}$$

We recognize here the definition of the Fisher information metric in the Gaussian case, thus:

$$\langle \boldsymbol{\zeta}, \boldsymbol{\eta} \rangle_{\boldsymbol{\theta}}^{\mathcal{M}_{p,N}} = \sum_{k=1}^N \langle D\phi_k(\boldsymbol{\theta})[\boldsymbol{\zeta}], D\phi_k(\boldsymbol{\theta})[\boldsymbol{\eta}] \rangle_{\phi_k(\boldsymbol{\theta})}^{\mathbb{S}_{\mathbb{H},|\bullet|}^p}.$$

Through cumbersome manipulations we find that the metric reads:

$$\begin{aligned} \langle \boldsymbol{\zeta}, \boldsymbol{\eta} \rangle_{\boldsymbol{\theta}}^{\mathcal{M}_{p,N}} &= \langle \boldsymbol{\zeta}_\xi, \boldsymbol{\eta}_\xi \rangle_\xi^{\mathbb{S}_{\mathbb{H},|\bullet|}^p} + \langle \boldsymbol{\zeta}_\tau, \boldsymbol{\eta}_\tau \rangle_\tau^{(\mathbb{R}^+)^N} + \text{Tr}\{\boldsymbol{\xi}^{-1} \boldsymbol{\zeta}_\xi\} (\boldsymbol{\eta}_\tau \odot \boldsymbol{\tau}^{-1})^T \mathbf{1}_{N,1} \\ &\quad + \text{Tr}\{\boldsymbol{\xi}^{-1} \boldsymbol{\eta}_\xi\} (\boldsymbol{\zeta}_\tau \odot \boldsymbol{\tau}^{-1})^T \mathbf{1}_{N,1}. \end{aligned} \quad (5.45)$$

Now since $\boldsymbol{\zeta}_\xi$ and $\boldsymbol{\eta}_\xi$ both belong to $T_{\boldsymbol{\xi}}\mathbb{S}_{\mathbb{H},|\bullet|}^p$, we have $\text{Tr}\{\boldsymbol{\xi}^{-1} \boldsymbol{\zeta}_\xi\} = 0$ and $\text{Tr}\{\boldsymbol{\xi}^{-1} \boldsymbol{\eta}_\xi\} = 0$, which concludes the proof. \square

5.B PROOF OF PROPOSITION 5.3.1 AT P. 179

Proposition. *The Riemannian gradient of $\log \mathcal{L}_{\mathcal{G}}(\mathbf{X}_t; \boldsymbol{\theta})$ at $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\tau})$ is:*

$$\text{grad}_{\mathcal{M}_{p,N}} \log \mathcal{L}_{\mathcal{G}}(\mathbf{X}_t; \boldsymbol{\theta}) = \left(\mathcal{P}_{\mathbb{S}_{\mathbb{H},|\bullet|}^p}^{\mathbb{S}_{\mathbb{H},|\bullet|}^p} \left(\sum_{k=1}^N \tau_k^{-1} \mathbf{x}_k^t \mathbf{x}_k^{t^H} \right), \mathbf{g}_\tau \right), \quad (5.46)$$

where for $1 \leq k \leq K$, the k -th element of \mathbf{g}_τ is $(\mathbf{g}_\tau)_k = \text{Tr}(\boldsymbol{\xi}^{-1} \mathbf{x}_k^t \mathbf{x}_k^{t^H}) - \tau_k p$.

Proof. As for the proof of proposition 5.2.1, we define for $k \in \llbracket 1, N \rrbracket$, the functions $\phi_k(\boldsymbol{\theta}) = \tau_k \boldsymbol{\xi}$.

Let $k \in \llbracket 1, N \rrbracket$, the directional derivative $D \log \mathcal{L}_{CG}(\mathbf{X}_t; \boldsymbol{\theta})[\boldsymbol{\eta}]$ of $\log \mathcal{L}_{CG}(\mathbf{X}_t; \boldsymbol{\theta})$ at point $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\tau})$ towards direction $\boldsymbol{\eta} = (\boldsymbol{\eta}_{\boldsymbol{\xi}}, \boldsymbol{\eta}_{\boldsymbol{\tau}}) \in T_{\boldsymbol{\theta}} \mathcal{M}_{p,N}$ is given by:

$$\begin{aligned} D \log \mathcal{L}_{CG}(\mathbf{X}_t; \boldsymbol{\theta})[\boldsymbol{\eta}] &= \sum_{k=1}^N \text{Tr} \left(\mathbf{x}_k^t \mathbf{x}_k^{t^H} (\phi_k(\boldsymbol{\theta}))^{-1} D\phi_k(\boldsymbol{\theta})[\boldsymbol{\eta}] (\phi_k(\boldsymbol{\theta}))^{-1} \right) \\ &\quad - \sum_{k=1}^N \text{Tr} ((\phi_k(\boldsymbol{\theta}))^{-1} D\phi_k(\boldsymbol{\theta})[\boldsymbol{\eta}]) \\ &= \left\langle \sum_{k=1}^N \tau_k^{-1} \mathbf{x}_k^t \mathbf{x}_k^{t^H}, \boldsymbol{\eta}_{\boldsymbol{\xi}} \right\rangle_{\boldsymbol{\xi}}^{\mathbb{S}_{\mathbb{H}, |\bullet|}^p} + \\ &\quad \sum_{k=1}^N \text{Tr} \left(\boldsymbol{\xi}^{-1} (\mathbf{x}_k^t \mathbf{x}_k^{t^H} - \tau_k \boldsymbol{\xi}) \right) \tau_k^{-1} (\boldsymbol{\eta}_{\boldsymbol{\tau}})_k \tau_k^{-1} \\ &= \left\langle \sum_{k=1}^N \tau_k^{-1} \mathbf{x}_k^t \mathbf{x}_k^{t^H}, \boldsymbol{\eta}_{\boldsymbol{\xi}} \right\rangle_{\boldsymbol{\xi}}^{\mathbb{S}_{\mathbb{H}, |\bullet|}^p} + \langle \mathbf{g}_{\boldsymbol{\tau}}, \boldsymbol{\eta}_{\boldsymbol{\tau}} \rangle_{\boldsymbol{\tau}}^{(\mathbb{R}^+)^N}. \end{aligned}$$

The result is obtained by identifying the directional derivative to the Riemannian gradient:

$$D \log \mathcal{L}_{CG}(\mathbf{X}_t; \boldsymbol{\theta})[\boldsymbol{\eta}] = \langle \text{grad}_{\mathcal{M}_{p,N}} \log \mathcal{L}_{CG}(\mathbf{X}_t; \boldsymbol{\theta}), \boldsymbol{\eta} \rangle_{\boldsymbol{\theta}}^{\mathcal{M}_{p,N}}.$$

□

CONCLUSIONS

The present thesis has considered the problem of change detection in time series of high-resolution SAR images. As presented in Chapter 1, these type of images raise issues towards their processing: on the first hand, there is a need to obtain a diversity in order to use efficient multivariate based methodologies. On the other hand, the non-Gaussian distribution of high-resolution SAR images has to be taken into account to develop efficient methodologies.

The diversity issue was tackled in Chapter 2 where a wavelet packet was proposed to decompose the SAR images in order to obtain a spectro-angular diversity. Then, we have showed that the wavelet decomposition can be used in classic statistical detection methodologies. Moreover, the packet proposed allows controlling the sidelobes level while acting similar to the standard STFT. The benefit of sidelobes reduction has been illustrated in a target detection where the classic ANMF detector has better performance of detection with lower sidelobes.

The wavelet packet we proposed, does, however, split the spectra of the SAR image uniformly in both frequencies and angles domain. This can be a problem for images, where the spectral density is not uniformly distributed: some wavelet coefficients may contain a very small portion of the signal while other coefficients could still be split to improve the representation. This problem can partly explain why low-rank methodologies presented in 3.B and 3.C have better detection performance than their full-rank counterparts.

In order to improve the situation, it may be interesting to use the multi-resolution framework to select a decomposition which yields the best possible diversity of a given image while keeping the size of vector low. To this end, a solution can be for example the use of an entropy-based criterion when doing the decomposition. Indeed, by selecting a basis thanks to criterion such as developed in [Coifman and Wickerhauser, 1992], it would be possible to obtain more coefficients in dense energy zones and fewer in sparse energy zones.

Concerning the non-Gaussian issue, we have considered, in Chapter 3, the development of new statistics for testing the homogeneity of covariances matrices. We assumed both elliptical and compound-Gaussian models in order to bring robustness to the covariance equality problem. Through theoretical and experimental analysis, we have shown the attractiveness of the approach based on the deterministic compound-Gaussian model. Indeed, the model has allowed extending the Gaussian model by adding more degrees of freedom while also being less restrictive than the elliptical families which require the knowledge of the density generator function. We have also seen that a detection of the shape matrix alone is not satisfactory for a change detection application since the relative power between the images of the time series is an important factor to account for the changes.

A possible extension of this work consists of obtaining an approximation of the distribution of the statistic $\hat{\Lambda}_{MT}$ for testing a change in both texture and shape. We have seen that the Wilks theorem does not apply since the number of parameters grows as much as the

number of samples. As a first step, we have shown the unbiasedness and consistency of the shape matrices estimates used in the statistic. This result can be used in delta method techniques to obtain an approximation. However, since the expression of the statistic involves a ratio of dependent terms, the derivation is not trivial.

Another possible extension has been partially explored in 3.B and 3.C, where a low-rank structure on the matrices has been assumed. The new statistics obtained under this assumption have shown promising results on real SAR images. As such it might be interesting to study the problem of covariance homogeneity testing to other structured models such as Toeplitz or persymmetric matrices or Kronecker models.

We have also considered the problem of joint detection and estimation of change-points in Chapter 4. Based on the work of [Conradsen et al., 2016], we have adapted the statistics of Chapter 3 in order to obtain a robust methodology which has presented interesting results with regards to Gaussian-based methodology.

This work could benefit from more comparison to other change-points estimation algorithms. Notably, adapting the concept of the Bernoulli detector of [Harlé et al., 2016] by assigning a prior on the covariance matrices of the series, is an interesting possibility.

Thanks to a collaboration with Lucien Bacharach, we have considered the change-point problem in a Bayesian context. We developed a lower-bound on the MSE for the estimation of change points in Gaussian context. This lower bound has allowed considering the interesting problem which consists of tuning the size of the spatial window used for the estimation in order to obtain a good trade-off between spatial resolution and performance of estimation. This work has been done in the Gaussian context due to the difficulty associated with computing some of the bound elements in the elliptical case. Indeed, for most of the density generators, the derivation of the Renyi distance, which is needed in the bound, is still an open problem. However, an extension of the bound to \mathbb{CAE} distribution may be tractable since the density generator is not needed.

Finally, we have seen in Chapter 5, some perspectives involving the use of Riemannian geometry. We have developed a new geometry adapted to deterministic compound-Gaussian models based in order to consider two issues:

- With the new geometry, we have explored the use of stochastic gradient descent as a mean of optimization to compute the statistics under compound-Gaussian assumption. This allowed considering a recursive implementation of the change-point detection algorithm that is usable in an online context.

The preliminary study has showed a potential interest of this approach in order to improve the complexity of change-point estimation. The study is, however, still on going. Notably, an issue with the tuning of the stepsize has been raised. Improving the situation by taking an adaptive stepsize (based for example on the Fisher Information Matrix of the model) could raise the attractiveness of the proposed approach. Moreover, the study of change-point estimation performance has not been considered yet. This point will be considered in collaboration with Jialun Zhou, Salem Said and Yannick Berthoumieu of the IMS laboratory in Bordeaux.

- In order to bring more information about the changes, we have considered the problem of clustering the image time series according to its temporal evolution. We have thus considered two problems consisting of clustering using spatial features or temporal features. In the first case, the ordering of texture parameters has raised an issue on borders which may be solved by a costly reordering. For the second case, the

proposed methodology has shown very interesting results allowing to discriminate zones according to their temporal evolution.

Several extension of this preliminary work can be considered:

- The use of an elastic distance such as dynamic time warping [Petitjean et al., 2012] can be considered on texture parameters in order to compare series with same modalities with a time-shift.
- Other manifolds with respect to low-rank or other structures on the shape matrix can be considered. By developing a new geometry adapted to these models, the clustering could be improved similarly to the change detection.
- It might be more relevant to consider the geometry associated with the parameters over the whole time series rather than considering the geometry of a single parameter. Indeed, for each pixel, the time series of parameters can be seen as a sequence defining a curve in a manifold. Then, methodologies to compare curves, such as done in [Le Brigant, 2017], can be adapted to work for clustering purposes.

RÉSUMÉ EN FRANÇAIS

En télédétection, l'imagerie par Radar à Synthèse d'ouverture (SAR) est une technique prometteuse de par sa capacité à produire des images de bonne qualité à une très haute résolution. Les systèmes SAR sont connus pour leur fonctionnement dans toute sortes de conditions d'illumination et météo. Les dernières décennies ont vu l'avènement de missions spatiales, telles que Sentinel-1 ou TerraSAR-X, qui permettent d'avoir accès à un grand volume de données acquises régulièrement. L'augmentation du nombre de missions permet d'obtenir des images multi-temporelles et multi-dimensionnelles de la surface Terrestre, avec de grandes résolutions spatiales et temporelles. Dans ce contexte, l'utilisation du traitement des séries temporelles est un axe intéressant pour diverses applications qui vont de la détection de changement, à de la surveillance de trafic maritime.

Ce nouveau scénario de données massives engendre toutefois une nouvelle complexité qui constitue une recherche active dans les domaines de la télédétection et du traitement du signal. Dans ce contexte, les objectifs de cette thèse sont la prise en compte des spécificités des nouveaux systèmes radar (Très haute résolution) afin de développer des méthodologies pour la détection de changement dans des séries d'images SAR multi-temporelles.

L'axe de travail porte d'une part, sur l'analyse temps-fréquence des images SAR haute résolution afin d'extraire une information physique de dispersivité et d'anisotropie sur les diffuseurs présent dans une scène. D'autre part, le vecteur multivarié contenant cette information peut être modélisé par une loi de probabilité. En particulier, le travail se concentre sur la comparaison des matrices de covariance des données entre plusieurs dates. Des outils statistiques sont ainsi mis en œuvre à cet effet.

La thèse s'organise en cinq chapitres considérant chacun un aspect de l'analyse des changements dans une série temporelle d'images SAR. Dans le premier chapitre, nous introduisons les principales problématiques et enjeux que nous considérerons tout au long de la thèse. Nous résumons également les résultats de base de la littérature concernant l'analyse de changements pour les images SAR. Dans le second chapitre, nous nous intéressons à l'utilisation de transformées en ondelettes pour extraire une information de dispersivité et d'anisotropie qui permet de caractériser les diffuseurs présents sur les images. Cette information peut ainsi être utilisée dans un cadre statistique comme nous le verrons au chapitre 3. Dans ce chapitre, nous considérons le problème de détection de changements par le prisme d'un test d'égalité de matrices de covariance. Dans ce cadre, nous développons de nouveaux tests statistiques basés sur un modèle de gaussienne-composée, plus adaptée aux images SAR que la gaussienne, et nous analysons leurs propriétés théoriques. Nous étendons ces tests statistiques dans un cadre qui permet la détection des points de ruptures en chapitre 4. Enfin, nous proposons une ouverture en chapitre 5 basée sur la géométrie Riemannienne afin de considérer le problème de la détection en ligne et un problème de classification non supervisée des images.

Les différentes contributions de cette thèse sont résumées à présent ici-bas.

ANALYSE DES IMAGES SAR PAR TRANSFORMÉES EN ONDELETTES

Dans un premier temps, nous nous sommes intéressés aux images SAR très haute résolution monovariés. Celle-ci sont obtenus par des algorithmes tels que Range Migration Algorithm (RMA) à travers l'intégration d'un coefficient de rétro-diffusion $\tilde{I}(\mathbf{k})$ qui correspond à la réponse de la scène pour chaque vecteur d'onde \mathbf{k} donné:

$$I(\mathbf{r}) = \int_{\mathcal{D}} \tilde{I}(\mathbf{k}) \exp(2i\pi \mathbf{k}^T \mathbf{r}) d\mathbf{k}, \quad (5.47)$$

où \mathbf{r} correspond à la position du diffuseur et \mathcal{D} corresponds au support spectral et angulaire du système SAR.

On peut reconstruire une l'image sur une partition de \mathcal{D} afin de récupérer la réponse de la scène correspondant au bandes de fréquences et intervalles d'angles correspondants en considérant l'image formée par :

$$\tilde{I}_{\mathcal{E}}(\mathbf{r}) = \int_{\mathcal{E}} \tilde{I}(\mathbf{k}) \overline{\Psi}_{\mathcal{E}}^{\mathbf{S}}(\mathbf{k}, \mathbf{r}) d\mathbf{k}, \quad (5.48)$$

où $\Psi_{\mathcal{E}}^{\mathbf{S}}(\mathbf{k}, \mathbf{r})$ est une fonction ondelette avec un support \mathcal{E} . Lorsque nous considérons plusieurs partitions $\mathcal{E}_{1, \dots, M}$ de \mathcal{D} , on peut définir un paquet en ondelettes $\{\Psi_{\mathcal{E}_i}^{\mathbf{S}}(\mathbf{k}, \mathbf{r})/i = 1, \dots, M\}$.

Ainsi, nous avons proposé un nouveau paquet d'ondelettes avec un support sur le domaine polaire des angles de visé et des fréquences du système SAR. Pour cela nous considérons dans un premier temps les ondelettes de Shannon [Hess-Nielsen, 1994, Nielsen, 2002] dans le domaine cartésien que nous adaptons à ce domaine polaire. Enfin, nous considérons le développement d'une nouveau paquet basé sur des fonctions en cloche :

$$g_{a,b,c}^{\text{Bell}}(x) = \frac{1}{1 + \left| \frac{x - c}{a} \right|^{2b}}. \quad (5.49)$$

Ce nouveau paquet permet de réduire les lobes secondaires des diffuseurs forts, afin d'améliorer la résolution spatiale des images ainsi obtenues.

DÉVELOPPEMENT DE NOUVEAUX TESTS STATISTIQUES

Nous montrons dans le premier chapitre de cette thèse que le modèle de la gaussienne, souvent utilisé pour caractériser les images SAR polarimétriques, ne fonctionne plus pour des images très haute résolution. Ainsi le modèle des distributions elliptiques a été introduit pour modéliser les données. Ces distributions se caractérisent par la distribution de probabilité suivante :

$$p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\Sigma}, g) = \mathfrak{C}_{p,g} |\boldsymbol{\Sigma}|^{-1} g(\mathbf{x}^H \boldsymbol{\Sigma}^{-1} \mathbf{x}), \quad (5.50)$$

où $\mathfrak{C}_{p,g}$ est un facteur de normalisation, $\boldsymbol{\Sigma}$ est la matrice de covariance associée aux données et g une fonction respectant certaines conditions de régularité. Nous associons ainsi à chaque date une matrice $\boldsymbol{\Sigma}_t = \tau_t \boldsymbol{\xi}_t$. $\boldsymbol{\xi}_t$ correspond à une matrice normalisée ($\text{Tr}(\boldsymbol{\xi}_t) = p$) et τ_t un facteur d'échelle. Nous considérons le problème qui consiste, pour T images différentes, à décider entre les deux alternatives suivantes:

$$\begin{cases} H_0 : \boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_T = \boldsymbol{\Sigma}_0, \\ H_1 : \exists (t, t') \in \llbracket 1, T \rrbracket^2, \boldsymbol{\Sigma}_t \neq \boldsymbol{\Sigma}_{t'} \end{cases}$$

Dans ce problème, nous voulons détecteur un changement à la fois dans la forme de la matrice de covariance et du facteur d'échelle (correspondant à la puissance locale des diffuseurs) et si possible discriminer entre les deux. Cette approche a pour but de distinguer différents types de changements. En effet, un changement dans la puissance locale ne correspond pas au même phénomène qu'un changement uniquement dans la forme de la matrice de covariance.

Après avoir obtenu deux tests statistiques en utilisant la technique de Test de Vraisemblance Généralisé (TRVG), nous considérons un nouveau modèle de gaussienne-composée qui permet de s'abstraire de la connaissance de la fonction g qui dépend de l'image considérée.

Partant de ce nouveau modèle, trois statistiques ont été développées :

- une statistique sensible à un changement dans la forme de la matrice de covariance ainsi que dans l'échelle de puissance. Elle sera nommée $\hat{\Lambda}_{MT}$.
- une statistique sensible uniquement à la forme de la matrice de covariance. Elle sera nommée $\hat{\Lambda}_{MG}$.
- une statistique sensible uniquement à une variation dans la puissance locale. Elle sera nommée $\hat{\Lambda}_{Tex}$.

Ces statistiques nécessitent la résolution d'une équation de point fixe, pour lequel un algorithme a été proposé. Les propriétés de convergence de cet algorithme ainsi que les propriétés statistiques des nouvelles statistiques introduites ont également été étudiés d'un point de vue strictement théorique mais également lors de simulations numériques. Les propriétés statistiques démontrent une robustesse de nos détecteurs dans le cadre des gaussienne-composée du fait qu'elles respectent la propriété de taux de fausse alarme constante.

Enfin, des expérimentations sur données réelles ont permis de valider une meilleure détection en pratique du détecteur $\hat{\Lambda}_{MT}$ vis-à-vis de l'état de l'art.

DÉTECTION DE RUPTURES ET BORNE HYBRIDE

Dans le quatrième chapitre de cette thèse, nous considérons un algorithme de détection et d'estimation de points de ruptures développé par [Conradsen et al., 2016] basé sur des tests d'hypothèses dans un cadre gaussien. Nous extendons cette approche dans le cadre des distributions gaussienne-composées à l'aide des résultats du chapitre 3. Des expérimentations sur données réelles montrent l'intérêt de la méthode.

Dans un second temps, nous considérons à la dérivation d'une borne hybride de Cramér-Rao/Weiss-Weinstein sur l'erreur quadratique moyenne de l'estimation d'un point de rupture dans une série temporelle de matrices distribuées selon un modèle de Wishart. La borne fait appel à un résultat obtenu par [Bacharach et al., 2019] qui permet une dérivation rapide du résultat. Cette borne permet ainsi de s'intéresser à une problématique de design sur le compromis entre résolution spatiale de la détection de la rupture et performance d'estimation temporelles de celle-ci.

GÉOMÉTRIE RIEMANNIENNE

Dans ce dernier chapitre, nous proposons une ouverture basée sur le principe de la géométrie riemannienne afin de considérer deux problèmes :

- l'algorithme de détection de ruptures du chapitre 4 repose sur une procédure itérative de tests de détection. Or dans le cadre de la gaussienne-composée, ces tests statistiques

reposent sur la résolution d'une équation de point-fixe couteuse. Des travaux basés sur cette géométrie riemannienne ont permis dans un cadre similaire d'obtenir une estimation en ligne de la matrice de covariance sans considérer d'algorithme de point-fixe. Ainsi, nous considérons la même approche.

- il serait intéressant de pouvoir classifier les séries d'images SAR selon le type de changement qu'elles subissent. Dans le cadre de la classification des images SAR, l'utilisation de la géométrie riemannienne a permis dans la littérature d'améliorer la robustesse des méthodes de classification non-supervisée, ce qui nous amène à considérer le même cadre.

Ce chapitre considère alors la dérivation d'une métrique sur la variété riemannienne des paramètres d'une loi de gaussienne-composée afin de pouvoir adapter les approches de la littérature dans ce cadre non-gaussien robuste. Les premiers résultats obtenus concernant les deux problèmes sont prometteurs.

BIBLIOGRAPHY

- [Bar, 1937] (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 160(901):268–282.
- [Absil et al., 2009] Absil, P.-A., Mahony, R., and Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.
- [Achim et al., 2003] Achim, A., Tsakalides, P., and 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 et al., 2013] Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A., and 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.
- [Aminikhanghahi and Cook, 2017] Aminikhanghahi, S. and Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2):339–367. 28603327[pmid].
- [Anderson, 2003] Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley.
- [Anfinsen et al., 2007] Anfinsen, S. N., Jenssen, R., and Eltoft, T. (2007). Spectral clustering of polarimetric SAR data with wishart-derived distance measures. In *Proc. POLIn-SAR*, volume 7, pages 1–9.
- [Atto et al., 2016] Atto, A. M., Trouvé, E., Nicolas, J., and Lê, T. T. (2016). Wavelet operators and multiplicative observation models—application to SAR image time-series analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 54(11):6606–6624.
- [Atto et al., 2013] Atto, A. M., Trouve, E., Berthoumieu, Y., and Mercier, G. (2013). Multidate divergence matrices for the analysis of SAR image time series. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4):1922–1938.
- [Averbuch et al., 2006] Averbuch, A., Coifman, R. R., Donoho, D., Elad, M., and Israeli, M. (2006). Fast and accurate polar Fourier transform. *Applied and Computational Harmonic Analysis*, 21(2):145 – 167.
- [Bacharach et al., 2019] Bacharach, L., Korso, M. N. E., Renaux, A., and Tourneret, J. (2019). A hybrid lower bound for parameter estimation of signals with multiple change-points. *IEEE Transactions on Signal Processing*, 67(5):1267–1279.
- [Bacharach et al., 2017] Bacharach, L., Renaux, A., Korso, M. N. E., and Chaumette, E. (2017). Weiss–Weinstein bound on multiple change-points estimation. *IEEE Transactions on Signal Processing*, 65(10):2686–2700.

- [Bankó and Abonyi, 2012] Bankó, Z. and Abonyi, J. (2012). Correlation based dynamic time warping of multivariate time series. *Expert Systems with Applications*, 39(17):12814 – 12823.
- [Barber, 2015] Barber, J. (2015). A generalized likelihood ratio test for coherent change detection in polarimetric sar. *IEEE Geoscience and Remote Sensing Letters*, 12(9):1873–1877.
- [Basseville and Nikiforov, 1993] Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes - Theory and Application*. Prentice Hall, Inc. - <http://people.irisa.fr/Michele.Basseville/kniga/>.
- [Benjamini and Hochberg, 1995] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.
- [Bentoutou et al., 2005] Bentoutou, Y., Taleb, N., Kpalma, K., and Ronsin, J. (2005). An automatic image registration for applications in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 43(9):2127–2137.
- [Bertrand and Bertrand, 1996] Bertrand, J. and Bertrand, P. (1996). The concept of hyper-image in wide-band radar imaging. *Geoscience and Remote Sensing, IEEE Transactions on*, 34(5):1144–1150.
- [Besson et al., 2017] Besson, O., Coluccia, A., Chaumette, E., Ricci, G., and Vincent, F. (2017). Generalized likelihood ratio test for detection of Gaussian rank-one signals in Gaussian noise with unknown statistics. *Signal Processing, IEEE Transactions on*, 65(4):1082–1092.
- [Bhatia, 2009] Bhatia, R. (2009). *Positive definite matrices*, volume 16. Princeton university press.
- [Boyd and Vandenberghe, 2004] Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.
- [Brekke et al., 2013] Brekke, C., Anfinsen, S. N., and Larsen, Y. (2013). Subband extraction strategies in ship detection with the subaperture cross-correlation magnitude. *IEEE Geoscience and Remote Sensing Letters*, 10(4):786–790.
- [Breloy, 2015] Breloy, A. (2015). *Estimation and Detection Algorithms for Low Rank Heterogeneous Context*. Theses, Université Paris-Saclay.
- [Brigui et al., 2014] Brigui, F., Thirion-Lefevre, L., Ginolhac, G., and Forster, P. (2014). New SAR algorithm based on orthogonal projections for MMT detection and interference reduction. *Geoscience and Remote Sensing, IEEE Transactions on*, 52(7):3800–3811.
- [Brodsky and Darkhovsky, 2013] Brodsky, E. and Darkhovsky, B. S. (2013). *Non-parametric statistical diagnosis: problems and methods*, volume 509. Springer Science & Business Media.
- [Bruzzone and Prieto, 2000] 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 et al., 1995] Bucciarelli, T., Lombardo, P., Oliver, C. J., and Perrotta, M. (1995). 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*, volume 1, pages 181–183 vol.1.
- [Candès et al., 2006] Candès, E., Demanet, L., Donoho, D., and Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3):861–899.
- [Carincotte et al., 2006] Carincotte, C., Derrode, S., and Bourennane, S. (2006). Unsupervised change detection on SAR images using fuzzy hidden markov chains. *IEEE Transactions on Geoscience and Remote Sensing*, 44(2):432–441.
- [Carotenuto et al., 2015] Carotenuto, V., Maio, A. D., Clemente, C., and Soraghan, J. (2015). Unstructured versus structured GLRT for multipolarization SAR change detection. *IEEE Geoscience and Remote Sensing Letters*, 12(8):1665–1669.
- [Carotenuto et al., 2016] Carotenuto, V., Maio, A. D., Clemente, C., Soraghan, J. J., and Alfano, G. (2016). Forcing scale invariance in multipolarization SAR change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 54(1):36–50.
- [Carrara et al., 1995] Carrara, W. G., Goodman, R. S., and Majewski, R. M. (1995). *Spot-light Synthetic Aperture Radar: Signal Processing Algorithms*. Artech House Publishers.
- [Chen et al., 2012] Chen, G., Hay, G., Carvalho, L., and Wulder, M. (2012). Object-based change detection. *International Journal of Remote Sensing*, 33:4434–4457.
- [Chen et al., 2011] Chen, Y., Wiesel, A., and Hero, A. O. (2011). Robust shrinkage estimation of high-dimensional covariance matrices. *Signal Processing, IEEE Transactions on*, 59(9):4097–4107.
- [Christensen, 2002] Christensen, O. (2002). *An Introduction to Frames and Riesz Bases*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston.
- [Ciuonzo et al., 2017] Ciuonzo, D., Carotenuto, V., and Maio, A. D. (2017). On multiple covariance equality testing with application to SAR change detection. *IEEE Transactions on Signal Processing*, 65(19):5078–5091.
- [Cloude and Pottier, 1997] Cloude, S. R. and Pottier, E. (1997). An entropy based classification scheme for land applications of polarimetric sar. *IEEE Transactions on Geoscience and Remote Sensing*, 35(1):68–78.
- [Cohen et al., 1992] Cohen, A., Daubechies, I., and Feauveau, J. (1992). Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 45(5):485–560.
- [Coifman and Wickerhauser, 1992] Coifman, R. R. and Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on information theory*, 38(2):713–718.
- [Coluccia and Ricci, 2017] Coluccia, A. and Ricci, G. (2017). Adaptive radar detectors for point-like Gaussian targets in Gaussian noise. *Aerospace and Electronic Systems, IEEE Transactions on*, 53(3):1284–1294.

- [Conradsen et al., 2001] Conradsen, K., Nielsen, A. A., Schou, J., and Skriver, H. (2001). 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)*, volume 6, pages 2628–2630 vol.6.
- [Conradsen et al., 2003] Conradsen, K., Nielsen, A. A., Schou, J., and Skriver, H. (2003). A test statistic in the complex wishart distribution and its application to change detection in polarimetric SAR data. *Geoscience and Remote Sensing, IEEE Transactions on*, 41(1):4–19.
- [Conradsen et al., 2003] Conradsen, K., Nielsen, A. A., Schou, J., and 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.
- [Conradsen et al., 2016] Conradsen, K., Nielsen, A. A., and Skriver, H. (2016). Determining the points of change in time series of polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5):3007–3024.
- [De Grandi et al., 2007] De Grandi, G. D., Lee, J. S., and Schuler, D. L. (2007). Target detection and texture segmentation in polarimetric SAR images using a wavelet frame: Theoretical aspects. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(11):3437–3453.
- [De Grandi et al., 2009] De Grandi, G. D., Lucas, R. M., and Kropacek, J. (2009). Analysis by wavelet frames of spatial statistics in SAR data for characterizing structural properties of forests. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(2):494–507.
- [De Maio et al., 2010] De Maio, A., Kay, S. M., and Farina, A. (2010). On the invariance, coincidence, and statistical equivalence of the glrt, rao test, and wald test. *IEEE Transactions on Signal Processing*, 58(4):1967–1979.
- [Deng et al., 2017] Deng, X., López-Mart
- [Doulgeris et al., 2019] Doulgeris, A., Anfinsen, S., and Eltoft, T. (2019). Automated k-wishart clustering of PolSAR images.
- [Doulgeris et al., 2008] Doulgeris, A. P., Anfinsen, S. N., and Eltoft, T. (2008). Classification with a non-gaussian model for PolSAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 46(10):2999–3009.
- [Doulgeris et al., 2011] Doulgeris, A. P., Anfinsen, S. N., and Eltoft, T. (2011). Automated non-gaussian clustering of polarimetric synthetic aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3665–3676.
- [Drašković and Pascal, 2016] Drašković, G. and Pascal, F. (2016). New properties for Tyler's covariance matrix estimator. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 820–824.
- [Dunn, 1961] Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64.

- [Duquenoy et al., 2010] Duquenoy, M., Ovarlez, J.-P., Ferro-Famil, L., Pottier, E., and Vignaud, L. (2010). Scatterers characterisation in radar imaging using joint time-frequency analysis and polarimetric coherent decompositions. *IET Radar, Sonar and Navigation*, 4(3):384–402.
- [Durand et al., 2009] Durand, R., Ginolhac, G., Thirion, L., and Forster, P. (2009). New SAR processor based on matched subspace detectors. *Aerospace and Electronic Systems, IEEE Transactions on*, 45(1):221–236.
- [Efron, 2007] Efron, B. (2007). Size, power and false discovery rates. *Ann. Statist.*, 35(4):1351–1377.
- [Federer, 1951] Federer, W. T. (1951). Testing proportionality of covariance matrices. *Ann. Math. Statist.*, 22(1):102–106.
- [Feng et al., 2014] Feng, J., Cao, Z., and Pi, Y. (2014). Polarimetric contextual classification of PolSAR images using sparse representation and superpixels. *Remote Sensing*, 6:7158–7181.
- [Ferrari and Tourneret, 2003] Ferrari, A. and Tourneret, J.-Y. (2003). Barankin lower bound for change-points in independent sequences. In *IEEE Workshop on Statistical Signal Processing, 2003*, pages 557–560.
- [Fletcher and Joshi, 2004] Fletcher, P. T. and Joshi, S. (2004). Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, pages 87–98. Springer.
- [Formont et al., 2011] Formont, P., Ovarlez, J. P., Pascal, F., Vasile, G., and Ferro-Famil, L. (2011). On the extension of the product model in PolSAR processing for unsupervised classification using information geometry of covariance matrices. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 1361–1364.
- [Formont et al., 2011] Formont, P., Pascal, F., Vasile, G., Ovarlez, J.-P., and Ferro-Famil, L. (2011). Statistical classification for heterogeneous polarimetric SAR images. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):567–576.
- [Formont et al., 2013] Formont, P., Veganzones, M. A., Frontera-Pons, J., Pascal, F., Ovarlez, J.-P., and Chanussot, J. (2013). CFAR Hierarchical Clustering of Polarimetric SAR Data. In *2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2013)*, 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2013), pages 2461–2464, Melbourne, Australia.
- [Fotopoulos and Jandhyala, 2001] Fotopoulos, S. and Jandhyala, V. (2001). Maximum likelihood estimation of a change-point for exponentially distributed random variables. *Statistics & Probability Letters*, 51(4):423 – 429.
- [Fotopoulos et al., 2010] Fotopoulos, S. B., Jandhyala, V. K., and Khapalova, E. (2010). Exact asymptotic distribution of change-point MLE for change in the mean of gaussian sequences. *Ann. Appl. Stat.*, 4(2):1081–1104.
- [Foucher and Lopez-Martinez, 2014] Foucher, S. and Lopez-Martinez, C. (2014). Analysis, evaluation, and comparison of polarimetric SAR speckle filtering techniques. *IEEE Transactions on Image Processing*, 23(4):1751–1764.

- [Frambourg et al., 2013] Frambourg, C., Douzal-Chouakria, A., Gaussier, É., and Demangeot, J. (2013). Variance/Covariance extension for time series discrimination. Technical report.
- [Freitas et al., 2005] Freitas, C. C., Frery, A. C., and Correia, A. H. (2005). The polarimetric g distribution for SAR data analysis. *Environmetrics*, 16(1):13–31.
- [Frery et al., 2014] Frery, A. C., Nascimento, A. D. C., and Cintra, R. J. (2014). Analytic Expressions for Stochastic Distances Between Relaxed Complex Wishart Distributions. *IEEE Transactions on Geoscience and Remote Sensing*, 52:1213–1226.
- [F.R.S., 1901] F.R.S., K. P. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.
- [Gao, 2010] Gao, G. (2010). Statistical modeling of SAR images: A survey. *Sensors*, 10(1):775–795.
- [Garzelli and Zoppetti, 2017] Garzelli, A. and Zoppetti, C. (2017). Optimizing SAR change detection based on log-ratio features. In *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, pages 1–4.
- [Gens and Genderen, 1996] Gens, R. and Genderen, J. L. V. (1996). Review article SAR interferometry—issues, techniques, applications. *International Journal of Remote Sensing*, 17(10):1803–1835.
- [Gini and Greco, 2002] Gini, F. and Greco, M. (2002). Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter. *Signal Process.*, 82(12):1847–1859.
- [Gong et al., 2012] Gong, M., Cao, Y., and Wu, Q. (2012). A neighborhood-based ratio approach for change detection in SAR images. *IEEE Geoscience and Remote Sensing Letters*, 9(2):307–311.
- [Gong et al., 2016] Gong, M., Zhao, J., Liu, J., Miao, Q., and 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.
- [Goodman, 1976] Goodman, J. W. (1976). Some fundamental properties of speckle*. *J. Opt. Soc. Am.*, 66(11):1145–1150.
- [Goodnight and Schwartz, 1997] Goodnight, C. J. and Schwartz, J. M. (1997). A bootstrap comparison of genetic covariance matrices. *Biometrics*, 53(3):1026–1039.
- [Greco and De Maio, 2016] Greco, M. S. and De Maio, A., editors (2016). *Modern Radar Detection Theory*. SciTech Publishing.
- [Greco and Gini, 2007] 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.
- [Guo et al., 2015] Guo, S., Tian, Y., Li, Y., Chen, S., and Hong, W. (2015). Unsupervised classification based on h/alpha decomposition and wishart classifier for compact polarimetric sar. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1614–1617.

- [Gustafsson, 2000] Gustafsson, F. (2000). *Adaptive filtering and change detection*, volume 1. Wiley.
- [Hallac et al., 2017] Hallac, D., Vare, S., Boyd, S. P., and Leskovec, J. (2017). Toeplitz inverse covariance-based clustering of multivariate time series data. *CoRR*, abs/1706.03161.
- [Hallin and Paindaveine, 2007] Hallin, M. and Paindaveine, D. (2007). Parametric and semi-parametric inference for shape: the role of the scale functional. *Statistics & Decisions*, 24(3):327–350.
- [Hallin and Paindaveine, 2008] Hallin, M. and Paindaveine, D. (2008). Optimal rank-based tests for homogeneity of scatter. *Ann. Statist.*, 36(3):1261–1298.
- [Hallin and Paindaveine, 2009] Hallin, M. and Paindaveine, D. (2009). Optimal tests for homogeneity of covariance, scale, and shape. *Journal of Multivariate Analysis*, 100(3):422 – 444.
- [Harlé et al., 2016] Harlé, F., Chatelain, F., Gouy-Pailler, C., and Achard, S. (2016). Bayesian model for multiple change-points detection in multivariate time series. *IEEE Transactions on Signal Processing*, 64(16):4351–4362.
- [Harle, 2016] Harle, F. (2016). *Multiple change-point detection in multivariate time series : application to the inference of dependency networks*. Theses, Université Grenoble Alpes.
- [Hecheltjen et al., 2014] Hecheltjen, A., Thonfeld, F., and Menz, G. (2014). *Recent Advances in Remote Sensing Change Detection – A Review*, pages 145–178. Springer Netherlands, Dordrecht.
- [Hess-Nielsen, 1994] Hess-Nielsen, N. (1994). Control of frequency spreading of wavelet packets. *Applied and Computational Harmonic Analysis*, 1(2):157–168.
- [Hussain et al., 2013] Hussain, M., Chen, D., Cheng, A., Wei, H., and Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80:91 – 106.
- [Ingla and Mercier, 2007] Ingla, 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.
- [Irving et al., 1997] Irving, W. W., Novak, L. M., and Willsky, A. S. (1997). A multiresolution approach to discrimination in SAR imagery. *Aerospace and Electronic Systems, IEEE Transactions on*, 33(4):1157–1169.
- [Jambawalikar and Kumar, 2008] Jambawalikar, S. and Kumar, P. (2008). A note on approximate minimum volume enclosing ellipsoid of ellipsoids. In *2008 International Conference on Computational Sciences and Its Applications*, pages 478–487.
- [Jeuris et al., 2012] Jeuris, B., Vandebril, R., and Vandereycken, B. (2012). A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electronic Transactions on Numerical Analysis*, 39(ARTICLE):379–402.
- [Jia et al., 2014] Jia, L., Li, M., Wu, Y., Zhang, P., Chen, H., and An, L. (2014). Semisupervised SAR image change detection using a cluster-neighborhood kernel. *IEEE Geoscience and Remote Sensing Letters*, 11(8):1443–1447.

- [Kay, 1998] Kay, S. M. (1998). Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory. *Signal Process.*
- [Kay, 2010] Kay, S. M. (2010). Fundamentals of statistical signal processing: estimation theory.
- [Kay and Gabriel, 2003] Kay, S. M. and Gabriel, J. R. (2003). An invariance property of the generalized likelihood ratio test. *IEEE Signal Processing Letters*, 10(12):352–355.
- [Kervrann and Boulanger, 2006] Kervrann, C. and Boulanger, J. (2006). Optimal spatial adaptation for patch-based image denoising. *IEEE Transactions on Image Processing*, 15(10):2866–2878.
- [Kovacevic et al., 2002] Kovacevic, J., Dragotti, P. L., and Goyal, V. K. (2002). Filter bank frame expansions with erasures. *Information Theory, IEEE Transactions on*, 48(6):1439–1450.
- [Kraut et al., 2001] Kraut, S., Scharf, L., and Whorter, L. M. (2001). Adaptive Subspace Detector. *Signal Processing, IEEE Transactions on*, 49(1):1–16.
- [Le Brigant, 2017] Le Brigant, A. (2017). Computing distances and geodesics between manifold-valued curves in the SRV framework. *Journal of Geometric Mechanics*, 9(2):131–156.
- [LeCam, 1960] LeCam, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.*, 3:37–98.
- [Lehmann, 2009] Lehmann, E. (2009). *Testing Statistical Hypotheses*, volume 102.
- [Lesniewska-Choquet et al., 2017] Lesniewska-Choquet, C., Atto, A. M., Mauris, G., and Mercier, G. (2017). Image change detection by possibility distribution dissemblance. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6.
- [Li and Zelnio, 1996] Li, J. and Zelnio, E. G. (1996). Target detection with synthetic aperture radar. *Aerospace and Electronic Systems, IEEE Transactions on*, 32(2):613–627.
- [Liu et al., 2014a] Liu, B., Xu, L., Zheng, S., and Tian, G.-L. (2014a). A new test for the proportionality of two large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 131:293 – 308.
- [Liu et al., 2011] Liu, M., Zhang, H., and Wang, C. (2011). Change detection in urban areas of high-resolution polarization SAR images using heterogeneous clutter models. In *2011 3rd International Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*, pages 1–4.
- [Liu et al., 2014b] Liu, M., Zhang, H., Wang, C., and Wu, F. (2014b). Change detection of multilook polarimetric SAR images using heterogeneous clutter models. *IEEE Transactions on Geoscience and Remote Sensing*, 52(12):7483–7494.
- [Lopez-Martinez and Fabregas, 2002] Lopez-Martinez, C. and Fabregas, X. (2002). Modeling and reduction of SAR interferometric phase noise in the wavelet domain. *Geoscience and Remote Sensing, IEEE Transactions on*, 40(12):2553–2566.
- [Love, 1985] Love, A. W. (1985). In Memory of Carl A. Wiley. *Antennas and Propagation Society Newsletter, IEEE*, (June):17–18.

- [Ma et al., 2012] Ma, J., Gong, M., and Zhou, Z. (2012). Wavelet fusion on ratio images for change detection in SAR images. *IEEE Geoscience and Remote Sensing Letters*, 9(6):1122–1126.
- [MacQueen et al., 1967] MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA.
- [Mahot et al., 2013] Mahot, M., Pascal, F., Forster, P., and Ovarlez, J. (2013). Asymptotic properties of robust complex covariance matrix estimates. *IEEE Transactions on Signal Processing*, 61(13):3348–3356.
- [Maronna, 1976a] Maronna, R. A. (1976a). Robust M -estimators of multivariate location and scatter. *Annals of Statistics*, 4(1):51–67.
- [Maronna, 1976b] Maronna, R. A. (1976b). Robust m -estimators of multivariate location and scatter. *Ann. Statist.*, 4(1):51–67.
- [Maronna et al., 2006] Maronna, R. A., Martin, D. R., and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley Series in Probability and Statistics. Wiley, 1 edition.
- [Martin and Pierre,] Martin, B. and Pierre, D. Robust estimation of the sur model. *Canadian Journal of Statistics*, 28(2):277–288.
- [Mian et al., 2017] Mian, A., Ovarlez, J.-P., Ginolhac, G., and Atto, A. M. (2017). Multivariate change detection on high resolution monovariate SAR image using linear Time-Frequency analysis. In *25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece.
- [Moakher, 2005] Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747.
- [Muirhead and Wateraux, 1980] Muirhead, R. J. and Wateraux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika*, 67(1):31–43.
- [Muller, 1994] Muller, H. . (1994). K statistics of terrain clutter in high resolution SAR images. In *Proceedings of IGARSS '94 - 1994 IEEE International Geoscience and Remote Sensing Symposium*, volume 4, pages 2146–2148 vol.4.
- [Nagao, 1973] Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.*, 1(4):700–709.
- [Nascimento et al., 2010] Nascimento, A. D. C., Cintra, R. J., and Frery, A. C. (2010). Hypothesis testing in speckled data with stochastic distances. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1):373–385.
- [Nascimento et al., 2019] Nascimento, A. D. C., Frery, A. C., and 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.

- [Nielsen et al., 2016] Nielsen, A. A., Conradsen, K., and Skriver, H. (2016). Omnibus test for change detection in a time sequence of polarimetric SAR data. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3398–3401.
- [Nielsen, 2002] Nielsen, M. (2002). Highly nonstationary wavelet packets. *Applied and Computational Harmonic Analysis*, 12(2):209–229.
- [Novak, 2005] Novak, L. M. (2005). Coherent change detection for multi-polarization SAR. In *Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers, 2005.*, pages 568–573.
- [Oehlert, 1992] Oehlert, G. W. (1992). A note on the delta method. *The American Statistician*, 46(1):27–29.
- [Ollila and Tyler, 2014] Ollila, E. and Tyler, D. E. (2014). Regularized M-estimators of scatter matrix. *IEEE Transactions on Signal Processing*, 62(22):6059–6070.
- [Ollila et al., 2012a] Ollila, E., Tyler, D. E., Koivunen, V., and Poor, H. V. (2012a). Complex elliptically symmetric distributions: Survey, new results and applications. *IEEE Transactions on Signal Processing*, 60(11):5597–5625.
- [Ollila et al., 2012b] Ollila, E., Tyler, D. E., Koivunen, V., and Poor, H. V. (2012b). Compound-Gaussian clutter modeling with an inverse Gaussian texture distribution. *IEEE Signal Processing Letters*, 19(12):876–879.
- [Ovarlez et al., 2011] Ovarlez, J. ., Formont, P., Pascal, F., Vasile, G., and Ferro-Famil, L. (2011). Contribution of information geometry for polarimetric SAR classification in heterogeneous areas. In *2011 12th International Radar Symposium (IRS)*, pages 669–674.
- [Ovarlez et al., 2017] Ovarlez, J.-P., Ginolhac, G., and Atto, A. M. (2017). Multivariate linear time-frequency modeling and adaptive robust target detection in highly textured monovariate SAR image. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4029–4033.
- [Ovarlez et al., 2003] Ovarlez, J.-P., Vignaud, L., Castelli, J. C., Tria, M., and Benidir, M. (2003). Analysis of SAR images by multidimensional wavelet transform. *IEE Proceedings - Radar, Sonar and Navigation*, 150(4):234–241.
- [Page, 1954] Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika*, 41(1-2):100–115.
- [Pallotta et al., 2019] Pallotta, L., De Maio, A., and Orlando, D. (2019). A robust framework for covariance classification in heterogeneous polarimetric SAR images and its application to l-band data. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1):104–119.
- [Pascal et al., 2008a] Pascal, F., Chitour, Y., Ovarlez, J.-P., Forster, P., and Larzabal, P. (2008a). Covariance structure maximum-likelihood estimates in compound gaussian noise: Existence and algorithm analysis. *IEEE Transactions on Signal Processing*, 56(1):34–48.
- [Pascal et al., 2008b] Pascal, F., Chitour, Y., Ovarlez, J.-P., Forster, P., and Larzabal, P. (2008b). Covariance structure maximum likelihood estimates in compound Gaussian noise: Existence and algorithm analysis. *Signal Processing, IEEE Transactions on*, 56(1):34–38.

- [Pascal et al., 2014] Pascal, F., Chitour, Y., and Quek, Y. (2014). Generalized robust shrinkage estimator and its application to STAP detection problem. *Signal Processing, IEEE Transactions on*, 62(21):5640–5661.
- [Petitjean et al., 2010] Petitjean, F., Gançarski, P., Masseglia, F., and Forestier, G. (2010). Analysing satellite image time series by means of pattern mining. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 45–52. Springer.
- [Petitjean et al., 2012] Petitjean, F., Inglada, J., and Gancarski, P. (2012). Satellite image time series analysis under time warping. *IEEE Transactions on Geoscience and Remote Sensing*, 50(8):3081–3095.
- [Preiss and Stacy, 2006] Preiss, M. and Stacy, N. J. (2006). Coherent change detection: Theoretical description and experimental results. Technical report, DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION EDINBURGH (AUSTRALIA).
- [Prendes et al., 2015] Prendes, J., Chabert, M., Pascal, F., Giros, A., and Tourneret, J. (2015). 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)*, pages 1513–1517.
- [Radhakrishna Rao, 1948] Radhakrishna Rao, C. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 44(1):50–57.
- [Randles, 2000] Randles, R. H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *Journal of the American Statistical Association*, 95(452):1263–1268.
- [Ratha et al., 2017] Ratha, D., De, S., Celik, T., and 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.
- [Reigber et al., 2010] Reigber, A., Jäger, M., Neumann, M., and Ferro-Famil, L. (2010). Classifying polarimetric SAR data by combining expectation methods with spatial context. *International Journal of Remote Sensing*, 31(3):727–744.
- [Ren et al., 2015] Ren, C., Galy, J., Chaumette, E., Larzabal, P., and Renaux, A. (2015). Hybrid Barankin–Weiss–Weinstein bounds. *IEEE Signal Processing Letters*, 22(11):2064–2068.
- [Rignot et al., 1992] Rignot, E., Chellappa, R., and Dubois, P. (1992). Unsupervised segmentation of polarimetric SAR data using the covariance matrix. *IEEE Transactions on Geoscience and Remote Sensing*, 30(4):697–705.
- [Robey et al., 1992] Robey, F. C., Fuhrmann, D. R., Kelly, E. J., and Nitzberg, R. (1992). A CFAR adaptive matched filter detector. *Aerospace and Electronic Systems, IEEE Transactions on*, 28(1):208–216.
- [Rosa et al., 2010] Rosa, P. S. L., Renaux, A., Muravchik, C. H., and Nehorai, A. (2010). Barankin-type lower bound on multiple change-point estimation. *IEEE Transactions on Signal Processing*, 58(11):5534–5549.

- [Roy et al., 2018] Roy, S. K., Mhammedi, Z., and Harandi, M. (2018). Geometry aware constrained optimization techniques for deep learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4460–4469.
- [Roy, 1946] Roy, S. N. (1946). Multivariate analysis of variance: The sampling distribution of the numerically largest of the p-statistics on the non-null hypothesis. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 8(1):15–52.
- [Sansosti et al., 2006] Sansosti, E., Berardino, P., Manunta, M., Serafino, F., and Fornaro, G. (2006). Geometrical SAR image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 44(10):2861–2870.
- [Scarborough et al., 2010] Scarborough, S. M., Gorham, L., Minardi, M. J., Majumder, U. K., Judge, M. G., Moore, L., Novak, L., Jaroszewksi, S., Spoldi, L., and Pieramico, A. (2010). A challenge problem for SAR change detection and data compression. In *Proceedings Volume 7699, Algorithms for Synthetic Aperture Radar Imagery XVII*, volume 7699, pages 7699 – 7699 – 5.
- [Scharf, 1991] Scharf, L. L. (1991). *Statistical signal processing*, volume 98. Addison-Wesley Reading, MA.
- [Scheiber and Moreira, 2000] Scheiber, R. and Moreira, A. (2000). Coregistration of interferometric SAR images using spectral diversity. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5):2179–2191.
- [Schott, 2001] Schott, J. R. (2001). Some tests for the equality of covariance matrices. *Journal of Statistical Planning and Inference*, 94(1):25 – 36.
- [Simoncelli et al., 1992] Simoncelli, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J. (1992). Shiftable multiscale transforms. *Information Theory, IEEE Transactions on*, 38(2):587–607.
- [Skovgaard, 1984] Skovgaard, L. T. (1984). A riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, pages 211–223.
- [Smith, 2005] Smith, S. T. (2005). Covariance, subspace, and intrinsic crame/spl acute/r-rao bounds. *IEEE Transactions on Signal Processing*, 53(5):1610–1630.
- [Soloveychik and Wiesel, 2015] Soloveychik, I. and Wiesel, A. (2015). Performance analysis of Tyler’s covariance estimator. *IEEE Transactions on Signal Processing*, 63(2):418–426.
- [Subotic et al., 1994] Subotic, N. S., Collins, L. M., Gorman, J. D., and Thelen, B. J. (1994). A multiresolution approach to target detection in synthetic aperture radar data. In *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 122–126.
- [Tao et al., 2016] Tao, D., Anfinsen, S. N., and Brekke, C. (2016). Robust CFAR detector based on truncated statistics in multiple-target situations. *Geoscience and Remote Sensing, IEEE Transactions on*, 54(1):117–134.
- [Tatsuoka and Tyler, 2000] Tatsuoka, K. S. and Tyler, D. E. (2000). On the uniqueness of s-functionals and m-functionals under nonelliptical distributions. *Ann. Statist.*, 28(4):1219–1243.

- [Taylor et al., 2017] Taylor, A., Oriot, H., Forster, P., Brigui, F., Savy, L., and Daout, F. (2017). Reducing false alarm rate by testing proportionality of covariance matrices. In *International Conference on Radar Systems (Radar 2017)*, pages 1–4.
- [Terrell, 2002] Terrell, G. R. (2002). The gradient statistic. *Computing Science and Statistics*, 34(34):206–215.
- [Tourneret et al., 2003] Tourneret, J.-Y., Doisy, M., and Lavielle, M. (2003). Bayesian off-line detection of multiple change-points corrupted by multiplicative noise : application to SAR image edge detection. *Signal Processing*, 8(9):1871–1887. This publication is available at <http://www.sciencedirect.com/science/journal/01651684>.
- [Tria, 2005] Tria, M. (2005). *SAR Imaging By Multidimensional Continuous Wavelets Analysis*. Theses, Université Paris Sud - Paris XI.
- [Tria et al., 2007] Tria, M., Ovarlez, J.-P., Vignaud, L., Castelli, J. C., and Benidir, M. (2007). Discriminating real objects in radar imaging by exploiting the squared modulus of the continuous wavelet transform. *IET Radar, Sonar and Navigation*, 1(1):27–37.
- [Tyler, 1987] Tyler, D. E. (1987). A distribution-free m -estimator of multivariate scatter. *Ann. Statist.*, 15(1):234–251.
- [Unser et al., 2011] Unser, M., Chenouard, N., and Van De Ville, D. (2011). Steerable pyramids and tight wavelet frames in. *Image Processing, IEEE Transactions on*, 20(10):2705–2721.
- [Vaart, 1998] Vaart, A. W. v. d. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
- [Van Trees and Bell, 2007] Van Trees, H. L. and Bell, K. L. (2007). *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. IEEE.
- [Wald, 1943] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3):426–482.
- [Wang et al., 2013] Wang, F., Wu, Y., Zhang, Q., Zhang, P., Li, M., and Lu, Y. (2013). Unsupervised change detection on SAR images using triplet markov field model. *IEEE Geoscience and Remote Sensing Letters*, 10(4):697–701.
- [Weinstein and Weiss, 1988] Weinstein, E. and Weiss, A. J. (1988). A general class of lower bounds in parameter estimation. *IEEE Transactions on Information Theory*, 34(2):338–342.
- [Wen, 2001] Wen, X. (2001). *Performance bounds on matched-field methods for source localization and estimation of ocean environmental parameters*.
- [Wiesel, 2012] Wiesel, A. (2012). Geodesic convexity and covariance estimation. *IEEE Transactions on Signal Processing*, 60(12):6182–6189.
- [Wilks, 1932] Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, 24(3/4):471–494.
- [Wilks, 1938] Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, 9(1):60–62.

- [Willsky and Jones, 1976] Willsky, A. and Jones, H. (1976). A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control*, 21(1):108–112.
- [Wirtinger, 1927] Wirtinger, W. (1927). Zur formalen theorie der funktionen von mehr komplexen veränderlichen. *Mathematische Annalen*, 97:357–376.
- [Yanagihara et al., 2005] Yanagihara, H., Tonda, T., and Matsumoto, C. (2005). The effects of nonnormality on asymptotic distributions of some likelihood ratio criteria for testing covariance structures under normal assumption. *Journal of Multivariate Analysis*, 96(2):237 – 264.
- [Yao, 1973] Yao, K. (1973). A Representation Theorem and its Applications to Spherically Invariant Random Processes. *Information Theory, IEEE Transactions on*, 19:600–608.
- [Yeredor, 2000] Yeredor, A. (2000). The joint MAP-ML criterion and its relation to ML and to extended least-squares. *IEEE Transactions on Signal Processing*, 48(12):3484–3492.
- [Yohai, 1974] Yohai, V. J. (1974). Robust estimation in the linear model. *Ann. Statist.*, 2(3):562–567.
- [Yousif and Ban, 2013] 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 et al., 1989] Yueh, S., Kong, J., Jao, J., Shin, R., and Novak, L. (1989). K-distribution and polarimetric terrain radar clutter. *Journal of Electromagnetic Waves and Applications*, 3(8):747–768.
- [Zecchetto and Biasio, 2008] Zecchetto, S. and Biasio, F. D. (2008). A wavelet-based technique for sea wind extraction from SAR images. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(10):2983–2989.
- [Zhang and Boos, 1992] Zhang, J. and Boos, D. D. (1992). Bootstrap critical values for testing homogeneity of covariance matrices. *Journal of the American Statistical Association*, 87(418):425–429.
- [Zhang et al., 2015] Zhang, L., Sun, L., Zou, B., and Moon, W. M. (2015). Fully polarimetric SAR image classification via sparse representation and polarimetric features. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8):3923–3932.
- [Zhou and Said, 2018] Zhou, J. and Said, S. (2018). Fast, asymptotically efficient, recursive estimation in a Riemannian manifold. *arXiv e-prints*, page arXiv:1805.06811.
- [Zhu et al., 2002] Zhu, L.-X., Ng, K. W., and Jing, P. (2002). Resampling methods for homogeneity tests of covariance matrices. *Statistica Sinica*, 12(3):769–783.
- [Zoubir et al., 2018] Zoubir, A. M., Koivunen, V., Ollila, E., and Muma, M. (2018). *Robust Statistics for Signal Processing*. Cambridge University Press.

Titre : Contributions à l'analyse de séries temporelles d'images SAR

Mots-clés : Imagerie SAR; Séries temporelles; Détection robuste; Détection de changements; Ondelettes; Télédétection

Résumé : La télédétection par Radar à Synthèse d'Ouverture (RSO) offre une opportunité unique d'enregistrer, d'analyser et de prédire l'évolution de la surface de la Terre. La dernière décennie a permis l'avènement de nombreuses missions spatiales équipées de capteurs RSO (Sentinel-1, UAVSAR, TerraSAR X, etc.), ce qui a engendré une rapide amélioration des capacités d'acquisition d'images de la surface de la Terre. Le nombre croissant d'observations permet maintenant de construire des bases de données caractérisant l'évolution temporelle d'images, augmentant considérablement l'intérêt de l'analyse de séries temporelles pour caractériser des changements qui ont lieu à une échelle globale. Cependant, le développement de nouveaux algorithmes pour traiter ces données très volumineuses est un défi qui reste à relever. Dans ce contexte, l'objectif de cette thèse consiste ainsi à proposer et à développer des méthodologies relatives à la détection de changements dans les séries d'images RSO à très haute résolution spatiale.

Le traitement de ces séries pose deux problèmes notables. En premier lieu, les méthodes d'analyse statistique performantes se basent souvent sur des données multivariées caractérisant, dans le cas des images RSO, une diversité polarimétrique, interférométrique, par exemple. Lorsque cette diversité n'est pas disponible et que les images RSO sont monocanal, de nouvelles méthodologies basées sur la décomposition en ondelettes ont été développées. Celles-ci permettent d'ajouter une diversité supplémentaire spectrale et angulaire représentant le comportement physique de rétrodiffusion des diffuseurs présents la scène de l'image. Dans un second temps, l'amélioration de la résolution spatiale sur les dernières générations de capteurs engendre une augmentation de l'hétérogénéité des données obtenues. Dans ce cas, l'hypothèse gaussienne, traditionnellement considérée pour développer les méthodologies standards de détection de changements, n'est plus valide. En conséquence, des méthodologies d'estimation robuste basée sur la famille des distributions elliptiques, mieux adaptée aux données, ont été développées.

L'association de ces deux aspects a montré des résultats prometteurs pour la détection de changements.

Title: Contributions to SAR image time series analysis

Keywords: SAR imaging; Time series; Change detection; Robust detection; Wavelet; Remote Sensing

Abstract: Remote sensing data from Synthetic Aperture Radar (SAR) sensors offers a unique opportunity to record, to analyze, and to predict the evolution of the Earth. In the last decade, numerous satellite remote sensing missions have been launched (Sentinel-1, UAVSAR, TerraSAR X, etc.). This resulted in a dramatic improvement in the Earth image acquisition capability and accessibility. The growing number of observation systems allows now to build high temporal/spatial-resolution Earth surface images data-sets. This new scenario significantly raises the interest in time-series processing to monitor changes occurring over large areas. However, developing new algorithms to process such a huge volume of data represents a current challenge. In this context, the present thesis aims at developing methodologies for change detection in high-resolution SAR image time series.

These series raise two notable challenges that have to be overcome: On the one hand, standard statistical methods rely on multivariate data to infer a result which is often superior to a univariate approach. Such multivariate data is however not always available when it concerns SAR images. To tackle this issue, new methodologies based on wavelet decomposition theory have been developed to fetch information based on the physical behavior of the scatterers present in the scene. On the other hand, the improvement in resolution obtained from the latest generation of sensors comes with an increased heterogeneity of the data obtained. For this setup, the standard Gaussian assumption used to develop classic change detection methodologies is no longer valid. As a consequence, new robust methodologies have been developed considering the family of elliptical distributions which have been shown to better fit the observed data.

The association of both aspects has shown promising results in change detection applications.

