sid.inpe.br/mtc-m21c/2020/12.28.13.44-TDI

BISECTING STOCHASTIC CLUSTERING: A NEW ALGORITHM FOR POLSAR IMAGE UNSUPERVISED CLASSIFICATION

Naiallen Carolyne Rodrigues Lima Carvalho

Doctorate Thesis of the Graduate Course in Applied Computing, guided by Drs. Sidnei João Siqueira Sant'Anna e Leonardo Sant'Anna Bins, approved in November 10, 2020.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34R/43RU675>

INPE
São José dos Campos
2020

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Gabinete do Diretor (GBDIR)
Serviço de Informação e Documentação (SESID)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

BOARD OF PUBLISHING AND PRESERVATION OF INPE INTELLECTUAL PRODUCTION - CEPPII (PORTARIA Nº 176/2018/SEI-INPE):

Chairperson:

Dra. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos Climáticos (CGCPT)

Members:

Dra. Carina Barros Mello - Coordenação de Laboratórios Associados (COCTE)

Dr. Alisson Dal Lago - Coordenação-Geral de Ciências Espaciais e Atmosféricas (CGCEA)

Dr. Evandro Albiach Branco - Centro de Ciência do Sistema Terrestre (COCST)

Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia e Tecnologia Espacial (CGETE)

Dr. Hermann Johann Heinrich Kux - Coordenação-Geral de Observação da Terra (CGOBT)

Dra. Ieda Del Arco Sanches - Conselho de Pós-Graduação - (CPG)

Silvia Castro Marcelino - Serviço de Informação e Documentação (SESID)

DIGITAL LIBRARY:

Dr. Gerald Jean Francis Banon

Clayton Martins Pereira - Serviço de Informação e Documentação (SESID)

DOCUMENT REVIEW:

Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação (SESID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SESID)

ELECTRONIC EDITING:

Ivone Martins - Serviço de Informação e Documentação (SESID)

Cauê Silva Fróes - Serviço de Informação e Documentação (SESID)

sid.inpe.br/mtc-m21c/2020/12.28.13.44-TDI

BISECTING STOCHASTIC CLUSTERING: A NEW ALGORITHM FOR POLSAR IMAGE UNSUPERVISED CLASSIFICATION

Naiallen Carolyne Rodrigues Lima Carvalho

Doctorate Thesis of the Graduate Course in Applied Computing, guided by Drs. Sidnei João Siqueira Sant'Anna e Leonardo Sant'Anna Bins, approved in November 10, 2020.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34R/43RU675>

INPE
São José dos Campos
2020

Cataloging in Publication Data

Carvalho, Naiallen Carolyne Rodrigues Lima.

C253b Bisecting stochastic clustering: a new algorithm for PolSAR image unsupervised classification / Naiallen Carolyne Rodrigues Lima Carvalho. – São José dos Campos : INPE, 2020.

xxiv + 181 p. ; (sid.inpe.br/mtc-m21c/2020/12.28.13.44-TDI)

Thesis (Doctorate in Applied Computing) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2020.

Guiding : Drs. Sidnei João Siqueira Sant'Anna e Leonardo Sant'Anna Bins.

1. Stochastic distance. 2. Riemann geometry. 3. Divisive hierarchical clustering. 4. PolSAR image. 5. Unsupervised classification. I.Title.

CDU 004.021:528

Esta obra foi licenciada sob uma Licença Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada.

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

DEFESA FINAL DE TESE DE NAIALLEN CAROLYNE RODRIGUES LIMA CARVALHO

No dia 10 de novembro de 2020, às 14h, por videoconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Doutor em Computação Aplicada. O trabalho precisa da incorporação das correções sugeridas pela Banca Examinadora e revisão final pelo(s) orientador(es).

Título: "Bisecting Stochastic Clustering: a new algorithm for PolSAR image unsupervised classification"

Eu, Thales Sehn Korting, como Presidente da Banca Examinadora, assino esta ATA em nome de todos os membros.

Membros da Banca

Dr. Thales Sehn Korting, Presidente INPE.
Dr. Sidnei João Siqueira Sant'Anna, Orientador(a) INPE
Dr. Leonardo Sant'Anna Bins, Orientador(a) INPE
Dr. Elcio Hideiti Shiguemori, Membro da Banca, IEAv
Dr. Solon Venâncio de Carvalho, Membro da Banca, INPE
Dr. Antonio Henrique Correia, Convidado(a) DSG
Dr. Manoel de Araújo Sousa Júnior, Convidado(a) UFSM

sei! assinatura eletrônica

Documento assinado eletronicamente por **Thales Sehn Korting**, **Pesquisador**, em 16/11/2020, às 09:08 (horário oficial de Brasília), com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.mctic.gov.br/verifica.html, informando o código verificador **6066847** e o código CRC **69D5F4EE**.

Referência: Processo nº 01340.007797/2020-30 SEI nº 6066847

"If I have seen further it is by standing on the shoulders of Giants.".

Isaac Newton
1675

To my husband, Diego Carvalho, to my mother, Lucineide Rodrigues and to my grandmother, Marlene Rodrigues

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr. Sidnei João Siqueira Sant'Anna e Dr. Leonardo Sant'Anna Bins for all their help, advice and partnership in this work.

I would also like to thank the researchers Dr. Antonio Correia, Dr. Manoel de Araújo Sousa Júnior, Dr. Elcio Shiguemori, Dr. Solon Venâncio de Carvalho, and Dr. Thales Sehn Körting for accepting to participate in this Thesis defense and for all valuable contribution to this work.

Special thanks to INPE and all researchers, colleagues and employees, for all the learning, experience exchanges and technical support on my study, that allowed my professional and personal growth.

ABSTRACT

PolSAR (Polarimetric Synthetic Aperture Radar) images can be represented by a set of complex Hermitian positive definite matrices, which have a natural Riemannian metric tensor. PolSAR images are, also, known for following the Wishart distribution, and, by using the information theory contrast function, stochastic distances between Wishart distributions can be derived. This work addresses unsupervised classification strategies, explores the Riemann geometry and studies stochastic distances applied to PolSAR images. The proposed algorithm, named Bisecting Stochastic Clustering (BSC), is a combination between the Stochastic Clustering (SC) algorithm and the hierarchical divisive clustering algorithm. The SC algorithm is technique based on K-means, which uses stochastic distances as similarity metric. The SC algorithm can, usually, be trapped in a local minimum, what led to incorrect clustering results. Therefore, the choice of good initial parameter candidates is essential for the clustering quality. The BSC algorithm is a top-down procedure, it starts with all samples in an unique cluster, that are successively splitted into two new sub-clusters. This algorithm is mainly divided into three steps: the initial parameter determination, the cluster bi-partitioning procedure, and the choice of a suitable cluster to split. In this work, two algorithms for the initial parameter determination are tested: the Expectation-Maximization (EM) algorithm for Wishart Mixture Model and the Riemann Principal Direction Divisive Partitioning (RPDDP). The RPDDP is a new proposed algorithm, whose goal is to perform the bi-partition of a dataset. This algorithm estimates the dataset covariance matrix under the the Riemann geometry, in order to find the principal component, which is used to separate the input data in two sub-clusters. From the RPDDP two estimated sub-clusters, the BSC derives the initial parameters. The BSC second step is performed by the SC algorithm. The BSC builds a dendrogram in order to represent the dataset splitting. Each sub-cluster, or "node", links two successor sub-clusters in the dendrogram. When three or more nodes are available in one dendrogram level, the algorithm needs to choose a node to split. The BSC third step uses the information gain as the node choice rule. This work analyses the SC algorithm and two main variants of BSC. The first variant uses the RPDDP as initial parameter determiner, and the second, uses the EM algorithm as initial parameter determiner. The Bhattacharyya (B), Kullback-Leibler (KL) and Hellinger (H) stochastic distances are analysed in this work. In total, nine algorithms are evaluated: SC-B, SC-KL, SC-H, BSC-R-B, BSC-R-KL, BSC-R-H, BSC-EM-B, BSC-EM-KL, BSC-EM-H. The algorithms were analysed in a quantitative and qualitative way. The quantitative analysis consists in the confusion matrix and accuracy estimation, and the qualitative analysis explore the BSC dendrogram and the clusters scattering mechanism by inspecting the Plan $H - \alpha$.

Keywords: Stochastic distance. Riemann geometry. Divisive hierarchical clustering. PolSAR image. Unsupervised classification.

BI-DIVISÃO ESTOCÁSTICA DE AGRUPAMENTO: UM NOVO ALGORITMO PARA CLASSIFICAÇÃO NÃO SUPERVISIONADA DE IMAGENS POLSAR

RESUMO

As imagens PolSAR (*Polarimetric Synthetic Aperture Radar*) podem ser representadas por um conjunto de matrizes definidas positivas Hermitianas complexas, que possuem um tensor métrico Riemanniano. As imagens PolSAR também são conhecidas por seguir a distribuição de Wishart e, usando a função de contraste da teoria da informação, distâncias estocásticas entre as distribuições de Wishart podem ser derivadas. Este trabalho aborda estratégias de classificação não supervisionadas, explora a geometria de Riemann e estuda distâncias estocásticas aplicadas às imagens PolSAR. O algoritmo proposto, denominado *Bisecting Stochastic Clustering* (BSC), é uma combinação entre o algoritmo *Stochastic Clustering* (SC) e o algoritmo hierárquico divisivo. O algoritmo SC é uma técnica baseada no K-médias, que usa distâncias estocásticas como métrica de similaridade. O algoritmo SC pode, geralmente, ficar preso em um mínimo local, o que leva a agrupamentos incorretos. Por isso, a escolha de bons parâmetros iniciais é essencial para a qualidade do agrupamento. O algoritmo BSC é um procedimento *top-down*, ele começa com todas as amostras em um único cluster, que é sucessivamente dividido em dois novos subclusters. Este algoritmo é dividido em três etapas: a determinação do parâmetro inicial, o procedimento de bi-particionamento do cluster e a escolha de um cluster adequado para dividir. Neste trabalho, dois algoritmos para a determinação dos parâmetros iniciais são testados: o algoritmo *Expectation-Maximization* (EM) para o Modelo de Mistura de Wishart e o Particionamento Divisivo da Direção Principal de Riemann (RPDDP). O RPDDP é um novo algoritmo, proposto com objetivo de realizar a bi-partição de um conjunto de dados. Este algoritmo estima a matriz de covariância do conjunto de dados sob a geometria de Riemann, a fim de encontrar a componente principal, que é usada para separar os dados de entrada em dois subclusters. A partir dos dois subclusters estimados pelo RPDDP, o BSC deriva os parâmetros iniciais. A segunda etapa do BSC é realizada pelo algoritmo SC. O BSC constrói um dendrograma para representar a divisão do conjunto de dados. Cada sub-cluster, ou "nó", é ligado a dois sub-grupos sucessores no dendrograma. Quando há três ou mais nós disponíveis em um nível de dendrograma, o algoritmo precisa escolher um nó para ser dividido. A terceira etapa do BSC usa o ganho de informação como regra de escolha desse nó. Este trabalho analisa o algoritmo SC e as duas variantes principais do BSC. A primeira variante usa o RPDDP como determinador dos parâmetros iniciais e a segunda, usa o algoritmo EM. As distâncias estocásticas de Bhattacharyya (B), Kullback-Leibler (KL) e Hellinger (H) são analisadas neste trabalho. No total, nove algoritmos são avaliados: SC-B, SC-KL, SC-H, BSC-R-B, BSC-R-KL, BSC-R-H, BSC-EM-B, BSC-EM-KL, BSC-EM-H. Os algoritmos foram analisados de forma quantitativa e qualitativa. A análise quantitativa consiste no calculo da matriz de confusão e na estimativa da acuracia; a análise qualitativa explora o dendrograma e os mecanismos de espalhamento dos clusters através da inspeção do Plan $H-alpha$.

Palavras-chave: Distancia estocástica. Geometria de Riemann. Agrupamento divisivo hierarquico. Imagens PolSAR. Classificação não supervisionada.

LIST OF FIGURES

Page

2.1 Radar time and distance measurement. 9
2.2 Imaging radar geometry. 10
2.3 Imaging radar footprint representation. 11
2.4 Speckle formation. 13
2.5 Reflectiveness of different scattering mechanism. 14
2.6 Coherent Decomposition representation. 19
2.7 Freeman-Durden canonical mechanism and their contribution to the final covariance matrix. 21
2.8 Plan $H - \alpha$. 22
2.9 Coherent response of a resolution cell. 24
2.10 Representation of manifold $\mathcal{M}$ and the corresponding local tangent space $\mathcal{T}_{\mathbf{Z}}\mathcal{M}$ at $\mathbf{Z}$. 32

3.1 Clustering steps. 34
3.2 Clustering Taxonomy. 38
3.3 Hierarchical clustering dendrogram. 40
3.4 Tree generated by Bisecting K-Means. 41
3.5 Confusion Matrix. 44

4.1 BSC classifier modules. 49
4.2 BSC flowchart. 50
4.3 RPDDP flowchart. 52
4.4 Information Gain computation. 56
4.5 Dendrogram Structure and ID image. 57

5.1 SIR-C image of Bebedouro, with the components hh (R), hv (G) and vv (B), in C and L band. 61
5.2 Hellinger distances between all classes within SIR-C image in L band. . 63
5.3 Hellinger distances between all classes within SIR-C image in C band. . 66
5.4 Tapajós National Forest ALOS/PALSAR image with the components hh (R), hv (G) and vv (B). 68
5.5 Hellinger distances between all classes within Tapajós ALOS/PALSAR image. 70
5.6 San Francisco ALOS/PALSAR image, located in the state of California, with the components hh (R), hv (G) and vv (B). 71

5.7 Hellinger distances between all classes within San Francisco ALOS/PALSAR image. 72
5.8 Simulated PolSAR image with the components hh (R), hv (G) and vv (B). 75
5.9 Simulated PolSAR image square root Gamma distribution. 77
5.10 Initial parameter determination analysis flowchart. 79
5.11 Σ_{AvCM} versus Σ_{InCM} comparisson flowchart 80
5.12 Monte Carlo Simulation flowchart. 82
5.13 Confusion matrix. 84
5.14 Confusion matrix representation. 84
5.15 Simulated PolSAR image dendrogram, generated by the BSC-R-H algorithm. 86
5.16 Plan $H - \alpha$ of image first splitting. 88

6.1 Accuracy of the simulated PolSAR image classification using the SC algorithm with different initial parameters. 90
6.2 Classification result S01. 93
6.3 Classification result S02. 94
6.4 Classification result S03. 95
6.5 Classification result S04. 96
6.6 Classification result S05. 97
6.7 Classification result S06. 98
6.8 RPDDP and PDDP algorithm applied on Bebedouro SIR-C C band image. 100
6.9 RPDDP and PDDP algorithm applied on Bebedouro SIR-C L band image. 102
6.10 RPDDP and PDDP algorithm applied on Tapajós ALOS/PALSAR image. 104
6.11 RPDDP and PDDP algorithm applied on ALOS/PALSAR San Francisco image. 106
6.12 Monte Carlo simulation accuracy box plot. 108
6.13 Simulated PolSAR image classification using SC algorithm and randomly chosen initial parameter. 111
6.14 Simulated PolSAR image classification using EM and BSC algorithm. . 112
6.15 Simulated PolSAR image classification using RPDDP and BSC algorithm. 113
6.16 Bebedouro SIR-C in C band Plan $H - \alpha$ of all identified classes. 115
6.17 BSC-R-H Bebedouro SIR-C in C band classification result. 116
6.18 Bebedouro SIR-C in C band BSC-R-H dendrogram. 117

6.19 Bebedouro SIR-C image in C band BSC-R-H dendrogram. Node ID 1 and its children. 119
6.20 Hellinger distances between the classes clustered by BSC-R-H using Bebedouro SIR-C image in C band as input image. 120
6.21 Bebedouro SIR-C in L band Plan $H - \alpha$ of all identified classes. 122
6.22 BSC-R-H Bebedouro SIR-C in L band classification result. 123
6.23 Bebedouro SIR-C image in L band BSC-R-H dendrogram. 124
6.24 Bebedouro SIR-C image in L band BSC-R-H dendrogram. Node ID 1 and its children. 125
6.25 Hellinger distances between the classes clustered by BSC-R-H using Bebedouro PolSAR image in L band as input image. 127
6.26 Multifrequency covariance matrix. 128
6.27 BSC-R-H Bebedouro SIR-C multifrequency classification result. 128
6.28 Bebedouro SIR-C multifrequency BSC-R-H dendrogram. 129
6.29 Hellinger distances between the classes clustered by BSC-R-H using Bebedouro PolSAR multifrequency. 130
6.30 Tapajós ALOS/PALSAR Plan $H - \alpha$ of all identified classes. 131
6.31 BSC-R-H Tapajós ALOS/PALSAR classification result. 132
6.32 Tapajós ALOS/PALSAR image BSC-R-H dendrogram. 133
6.33 Tapajós ALOS/PALSAR image BSC-R-H dendrogram. Node ID 1 and its children. 135
6.34 Hellinger distances between the classes clustered by BSC-R-H using Tapajós ALOS/PALSAR image as input image. 136
6.35 San Francisco ALOS/PALSAR Plan $H - \alpha$ of all identified classes. . . . 137
6.36 BSC-R-H San Francisco ALOS/PALSAR classification result. 138
6.37 San Francisco PolSAR image BSC-R-H dendrogram. 140
6.38 San Francisco PolSAR image BSC-R-H dendrogram - Level 1. 141
6.39 San Francisco ALOS/PALSAR water splitting. 142
6.40 Hellinger distances between the classes clustered by BSC-R-H using Tapajós PolSAR image as input image. 143
6.41 Bebedouro SIR-C in C band image classification using the SC-H algorithm. 144
6.42 Bebedouro SIR-C in C band image classification using the BSC-EM-H algorithm. 144
6.43 Bebedouro SIR-C in C band image classification using the BSC-R-H algorithm. 145
6.44 Bebedouro SIR-C in L band image classification using the SC-H algorithm. 146

6.45 Bebedouro SIR-C in L band image classification using the BSC-EM-H algorithm. 146
6.46 Bebedouro SIR-C in L band image classification using the BSC-R-H algorithm. 147
6.47 Tapajós image classification using the SC-H algorithm. 148
6.49 Tapajós image classification using the BSC-R-H algorithm. 149
6.48 Tapajós image classification using the BSC-EM-H algorithm. 149
6.50 San Francisco image classification using the SC-H algorithm. 150
6.52 San Francisco image classification using the BSC-R-H algorithm. 151
6.51 San Francisco image classification using the BSC-EM-H algorithm. . . . 151

A.1 San Francisco Landsat 7. 174
A.2 BSC-PDDP-B San Francisco Landsat 7 classification result. 175
A.3 BSC-PDDP-B San Francisco Landsat 7 confusion matrix. 176
A.4 BSC-PDDP-B San Francisco Landsat 7 dendrogram. 177

B.1 BSC algorithm webpage. 179
B.2 BSC main.m file configuration. 180
B.3 BSC image name. 181

LIST OF TABLES

Page

2.1 Summary of statistical models applied on single polarization SAR image. 25
2.2 $h - \phi$ divergences and related functions. 28

5.1 SIR-C images parameters. 60
5.2 Bebedouro SIRC-C L band determinant and trace of the estimated covariance matrices. 62
5.3 Determinant and trace of the estimated covariance matrices from classes of C band image. 64
5.4 ALOS/PALSAR images parameters. 67
5.5 Determinant and trace of the average covariance matrices from Tapajós ALOS/PALSAR image classes. 69
5.6 Determinant and trace of the average covariance matrices from San Francisco ALOS/PALSAR image classes. 72
5.7 Algorithms used for the PolSAR images classification analysis. 81

6.1 Initial parameter scenarios accuracy [%]. 90
6.2 Bebedouro SIR-C C band image PDDP and RPDDP analysis. 101
6.3 Bebedouro SIR-C L band image PDDP and RPDDP analysis. 103
6.4 Tapajós ALOS/PALSAR image PDDP and RPDDP analysis. 105
6.5 San Francisco ALOS/PALSAR image PDDP and RPDDP analysis. . . . 107
6.6 Algorithms execution time (s), considering 5 iterations for SC. 108
6.7 Algorithms execution time (s), considering 15 levels with 5 iterations each for BSC algorithm. 108
6.8 Algorithms execution time (s), considering 15 levels with 5 iterations each for BSC algorithm. 109

CONTENTS

Page

1 INTRODUCTION 1
1.1 Motivation 3
1.2 Objectives 5
1.2.1 Specific objectives 6
1.3 Thesis hypothesis 6
1.4 Contributions 6
1.4.1 Research items 7
1.5 Thesis organization 8

2 SAR IMAGE 9
2.1 Speckle generation 12
2.2 PolSAR images 14
2.3 PolSAR target decomposition 19
2.3.1 Coherent decomposition 19
2.3.2 Incoherent decomposition 20
2.3.2.1 Freeman-Durden decomposition 20
2.3.2.2 Cloude-Pottier decomposition 21
2.4 SAR image statistical modeling 23
2.4.1 Single polarization SAR images 23
2.4.2 Fully PolSAR SAR images 26
2.5 Stochastic distances 27
2.6 Covariance matrices intrinsic mean 30

3 CLUSTERING 33
3.1 User's dilemma 34
3.2 Similarity metrics 36
3.3 Clustering taxonomy 38
3.3.1 Partitioning based clustering algorithms 38
3.3.2 Hierarchical clustering algorithms 39
3.3.3 Density-Based clustering algorithms 41
3.3.4 Model-Based clustering algorithms 41
3.4 Cluster validation 42
3.4.1 Internal cluster validation 42

3.4.2 External cluster validation 44
3.4.3 Relative cluster validation 45

4 NEW PROPOSED CLASSIFIER 47
4.1 Contextualization 47
4.2 Bisecting stochastic clustering algorithm 49
4.2.1 Initial parameters determination 50
4.2.1.1 Riemannian Principal Direction Divisive Partitioning (RPDDP) . . . 51
4.2.1.2 Expectation-Maximization (EM) 52
4.2.2 Splitting procedure: Stochastic Clustering 54
4.2.3 Determination of new nodes and clusters 54
4.2.4 BSC Dendrogram 56

5 METHODOLOGY AND DATA DESCRIPTION 59
5.1 Data set description 59
5.1.1 Real PolSAR images 59
5.1.1.1 SIR-C images 60
5.1.1.2 ALOS/PALSAR images 66
5.1.2 PolSAR simulated image 73
5.2 Experiments description 77
5.2.1 Experiment I 78
5.2.2 Experiment II 79
5.2.3 Experiment III 81
5.2.4 Experiments IV 83
5.3 Quantitative analysis 83
5.4 Qualitative analysis 85
5.4.1 Dendrogram analysis 85
5.4.2 Scattering mechanism analysis 87

6 RESULTS 89
6.1 Experiment I - initial parameter determination 89
6.2 Experiment II - covariance arithmetic mean versus covariance intrinsic mean 99
6.3 Experiment III - simulated PolSAR image classification 107
6.4 Experiment IV - real PolSAR image classification 114
6.4.1 Real PolSAR image qualitative analysis 114
6.4.1.1 Bebedouro SIR-C in C band image qualitative analysis 114
6.4.1.2 Bebedouro SIR-C in L band image qualitative analysis 121

6.4.1.3 Bebedouro SIR-C in C+L band image qualitative analysis 127
6.4.1.4 Tapajós ALOS/PALSAR image qualitative analysis 130
6.4.1.5 San Francisco ALOS/PALSAR image qualitative analysis 136
6.4.2 Real PolSAR data quantitative analysis 143
6.4.2.1 Bebedouro SIR-C in C band image quantitative analysis 143
6.4.2.2 Bebedouro SIR-C in L band image quantitative Analysis 145
6.4.2.3 Tapajós ALOS/PALSAR image quantitative analysis 147
6.4.2.4 San Francisco ALOS/PALSAR image quantitative analysis 150

7 CONCLUSIONS 153
7.1 Future works 156

REFERENCES 159

APPENDIX A - BSC FOR IMAGES FOLLOWING THE GAUSSIAN DISTRIBUTION. 171
A.1 Initial parameter determination 171
A.1.1 PDDP 171
A.1.2 EM 172
A.2 Stochastic distances 173
A.3 Multivariate Gaussian entropy 173
A.4 Results 174

APPENDIX B - BSC WEBPAGE AND HOW RUN THE CODE. . 179

1 INTRODUCTION

Remote sensing area has progressed rapidly due to the increasing availability of satellites for Earth observation and the continuous sophistication of sensor systems and algorithms for processing these data. Therefore, images nowadays have better spatial, temporal, spectral, and radiometric resolutions.

Currently, there are a great number of satellites operating in diverse spectral bands (for instance, optical, thermal, and microwave), supplying the databases with millions of images every day - the European Space Imaging agency alone collects over three million square kilometers of imagery daily (IMAGING, 2019). This abundance of information requires a constant improvement of algorithms for processing the data, and this availability of different types of sensors and image formats allow us to understand the physicochemical properties of objects in a more complete way, since each wavelength interacts in a particular manner with them. Moreover, different wavelengths can provide complementary features. For instance, while optical images have a visual response closer to the humans vision, microwave sensors can be used almost independently of weather conditions and sunlight, and, depending on the wavelength, clouds, dust, or even the vegetation canopy could be transparent to this frequency spectrum.

Synthetic Aperture Radar (SAR) is an active microwave sensor on board either space-borne or air-borne. SAR images have an important role in socioeconomic applications, helping on the environmental monitoring and surveillance activities. SAR systems have potential in a diverse range of applications, for instance, sea and ice monitoring (DIERKING; BUSCHE, 2006), oil pollution (ARCGIS, 2019), vessel monitoring (WU et al., 2016), flood monitoring (KUSSUL et al., 2011), urban growth monitoring (HENDERSON; XIA, 1997), classification of earth terrain (LIU et al., 2018), among others.

Usually SAR systems have two major product groups: basic image products and enhanced image products. The basic image product is the raw data, which consists of a In-Phase and Quadrature signal data that can be processed to become a single look complex (SLC) image. The enhanced image products represent a higher level of processing in terms of a more precise geometric correction, mosaicking, cloud thickness and, vegetation indices.

The SAR antenna can allow different polarization, such as linear - horizontal (h), vertical (v) - or circularly polarization in addition to a variety of polarimetric combi-

nations: single, dual or quad. The use of PolSAR (Polarimetric Synthetic Aperture Radar) images, which measures the target backscattering in different polarization, provides further information such as soil moisture, surface roughness, target shape, and object geometry.

A SAR system containing quad polarization, i.e. having the polarization hh, hv, vh, and vv, is known as Fully PolSAR. As already discussed, PolSAR systems can provide a rich set of information about objects on land surfaces, and high spatial resolution images. However, SAR images have an interference pattern called speckle, which causes a granular texture in SAR image. The speckle provides a rich statistical information that can be essential to good comprehension about the scattering mechanisms. Hence, a huge effort to characterize the statistical properties of PolSAR images has been done by (FRERY et al., 2013; FRERY et al., 1999; NASCIMENTO, 2012). The statistical analysis can improve the PolSAR image interpretation and help the development of smart algorithms for speckle filtering (TORRES et al., 2014), segmentation (SALDANHA, 2013; DOULGERIS, 2014; DOULGERIS; ELTOFT, 2014), feature extraction (YANG et al., 2012), and classification (BRAGA et al., 2015).

Classification is one of the main topics of scientific research, and image classification techniques appear as important tools for environmental preservation, national security and management of natural resources. Due to this topic importance, many Earth observation programs invest heavily in image classification. In Brazil, the National Institute for Space Research (INPE) develops and coordinates diverse remote sensing and geoprocessing programs that have in their scope several image classification techniques, such as PRODES (*Programa de Monitoramento da Floresta Amazônica Brasileira por Satélite*), the DETER (*Detecção de Desmatamento em Tempo Real*), and DEGRAD (*Sistema de Monitoramento de Áreas de Florestas Degradadas na Amazônia*), among others. In addition, INPE also develops software that contains image classification algorithms, such as TerraAmazon, Spring, and TerraView (INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. COORDENAÇÃO GERAL DE OBSERVAÇÃO DA TERRA - INPE/OBT, 2016).

A large number of approaches have been proposed for SAR image classification. Many methods are based on incomplete polarimetric information, such as the intensities or amplitude of hh, hv, vh or vv; or on the phase difference between hh, hv, vh or vv, among other variants. The use of PolSAR images for classification of the Earth terrain has been widely explored since 1980 decade. For instance, one of the first classification using the three complex polarimetric components hh, hv, and vv

was performed by Kong et al. (1988). They applied the Bayes classification using the a multivariate complex Gaussian distribution.

However, multilook PolSAR images are extensively known for following the Wishart distribution, for this reason, Lee et al. (1994) developed a maximum likelihood classifier based on the complex Wishart distribution. Another famous PolSAR image classifier is the H-α decomposition, proposed by Cloude e Pottier (1997). This algorithm is based on the eigenvalues and eigenvectors decomposition of the coherency matrix $\boldsymbol{T}$, and it has two parameter, the entropy (H) and α angle. The entropy can be related with the number of scattering mechanism within a PolSAR resolution cell and the α angle can be understood as the type of scattering mechanism. This two feature are combined into a two-dimensional space, known as Plan H-α, which is divided into scattering mechanism classification zones. Pottier e Cloude (1997) introduced another component to the H-α decomposition: the anisotropy (A). The anisotropy is used to discriminate the second and third eigenvalues, helping on the characterization of the scattering mechanisms.

There are many tools addressing PolSAR image processing, such as PolSARpro (POTTIER, 2019), and RAT (REIGBER; HELLWICH, 2004). Both PolSARpro and RAT offer as classification tools the supervised Wishart classification, the unsupervised H-α, A-α, H-A, and the H-A-α classifications.

1.1 Motivation

PolSAR images are usually represented by its second-order polarimetric representation, the covariance matrix. There are many applications for PolSAR covariance images; it can be used to estimate the scattering mechanism present at the image (FREEMAN; DURDEN, 1998), or it can be used as input for PolSAR image classification (LEE et al., 1999).

The use of covariance matrix as a feature has been widely explored in recent years, especially in the field of pattern recognition and machine learning. As a consequence, these matrices are explored exhaustively in relation to their space. According to Lim et al. (2019), real symmetric positive definite or complex Hermitian positive definite matrices have a natural Riemannian metric tensor, therefore covariance PolSAR images are better clustered by a non-flat geometry metric, as some hierarchical algorithms. Besides the advantage of working with arbitrary cluster shapes, hierarchical algorithms do not need to specify the number of clusters, and the dendrogram produced is very useful in understanding the data distribution.

In literature, there are some works approaching the semi-supervised hierarchical clustering for SAR image (YAO et al., 2016; SAPUCCI; NEGRI, 2019) and unsupervised hierarchical clustering for SAR images (ZHAO et al., 2017; CERRA; DATCU, 2010). In all these works, the hierarchical clustering showed to have a great potential to deal with SAR images, with particular regards to the algorithms that use the global merging criterion. Improvements in this field are still needed, particularly for more efficient and accurate approaches.

Another interesting feature of PolSAR images are their speckle statistical information, which can be explored by attributes derived from Information Theory, such as divergences (SONG et al., 2014), entropy (FRERY et al., 2012), and stochastic distances (SILVA et al., 2012). Stochastic distances are dissimilarity measures used for evaluating the contrast between two Probability Density Functions (PDF). Since PolSAR images are widely known for following the Wishart distribution, we consider the distances as stochastic distances between Wishart Distribution.

Stochastic distances are derived from divergence, which is a concept introduced by Kullback e Leibler (1951). There are many divergence families, and Nascimento (2012) used the divergence class $h - \phi$ to develop five stochastic distance between Wishart distributions: Bhattacharyya, Kullback-Leibler, Hellinger, Rényi of order β, and Chi-square.

In the study conducted in Carvalho et al. (2019), these five stochastic distances, plus the Euclidean distance, and the Wishart mixture model were compared. The K-means algorithm, when using stochastic distances as similarity metric, were named Stochastic Clustering (SC). In order to handle PolSAR images, the Expectation-Maximization (EM) algorithm were applied to the Wishart mixture. A Monte Carlo simulation was performed over a set of simulated PolSAR images to find all possible classification outcomes for each algorithm (K-means using the Euclidean distance, SC using the Bhattacharyya distance, SC using the Kullback-Leibler distance, SC using the Hellinger distance, SC using the Rényi of order β distance, SC using the Chi-square distance, and EM for Wishart mixture model). The Monte Carlo simulation had the goal of computing the classification overall accuracy in order to produce a quantitative analysis of the results.

According to the Monte Carlo simulation results, the EM showed to have great potential, classifying the homogeneous areas with small errors, while the SC algorithm accuracy performance was highly dependent on the stochastic distance choice. The stochastic distances Bhattacharyya, Kullback-Leibler, and Hellinger had strong and

similar results, having a clearly superior performance over the Euclidean distance. However, the Rényi of order β and Chi-square distances had the lowest classification accuracy results, primarily due to numerical instabilities presented by both distances.

The Monte Carlo simulation also exposed the dependency of all algorithms upon the correctness of the initial centroids. Bad initial parameter make the algorithms terminate at a local maximum or minimum, resulting in the non-optimized estimation. Therefore, a better initial guess would make the clustering convergence faster. There are a large number of methods for choosing initial centroids. Celebi et al. (2013) presented a comparative of commonly used methods for clustering initialization, such as the Ball and Hall's method, Simple Cluster Seeking method, Maximin method, among others.

PolSAR image classification using stochastic distance as a similarity measure is a relatively new field in SAR image processing, and has been growing in recent years. One of most significant works is this area was presented by Silva et al. (2013). In this work, the authors proposed a supervised classification algorithm to classify segments of PolSAR images using hypothesis test statistics derived from stochastic distances between Wishart distributions. Gomez et al. (2017) presented a supervised classification algorithm using the maximum likelihood and stochastic distance for classification of PolSAR images. Negri et al. (2019) combined support vector machines and stochastic distances to perform a region-based classification. Although great progress has been achieved in this field, there is still room for investigating the potential of stochastic distances applied on PolSAR data, especially when allied with unsupervised classification.

1.2 Objectives

This work has as its main goal the development and presentation of a new algorithm for PolSAR image unsupervised classification, named Bisecting Stochastic Clustering (BSC). The BSC is a hybrid algorithm of hierarchical and partitional algorithms, and it uses stochastic distances as similarity metric. It explores the handling of PolSAR covariance matrix under the Riemannian geometry and applies the information gain to define the best cluster to split.

1.2.1 Specific objectives

- Demonstration of the initial parameter estimation importance for unsupervised classification algorithms;
- Proposing and developing a new approach named Riemannian Principal Direction Divisive Partitioning (RPDDP) to determine the initial parameter on hierarchical divisive clustering algorithm for PolSAR images;
- Development of a new algorithm named Stochastic Clustering (SC);
- Study about the stochastic clustering robustness, showing the distances with better performance when applied on PolSAR image;
- Comparison between the covariance matrix handling in the Riemann geometry against the Euclidean geometry.

1.3 Thesis hypothesis

Based on the objectives of this work, two hypothesis have been defined:

- it is possible to employ the Riemannian geometry in determining initial parameters of PolSAR data represented by their covariance matrices, with the aim of mitigating the effects of the arithmetic mean of covariance matrices, and improving the quality and efficiency of divisive hierarchical groupings, for the development of unsupervised PolSAR image classifiers.
- it is possible to develop a robust unsupervised classifier for PolSAR data based on divisive hierarchical clustering algorithms and using stochastic distances.

1.4 Contributions

This work contributes to the PolSAR image unsupervised classification characterization by:

- Doing a quantitative comparison of several stochastic distances between Wishart distribution, against the mixture models, exploring the stochastic distances weaknesses and strengths.

- Studying the impacts of initial parameter determination on the clusters quality and efficiency;
- Analysing of the pros and cons of Riemann geometry usage in the PolSAR covariance matrices manipulation;
- Proposing and developing a new PolSAR unsupervised classifier, based on hierarchical divisive clustering algorithms, using as dissimilarity metric the stochastic distances (This algorithm handles the polSAR covariance matrices in the Riemann space);
- Analysing the hierarchical divisive clustering algorithms classification outputs and dendrogram using the H-α plan, in order to have a better understanding of the PolSAR scattering mechanism relationship with clustering.

1.4.1 Research items

Published journal paper:

CARVALHO, N. C. R. L.; BINS, L. S.; SANT'ANNA, S. J. S. Analysis of Stochastic Distances and Wishart Mixture Models Applied on PolSAR Images. **Remote Sensing**, V. 11, 2994, 2019. Disponível em: https://doi.org/10.3390/rs11242994.

Published conference article:

CARVALHO, N. C. R. L.; BINS, L. S.; SANT'ANNA, S. J. S. Monte Carlo Simulation study of stochastic distances applied on k-means algorithm for Fully Polarimetric SAR images. In: POLinSAR, Frascati (Rome), Italy, 2019.

CARVALHO, N. C. R. L.; BINS, L. S.; SANT'ANNA, S. J. S.Incoherent Polarimetric target scattering decomposition: An overview and their implementation in TerraLib system. In: POLinSAR, Frascati (Rome), Italy, 2019.

Mini-Course:

Fundamentos de Imagens SAR, avilable at PGCAP youtube chanel: https://youtu.be/nWh12Y89JUs.

Journal paper work in progress:

CARVALHO, N. C. R. L.; BINS, L. S.; SANT'ANNA, S. J. S. Riemann PDDP for PolSAR image classification.

CARVALHO, N. C. R. L.; BINS, L. S.; SANT'ANNA, S. J. S. Bisecting SC: a new algorithm for PolSAR image classification.

CARVALHO, N. C. R. L.; BINS, L. S.; SANT'ANNA, S. J. S. Analysis of Bisecting SC using Wishat Mixture Model as the initial parameter estimator.

1.5 Thesis organization

This work is organized as follows. Chapter 2 describes the PolSAR image description and its representation, the speckle generation, its statistical representation and the stochastic distance formulation. In Chapter 3 the clustering algorithms are described, we discuss about the user's dilemma, cluster taxonomy and pros and cons of some clustering techniques. Chapter 4 describes the proposed algorithm, BSC, we present all the needed steps in this algorithm and some variants. Chapter 5 presents the methodology and describes the PolSAR images used to evaluate the BSC classifier. Chapter 6 shows the comparison between the Riemann geometry and Euclidean geometry, presents the initial parameter estimation importance and shows the classification results using SC and BSC algorithms applied on the set of simulated images and real PolSAR images. Finally, our conclusions and future work proposal are presented in Chapter 7.

2 SAR IMAGE

Radar (Radio Detection and Ranging) is an electromagnetic sensor, which operates in the microwave spectrum, for detection and localization of objects in space (SKOLNIK, 2008). As presented in Figure 2.1, the radar radiates electromagnetic energy from its antenna - a common waveform radiated by this sensor is a series of narrows rectangular pulses - to propagate through space. A parcel of the radiated energy is intercepted by a reflective target and re-radiated in many directions, and a small portion of the reflected energy returns to the radar. The returned signal is amplified, and based on the time difference between the transmitted and received signal, the target radial position is determined.

Figure 2.1(a) shows the radar operation, where r is the target range distance and t is the time delay between the received and transmitted signal. Figure 2.1(b) shows the transmitted, received and processed signals. Based on time delay between transmitted and received signals, the target radial position is calculated.

Figure 2.1 - Radar time and distance measurement.

(a) The radar operation. (b) Transmitted, received and processed signal.

SOURCE: Author's production.

Radars are used for many applications, one of the most important being the imag-

ing. Usually, imaging radar operates with lateral or side-looking geometry. In this configuration, the platform travels with speed V_p towards the flight path (in x-axis) with the nadir directly below the platform (on the z-axis), while electromagnetic waves are transmitted obliquely to the flight direction (WOLFF, 2018), as showed in Figure 2.2.

Figure 2.2 - Imaging radar geometry.

SOURCE: Author's production.

The radar antenna's footprint size, i. e. the illuminated area on the ground, depends on the azimuth ψ_a and elevation ψ_e aperture angles, the sensor's platform height H, and the angle α between slant range and nadir, as shown in Figure 2.3. The imaging radar produces a two-dimensional image representing the footprint of its antenna. The amount of detail on this image depends on the radar azimuth and range resolution.

The azimuth resolution of real aperture radars with antenna length d_a and wavelength λ is given by the equation (LEE; POTTIER, 2009):

$$\delta_{ra} = \frac{\lambda}{d_a}.R \tag{2.1}$$

where R is the radial distance between the platform and ground.

Figure 2.3 - Imaging radar footprint representation.

SOURCE: Author's production.

It can be seen from Equation (2.1) that the azimuth resolution is inversely proportional to d_a, meaning that the bigger the antenna length, the better the azimuth resolution. However, due to physical limitations, the enlarging of antenna length could be impractical in many remote sensing situations. In order to overcome this problem, Wiley (1985) proposed the use of an antenna with modest size, moving along path sending pulses, whose echos would be compressed, proving a finer azimuth resolution. In this way, a small moving antenna electronically simulates a large antenna; this process is called SAR. The azimuth resolution for SAR system becomes:

$$\delta_{sa} = \frac{d_a}{2} \tag{2.2}$$

and therefore, the azimuth resolution is no longer dependent on wavelength or target range (MOREIRA et al., 2013), and the smaller the antenna length, the better the azimuth resolution.

Even thought the SAR method was developed to improve the azimuth resolution, in order to have an image with great level of detail, the range resolution should be improved as well. In order to have a finer range resolution, the pulse bandwidth shall be wide. One way to achieve a fairly wide bandwidth is to use a step frequency waveform, also known as a chirp pulse. The slant range resolution of a chirp waveform with bandwidth B is defined as (SKOLNIK, 2008):

$$\delta_r = \frac{c}{2.B} \tag{2.3}$$

where c is the light speed.

A SAR image is 2D raster, where each pixel represents a small portion of the Earth's surface, named resolution cell, whose size depends on range and azimuth resolutions. Each pixel carries the amplitude and phase information of the backscattered signal. These values depend on the target radar cross-section, reflectivity coefficient, surface topography, roughness, dielectric properties of the medium, and moisture; but also depends on radar parameters such as frequency, polarization, and incident angle, among others (LEE; POTTIER, 2009).

2.1 Speckle generation

The resolution cell is the smallest area of Earth's surface that can be represented by one pixel in a SAR image. Within this area, there are many different targets or scatterers located in random positions. The SAR emitted wave interacts with the scatterers and the total received signal S is the coherent sum of the s_w signals backscattered by the W scatterers within a resolution cell, as defined by Equation (2.4):

$$S = \sum_{w=1}^{W} s_w e^{\phi_w} = s_a + j s_b \tag{2.4}$$

where $j = \sqrt{-1}$, $s_a = \sum_{w=1}^{W} s_w \cos(\phi_w)$ is the total received signal in-phase component, while $s_b = \sum_{w=1}^{W} s_w \sin(\phi_w)$ is the quadrature component; s_w and ϕ_w are the wave amplitude and phase of individual scatterers.

Since the scatterers' distance to SAR sensor is random, the received waves are coherent in frequency but not in phase (LEE; POTTIER, 2009). Therefore, if the received waves are constructively added, the signal will be strong. However, if the received waves are out of phase, then the signal will be destructively added, resulting in

a weak signal, as shown in Figure 2.4. This phenomenon, called speckle, causes a granular pattern in SAR images.

Figure 2.4 - Speckle formation.

SOURCE: Adapted from Palumbo (2017).

The pixel-to-pixel variation in intensities that appear on SAR images due to speckle can lead to a reduction in classification accuracy and segmentation effectiveness. Over the decades many methods were developed in order to mitigate speckle. For instance, the multi-look technique can be done in spatial domain (non-coherent averaging of adjacent pixels intensity of an image), in time domain (divide the synthetic aperture and process each segment separately), or in frequency domain (HELLMAN, 2000). Further algorithms in spatial domain are the Local Statistics Filter (LEE, 1980), the Refined Local Statistics Filter (LEE, 1981), the Sigma Filter (LEE, 1983), the Frost Filter (FROST et al., 1982), the Kalman filtering (AZIMI-SADJADI; BANNOUR, 1991), and more recently the Nonlocal Means (TORRES et al., 2014), among others.

Another factor that influences the speckle statistical modeling is the scattering mechanism. Roughly, the SAR backscattering can be defined into three scattering mechanisms: smooth surface scattering (such as lakes or pasture), double bounce (such as urban areas) and the volume scattering (such as forests).

Surface roughness is defined by the Rayleigh Criterion, which says that a surface is smooth if:

$$\delta_h < \frac{\lambda}{8 \cos \theta_i} \tag{2.5}$$

where δ_h is the standard deviation of surface roughness, λ is the radar wavelength, and θ_i is the incident angle (LÓPEZ-MARTÍNEZ et al., 2005). The signal which interacts with smooth surfaces is mostly reflected away from the radar and only a slight backscatter is received by the radar, as shown in Figure 2.5. Therefore the resolution cells of smooth surfaces usually have a dark tonality on radar images (EUROPEAN SPACE AGENCY - ESA, 2017).

Among diverse factors, the backscattering intensity can be seem as a factor of scattering mechanism type, for instance, resolution cells with volume scattering have an intermediate backscatter, and resolution cells with double bounce scattering have very strong backscatter. Therefore, the higher the backscattered intensity, the rougher the imaged surface, as presented in Figure 2.5.

Figure 2.5 - Reflectiveness of different scattering mechanism.

SOURCE: Adapted from Nascimento (2012).

2.2 PolSAR images

SAR systems use electromagnetic waves, in microwave spectrum, to determine the pixel value of a radar image; therefore, the electromagnetic wave carries all the important information between the radar and observed target. Electromagnetic waves can be characterized by their frequency, amplitude, phase, direction of propagation, and polarization. Wave polarization happens when the electrical field distorts the electrons' cloud in a particular direction, i. e., the polarization refers to electric ($\vec{E}$)

and magnetic ($\vec{B}$) field alignment in a plane perpendicular to propagation direction.

Under assumption that, in the Cartesian coordinate system, the electromagnetic waves propagate in $z-$direction, the electromagnetic field is located in the plan $x - y$, where:

$$\vec{E} = E_x + E_y \tag{2.6}$$

with

$$\vec{E_x} = E_{x_0} \cos(\omega t - kz + \delta_x) \tag{2.7}$$

$$\vec{E_y} = E_{y_0} \cos(\omega t - kz + \delta_y) \tag{2.8}$$

where $\omega = 2\pi f$, $k = \frac{2\pi}{\lambda}$ is the wave number, f is the signal frequency, λ is the wavelength, and δ is phase angles.

Depending on the behavior of amplitudes E_0 and angular phase δ in a plane perpendicular to the propagation direction, three types of polarization can be derived:

- Elliptical polarization: the general case of electromagnetic wave polarization with $E_{x0} \neq 0$ and $E_{y0} \neq 0$ and $\delta_x - \delta_y \neq 0$;
- Circular polarization: with $E_{x0} = E_{y0}$ and $\delta_x - \delta_y = \dfrac{\pi}{2}$;
- Linear polarization: with $\delta_x = \delta_y$, meaning that the electric field is confined to a single plane along the propagation direction.

The most common forms of linear polarization are horizontal linear h and vertical linear v. The SAR antenna may be designed to transmit and receive waves at more than one polarization, therefore, the SAR system can have the following channels (CANADA, 2014):

- hh - for horizontal transmit and horizontal receive;
- vv - for vertical transmit and vertical receive;
- hv - for horizontal transmit and vertical receive;

- vh - for vertical transmit and horizontal receive.

A SAR system that incorporates the linear polarization combinations between vertical and horizontal directions for transmitted and received waves is named PolSAR. Depending on the levels of polarization complexity, a PolSAR system can be classified as:

- Single polarized - hh or vv or hv or vh;
- Dual polarized - different channels combination, for instance, hh and vv, hh and hv, among others;
- Full polarization having hh, vv, hv, and vh;

In linear polarization context, a scatterer is illuminated by a transmitted wave, such that the incident wave is given by (HELLMAN, 2000):

$$\vec{E}^t = \vec{E}^t_h + \vec{E}^t_v \tag{2.9}$$

and the received wave is given by (LEE; POTTIER, 2009):

$$\vec{E}^r = \boldsymbol{S}\vec{E}^t = \frac{e^{jk_0 r}}{r} \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \begin{bmatrix} E^t_h \\ E^t_v \end{bmatrix} \tag{2.10}$$

where term $\frac{e^{jk_0 r}}{r}$ refers to the propagation effects both in amplitude and phase, $\boldsymbol{S}$, named as scattering matrix or Jones matrix, is a matrix of complex scattering amplitudes, with $S_{xy} = \{S_{xy}\{e^{j\phi_{xy}}\} \mid x, y \in h, v\}$. The elements on $\boldsymbol{S}$ main diagonal are named 'co-pol', and the elements on secondary diagonal are named 'cross-pol'.

The scattering behavior of targets in matrix $\boldsymbol{S}$ depends on, among other things, the coordinate systems. In the coordinate system named Back Scattering Alignment (BSA), the transmitting and receiving antennas are collocated in space. Due to this, electromagnetic waves have a reciprocal medium. In the BSA coordinate system, the reciprocity theorem says that the cross-pol channels of the scattering matrix are equal, i.e., $S_{hv} = S_{vh}$.

Usually, for the studying of PolSAR behavior, it is more convenient to represent the matrix $\boldsymbol{S}$ in vector basis, as defined in Equation (2.11).

$$\vec{k} = \frac{1}{2} Tr(\boldsymbol{S}\boldsymbol{\Psi}) \tag{2.11}$$

where $Tr(.)$ is the matrix trace and $\boldsymbol{\Psi}$ is a set of 2×2 complex basis matrices. There are two basis widely known in PolSAR images field: the Borgeaud basis and the Pauli basis. The Borgeaud basis follows the lexicographic expansion of $\boldsymbol{S}$, with:

$$\boldsymbol{\Psi_b} = \left\{ 2\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, 2\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, 2\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, 2\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \tag{2.12}$$

and the corresponding vector $\vec{k_b}$:

$$\vec{k_b} = \begin{bmatrix} S_{hh} & S_{hv} & S_{vh} & S_{vv} \end{bmatrix}^{\dagger} \tag{2.13}$$

considering that $S_{hv} = S_{vh}$, $\vec{k_b}$ becomes:

$$\vec{k_b} = \begin{bmatrix} S_{hh} & \sqrt{2}S_{hv} & S_{vv} \end{bmatrix}^{\dagger} \tag{2.14}$$

where † denotes the conjugate transpose.

The latter group is the complex Pauli spin matrix basis set given by:

$$\boldsymbol{\Psi_p} = \left\{ \sqrt{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \sqrt{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \sqrt{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sqrt{2}\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \right\} \tag{2.15}$$

and the corresponding vector $\vec{k_p}$:

$$\vec{k_p} = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{hh} + S_{vv} & S_{hh} - S_{vv} & S_{hv} + S_{vh} & i(S_{vh} - S_{hv}) \end{bmatrix}^{\dagger} \tag{2.16}$$

considering that $S_{hv} = S_{vh}$, $\vec{k_p}$ becomes:

$$\vec{k_p} = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{hh} + S_{vv} & S_{hh} - S_{vv} & 2S_{hv} \end{bmatrix}^{\dagger} \tag{2.17}$$

One of the most usual vector basis of $\boldsymbol{S}$ in polarimetric applications are the Pauli basis ($\vec{k_p}$), from which the coherence matrix $\boldsymbol{T}$ is determined (Equation (2.19)). The coherence matrix of a PolSAR image is computed by the product between the Pauli vector $\vec{k_p}$ and its transposed conjugate complex, denoted by $\vec{k_p}^T$, as presented in

Equation (2.18).

$$\boldsymbol{T} = \vec{k_p}.\vec{k_p}^{\dagger} \tag{2.18}$$

$$\boldsymbol{T} = 0.5 \begin{bmatrix} |S_{hh}+S_{vv}|^2 & (S_{hh}+S_{vv}).(S_{hh}-S_{vv})^* & 2(S_{hh}+S_{vv}).(S_{hv})^* \\ (S_{hh}-S_{vv}).(S_{hh}+S_{vv})^* & |S_{hh}-S_{vv}|^2 & 2(S_{hh}-S_{vv}).(S_{hv})^* \\ 2S_{hv}.(S_{hh}+S_{vv})^* & 2S_{hv}.(S_{hh}-S_{vv})^* & 4|S_{hv}|^2 \end{bmatrix} \tag{2.19}$$

where $(.)^*$ represents the complex conjugate and $|.|$ means the absolute value of a given number.

Usually, for the studying of PolSAR statistical behavior, it is more convenient to represent the matrix $\boldsymbol{S}$ by the lexicographic basis vector $\vec{k_b}$ (DENG et al., 2017). The covariance matrix of a PolSAR image is computed by the product between the scattering vector $\vec{k_b}$ and its transposed conjugate complex, denoted by $\vec{k_b}^T$, as presented in Equation (2.20).

$$\boldsymbol{Z} = \vec{k_b}.\vec{k_b}^{\dagger} \tag{2.20}$$

The matrix $\boldsymbol{Z}$ is a Hermitian positive semidefinite matrix, where the elements of its main diagonal are positive real numbers and they correspond to the values of intensity. Elements outside the main diagonal are complex numbers and contain, in addition to the product information between the amplitudes of the polarimetric components, the phase difference information between the transmitted and received signals in the different polarization.

Since PolSAR data is affected by speckle, a common approach to reduce it is to perform multi-look processing. From the statistical point of view, multi-look processing is defined as the averaging of L neighboring samples, resulting in the covariance matrix $\boldsymbol{Z}$ (LEE; POTTIER, 2009), as defined in Equation (2.21):

$$\boldsymbol{Z} = \frac{1}{L}\sum_{\ell=1}^{L} \begin{bmatrix} |S_{hh}(\ell)|^2 & \sqrt{2}S_{hh}(\ell).S_{hv}^*(\ell) & S_{hh}(\ell).S_{vv}^*(\ell) \\ \sqrt{2}S_{hv}(\ell).S_{hh}^*(\ell) & 2|S_{hv}(\ell)|^2 & \sqrt{2}S_{hv}(\ell).S_{vv}(\ell)^* \\ S_{vv}(\ell).S_{hh}^*(\ell) & \sqrt{2}S_{vv}(\ell).S_{hv}^*(\ell) & |S_{vv}(\ell)|^2 \end{bmatrix} \tag{2.21}$$

where $(.)^*$ represents the complex conjugate and $|.|$ means the absolute value of a given number.

2.3 PolSAR target decomposition

The PolSAR images target decomposition consists of identifying and isolating different types of scattering mechanisms. This is an important field of study for targets characterization and feature extraction. Generally, the target decomposition could be split into two main groups: coherent decomposition and incoherent decomposition.

2.3.1 Coherent decomposition

The so-called coherent targets are those which do not depolarize the wave when interacting with it, such as the corner reflectors used for radar calibration. In the coherent decomposition, the scattering matrix $\boldsymbol{S}$ is decomposed into a sum of complex elements and each element represents a certain canonical scattering mechanism, as described in the Equation (2.22).

$$\boldsymbol{S} = \sum_{i=1}^{k} c_i \boldsymbol{S}_i \tag{2.22}$$

where where $\boldsymbol{S}_i$ represents the scattering matrix of a canonical target and c_i is the contribution of the canonical target to obtain the final target.

Examples of coherent target decomposition are Cameron decomposition and Krogager decomposition. Figure 2.6 shows a representation of the coherent decomposition. In this figure, the target is composed by a cylinder, a sphere, and a dihedral.

Figure 2.6 - Coherent Decomposition representation.

SOURCE: Author's production.

2.3.2 Incoherent decomposition

Incoherent targets are those with random backscatter, which produces partially polarized or completely depolarized waves. Natural targets in the scene, as trees or buildings, are know as incoherent targets, and they are better analyzed by the statistical point of view. The incoherence target decomposition can be split into two groups: the first one are the decomposition based in models, such as the Freeman-Durden algorithm (FREEMAN; DURDEN, 1998); and the second group is the one based on eigenvalues and eigenvectors analyzes, for example, the Cloude-Pottier decomposition (CLOUDE; POTTIER, 1996).

2.3.2.1 Freeman-Durden decomposition

The Freeman-Durden decomposition seeks to estimate the canonical model's contribution in the total backscattered answer of PolSAR images (FREEMAN; DURDEN, 1998). The three mechanisms that could be identified by this model are: volume or canopy scatter from a cloud of randomly oriented dipoles, even or double-bounce scatter from a pair of orthogonal surfaces with different dielectric constants, and Bragg scatter from a moderately rough surface (LEE; POTTIER, 2009). This model could be described by the set of Equations (2.23)

$$\begin{aligned}
\langle |S_{hh}|^2 \rangle &= f_s|\beta|^2 + f_d|\alpha|^2 + fv \\
\langle |S_{vv}|^2 \rangle &= f_s + f_d + fv \\
\langle S_{hh}S_{vv}^* \rangle &= f_s\beta + f_d\alpha + fv/3 \\
\langle |S_{hv}|^2 \rangle &= f_v/3 \\
\langle S_{hh}S_{hv}^* \rangle &= \langle S_{hv}S_{vv}^* \rangle = 0
\end{aligned} \tag{2.23}$$

where f_v corresponds to the volume scattering component contribution, f_d corresponds to the double-bounce scattering component contribution, f_s corresponds to the surface scattering component contribution. If $\Re(S_{hh}S_{vv}^*)$ is positive, then $\alpha = -1$. If $\Re(S_{hh}S_{vv}^*)$ is negative, then $\beta = 1$.

The Freeman-Durden models the covariance matrix $\boldsymbol{C}$ as a composition of f_s, f_d, and f_v, as presented in Figure 2.7. The output of this decomposition will be the contribution of each canonical mechanism in each cell of the analyzed image, in other words, the output will be three parameters defined as P_s, P_d and P_v, where:

$$
\begin{aligned}
P_s &= f_s(1+|\beta|^2) \\
P_d &= f_d(1+|\alpha|^2) \\
P_v &= 8.f_v/3
\end{aligned}
\tag{2.24}
$$

Figure 2.7 - Freeman-Durden canonical mechanism and their contribution to the final covariance matrix.

SOURCE: Author's production.

2.3.2.2 Cloude-Pottier decomposition

The Cloude-Pottier decomposition performs the analysis of a 3×3 coherency matrix $\boldsymbol{T}$ eigenvalues to decide how many mechanisms contributes to the total backscattering of a cell and its magnitude (LEE; POTTIER, 2009). The decomposition output are two polarimetric parameters: Entropy (H) (Equation (2.25)), and the Alpha Angle (α) (Equation (2.26)) (CLOUDE et al., 2002).

$$
H = -\sum_{i=1}^{3}(P_i \log_3 P_i), \quad P_i = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \lambda_3} \tag{2.25}
$$

where λ_i are the eigenvalues of $\boldsymbol{T}$ and P_i is a pseudo-probability determine by the Bernoulli process.

$$
\alpha = \sum_{i=1}^{3}(P_i \cos^{-1}(k_{xi})) \tag{2.26}
$$

where k_{xi} is the first element of the eigenvector associated with the eigenvalue λ_i.

The Entropy H represents the scattering randomness, it varies from 0 to 1. $H = 0$ means that only one eigenvalue is nonzero, and $H = 1$ means that the three eigenvalues are equal, i. e., they equally contribute to the scattering formation. The alpha angle describes the target geometry, it varies from 0° to 90°, where $\alpha = 0^o$ represents a surface scattering mechanism, $\alpha = 45^o$ represents a volumetric mechanism and $\alpha = 90^o$ represents the double bounce scattering mechanisms. The entropy and angle alpha information can be combined into a graphical representation named Plan $H - \alpha$, as described in Figure 2.8.

Figure 2.8 - Plan $H - \alpha$.

SOURCE: Author's production.

The Plan $H - \alpha$ is divided into nine zones: z1, z2, z3, z4, z5, z6, z7, z8, and z9. The z1 region distinguishes double bounce mechanisms in a high entropy environment. The z2 region groups high entropy volume scattering mechanism. The z3 region is out of $H - \alpha$ feasible region. In the z4 region, dihedral scattering mechanism with moderate entropy can be found. The z5 region have moderate entropy with dominant dipole scattering mechanisms. The z6 region have random surface, meaning that the mechanism is like Bragg scatterer with moderate entropy. The z7 corresponds to low entropy and double or 'even' bounce scattering mechanisms. The z8 have low entropy dipole scattering mechanisms. Finally, the z9, have low entropy and Bragg

surface mechanisms.

Cloude e Pottier (1997) describes these region in more details, and they give some examples of real scenarios where the mechanisms can be found. For instance, the z1 and z2 regions mechanisms can be observed in forestry areas; z4, z5, and z8 mechanisms are related with vegetation areas, while in the z7 zone, the mechanism are provided by urban ares, for example. Lakes and very smooth land surfaces are related to z9.

2.4 SAR image statistical modeling

Although speckle causes a granular pattern in SAR images, it contains rich statistical information. Understanding SAR speckle statistic is essential to have a scattering mechanisms quality comprehension. Moreover, the statistical analysis can improve the PolSAR image interpretation by presenting the proper statistical distribution to model it, helping to develop smart algorithms for speckle filtering (TORRES et al., 2014), segmentation (DOULGERIS, 2014; DOULGERIS; ELTOFT, 2014; SALDANHA, 2013), feature extraction (YANG et al., 2012), and classification (SILVA et al., 2013; BRAGA et al., 2015; FORMONT et al., 2010; NEGRI et al., 2019). In the following subsections, the statistical models of SAR images will be explored, and the application and limitation of each kind of statistical model will be discussed.

2.4.1 Single polarization SAR images

The speckle statistic field has been explored since the 1970s, with one of the first models being proposed by Arsenault (ARSENAULT; APRIL, 1976). In this model, the resolution cell is assumed to have homogeneous surface and stationary targets, and the speckle is modeled considering the following hypotheses (GAO, 2010):

- the number of scatterers within the resolution cell is large, provided that s_a and s_b (Equation (2.4)) satisfy the Central Limit theorem;
- s_a and s_b are statistically independent;
- s_a and s_b phases are normally distributed in the interval $(-\pi, \pi)$;
- inside a resolution cell, there are no dominant scatterer (Figure 2.9(a));
- the size of a resolution cell is large enough, compared with the size of a scatterer.

The speckle formation is modeled as a random walk process in complex domain (SOBRINO, 2002), which allows the magnitude of the signal to be described as a finite sum of random phasors (LENCINA et al., 2003). Therefore, the contributions of many scatterers in a resolution cell is statistically independent in amplitude and phase. Under theses conditions, the returned signal, s_a and s_b, are assumed to be identically Gaussian distributed, with zero mean and a variance denoted as $\sigma^2/2$ (LEE; POTTIER, 2009), as presented Table 2.1.

The quantities s_a and s_b are the real and imaginary part of returned signal, and they represent the In-phase and Quadrature images, respectively. From these images a variety of other images can be derived, for instance, the image in amplitude defined as $A = \sqrt{s_a^2 + s_b^2}$, and the image in intensity defined as $I = s_a^2 + s_b^2$. The image in amplitude follows the Rayleigh distribution, with mean value $\sqrt{(\pi\sigma)}/_2$ and variance defined as $(1 - {}^{\pi}/_{4})$. The intensity image has a negative exponential distribution, with mean value equals to σ and variance σ^2.

However, in resolution cells where a strong scatterer dominates the signal response, as presented in Figure 2.9, the backscattering is represented by the coherent combination of the dominant scatterer and the clutter (LÓPEZ-MARTÍNEZ et al., 2005). The intensity image of heterogeneous regions is better described by Rician distribution (Table 2.1) (GAO, 2010), with variance $\sigma^2(1+2\times SRC)$, where SRC is the Signal to Clutter Ratio. Figure 2.9(a) shows the resolution cell without a dominant scatterer. Figure 2.9(b) shows the resolution cell with dominant scatterer (blue arrow) and the clutter (black arrows).

Figure 2.9 - Coherent response of a resolution cell.

(a) No dominant scatterer.

(b) One dominant scatterer.

SOURCE: Adapted from López-Martínez et al. (2005).

Table 2.1 - Summary of statistical models applied on single polarization SAR image.

Statistical Models	**Probability Density Function**	**Application cases**
Gaussian ($\mathcal{N}$)	$f(s;\sigma) = \frac{1}{\sqrt{\pi\sigma^2}} exp\left(-\left(\frac{s}{\sigma}\right)^2\right)$	Homogeneous region, in-phase or quadrature image, single-look
Rayleigh ($\mathcal{R}$)	$f(A;\sigma) = \frac{2A}{\sigma} exp\left(-\frac{A^2}{\sigma}\right)$	Homogeneous region, image in amplitude, single-look
Exponential ($\mathcal{E}$)	$f(I;\sigma) = \frac{1}{\sigma} exp\left(-\frac{I}{\sigma}\right)$	Homogeneous region, image in intensity, single-look
Rician ($\mathcal{R}_{\rangle}$)	$f(I) = \frac{I}{\sigma} exp\left(-\frac{I}{\sigma} - SCR\right) I_0\left(2\sqrt{\frac{I \times SCR}{\sigma}}\right)$	Heterogeneous region, image in intensity, single-look
Square root Gamma ($\Gamma^{\frac{-1}{2}}$)	$f(A;L) = \frac{2L^L}{\Gamma(L)} A^{2L-1} exp\left(LA^2\right)$	Homogeneous region, image in amplitude, multi-look
Gamma (Γ)	$f(I;L) = \frac{L^L}{\Gamma(L)} I^{L-1} exp\left(-LI\right)$	Homogeneous region, image in intensity, multi-look
$\mathcal{K}$	$f(A;l,\alpha,\lambda) = \frac{4\lambda LA}{\Gamma(\alpha)\Gamma L}(\lambda LA^2)^{\alpha+L-2}\mathcal{K}_{\alpha-L}\left(2A\sqrt{\lambda L}\right)$	Heterogeneous region, image in amplitude, multi-look
$\mathcal{G}^0$	$f(A,L,\alpha) = \frac{2L^L\Gamma(L-\alpha)A^{2L-1}}{\gamma\Gamma(L)\Gamma(-\alpha)(\gamma+LA^2)^{l-\alpha}}$	Extremely heterogeneous region, image in amplitude, multi-look

SOURCE: Adapted from Gao (2010).

Frery (FRERY et al., 1997) proposed one of the greatest accomplishments in the field of speckle statistic by introducing the idea of degrees of homogeneity, where a multi-look SAR image can be divided into homogeneous, heterogeneous, and extremely heterogeneous regions. Multi-look images in amplitude having homogeneous targets are modeled as square root Gamma distribution, while the images in intensity are modeled by the Gamma distribution. For heterogeneous regions, for instance urban areas, the $\mathcal{K}$-distribution fits better, and for extremely heterogeneous regions the $\mathcal{G}^0$ is more suitable.

2.4.2 Fully PolSAR SAR images

Considering the reciprocal medium, the polarimetric complex vector can be defined as:

$$\vec{s} = \begin{bmatrix} S_{hh} & S_{hv} & S_{vv} \end{bmatrix}^{\dagger} \tag{2.27}$$

and it follows the multivariate complex Gaussian distribution:

$$f(\vec{s}; \mathbf{\Sigma}) = \frac{1}{\pi^3 |\mathbf{\Sigma}|} exp(-\vec{s}^{\dagger} \mathbf{\Sigma}^{-1} \vec{s}) \tag{2.28}$$

where $\Sigma = E(\vec{s} \times \vec{s}^{\dagger})$ is the complex covariance matrix and $\dagger$ is denotes the complex conjugate transpose. The matrix $\mathbf{\Sigma}$ contains the information needed to analyse the data from the statistical point of view.

Multilook PolSAR images containing areas with limited roughness and homogeneous regions can be modeled by the Wishart distribution. Let $Z = \{\boldsymbol{Z_1}, ..., \boldsymbol{Z_N}\}$ be a random variable data set, with N samples and L number of looks. The matrix $\boldsymbol{Z}$ follows the complex multivariate Wishart distribution (Equation (2.29)). Therefore, the PolSAR data can be described by the two Wishart parameters: the covariance matrix, where $|\boldsymbol{Z}|$ is related to image brightness, and the number of looks L, which is related to signal to noise ratio.

$$f(\boldsymbol{Z}; \mathbf{\Sigma}, L) = \frac{L^{qL} |\boldsymbol{Z}|^{L-q}}{|\mathbf{\Sigma}|^L \Gamma_q(L)} exp\{-L.Tr(\mathbf{\Sigma}^{-1} \boldsymbol{Z})\} \tag{2.29}$$

where $\mathbf{\Sigma} = E(\boldsymbol{Z})$, $E(.)$ is the expectation operator, $\Gamma_q(.)$ is the Gamma function, q is the covariance matrix order, $Tr(.)$ is the matrix trace, and $|.|$ is the matrix determinant.

According to Frery et al. (2007), the complex Wishart law can be generalized into

the $\mathcal{K}_p$ distribuiton, in order to deal with PolSAR images from heterogeneous areas. The extremely heterogeneous areas in PolSAR images can be modeled by the $\mathcal{G}_p^0$, which is a particular case of the polarimetric $\mathcal{G}_p$ law.

2.5 Stochastic distances

The information theory is the mathematical branch that deals with concepts, parameters, and rules of message transmission through communication systems. This field has been explored since 1948, when Shannon published the paper titled "A Mathematical Theory of Communication" (SHANNON, 1948). The main concepts of information theory are the entropy and information, and the primary goal of entropy is to quantify the amount of information in a data set. Given the random variable X with m different values for x_i and having a probabilistic distribution function $f(x_i)$, the entropy is defined as:

$$H(X) = \int f(x_i) \log_b \left(\frac{1}{f(x_i)} \right) dx \tag{2.30}$$

If the log has base $b = 2$, the entropy can be interpreted as the minimum number of bits it would take to encode the information (KURT, 2017). Although the entropy can provide the average rate of information produced from a stochastic process, it cannot indicate the optimal encoding scheme for data compression or whether a given PDF is more suitable to data set than another.

The information theory concepts were introduced into the statistic field by Kullback e Leibler (1951), in 1951. They developed the relative entropy concept, which computes the logarithmic difference between two distributions expectations, i. e., it measures the similarity between two PDFs. Given the random variable X, which follows the distribution f, but also fits in distribution g, the relative entropy of g relative to f is defined as:

$$H_{KL}(X) = \int f(x_i) \log_b \left(\frac{f(x_i)}{g(x_i)} \right) dx \tag{2.31}$$

The relative entropy can have different notations and names, and is often named as Kullback–Leibler divergence. In this way, a divergence is any non-negative function that gauges the contrast between two PDF(FRERY et al., 2013). Following Kullback and Leibler's work, a number of divergence classes were developed. Among then,

Salicru et al. (1993) defined the class of entropy $(h-\phi)$, defined as:

$$H_{\phi}^{h}(\boldsymbol{\theta}) = h\left(\int_{\boldsymbol{\mathcal{A}}} \phi\left(f_{\boldsymbol{X}}(\boldsymbol{X},\boldsymbol{\theta})\right) d\boldsymbol{X}\right) \tag{2.32}$$

where $\boldsymbol{X}$ is a random matrix with probabilistic density $f_{\boldsymbol{X}}(\boldsymbol{X},\boldsymbol{\theta})$, $\boldsymbol{\mathcal{A}}$ is the sample space, h is a strictly increasing function, and ϕ is a convex function.

In the context of PolSAR data, and considering that it follows the Wishart distribution, the $(h-\phi)$ class can be rewritten as (NASCIMENTO, 2012):

$$D_{\phi}^{h}(\boldsymbol{X},\boldsymbol{Y}) = h\Big(\int_{A} \phi\Big(\frac{f_{\boldsymbol{X}}(\boldsymbol{Z};\boldsymbol{\Sigma}_{\boldsymbol{X}},L)}{f_{\boldsymbol{Y}}(\boldsymbol{Z};\boldsymbol{\Sigma}_{\boldsymbol{Y}},L)}\Big) f_{\boldsymbol{Y}}(\boldsymbol{Z};\boldsymbol{\Sigma}_{\boldsymbol{Y}},L) d\boldsymbol{Z}\Big) \tag{2.33}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are random covariance matrices having density $f_{\boldsymbol{X}}(\boldsymbol{Z};\boldsymbol{\Sigma}_{\boldsymbol{X}},L)$ and $f_{\boldsymbol{Y}}(\boldsymbol{Z};\boldsymbol{\Sigma}_{\boldsymbol{Y}},L)$, respectively, with parameter $\{\boldsymbol{\Sigma}_{\boldsymbol{X}},L\}$ and $\{\boldsymbol{\Sigma}_{\boldsymbol{Y}},L\}$.

Through the combination of the functions h and ϕ, several divergences can be derived from Equation 2.32. For instance, Nascimento (2012) explored the followings measures: the Bhattacharyya, Kullback-Leibler, Hellinger, Rényi of order β, and Chi-square, as presented in Table 2.2.

Table 2.2 - $h-\phi$ divergences and related functions.

Divergence	$h(y)$	$\phi(x)$
Bhattacharyya	$-\log(-y+1)$, $0<y<1$	$-\sqrt{(}x)+{}^{x+1}/_{2}$
Kullback-Leibler	${}^{y}/_{2}$	$(x-1)\log(x)$
Hellinger	${}^{y}/_{2}$, $0<y<2$	$(\sqrt{x}-1)^2$
Rényi of order β	${}^{1}/_{\beta-1}\log((\beta-1)y+1)$, $0<y<{}^{1}/_{\beta-1}$	$\frac{x^{\beta-1}+x^{\beta}-\beta(x-1)-2}{2(\beta-1)}$, $0<\beta<1$
Chi-squared	${}^{y}/_{4}$	${}^{(x-1)^2(x+1)}/_{x}$

SOURCE: Nascimento (2012).

Although these divergences work as a good contrast estimator, they cannot be defined as similarity metric, because the symmetry property is not necessarily satisfied. To address the symmetry problem, the simplest solution to generate a symmetric

divergence, defined as (SEGHOUANE; AMARI, 2007):

$$d(\boldsymbol{X},\boldsymbol{Y}) = d(\boldsymbol{Y},\boldsymbol{X}) = \frac{D_{\phi}^{h}(\boldsymbol{X},\boldsymbol{Y}) + D_{\phi}^{h}(\boldsymbol{Y},\boldsymbol{X})}{2} \tag{2.34}$$

The function $d(\boldsymbol{X},\boldsymbol{Y})$ is defined as a distance once the following properties are satisfied:

- Non-negativity: $d(\boldsymbol{X},\boldsymbol{Y}) > 0,\ \boldsymbol{X} \neq \boldsymbol{Y}$
- Self-identity: $d(\boldsymbol{X},\boldsymbol{Y}) = 0,\ \boldsymbol{X} = \boldsymbol{Y}$;
- Symmetry: $d(\boldsymbol{X},\boldsymbol{Y}) = d(\boldsymbol{Y},\boldsymbol{X})$;

According to Frery et al. (2013), distances between random variables which follow the same PDF could be indexed by their parameters. Therefore, stochastic distances can be simplified into distances between parameters $d(\{\boldsymbol{\Sigma_X}, L\}, \{\boldsymbol{\Sigma_Y}, L\})$. Assuming an equal number of looks L for the entire analysed data set, the distance could be summarized as a distance between covariance matrices $d(\boldsymbol{\Sigma_X}, \boldsymbol{\Sigma_Y})$. Based on this notion, Nascimento (2012) derived the stochastic distances between Wishart distributions parameters presented hereafter:

a) Bhattacharyya;

$$d_{W_B}(\boldsymbol{\Sigma_X}, \boldsymbol{\Sigma_Y}) = L\left[\frac{log|\boldsymbol{\Sigma_X}| + log|\boldsymbol{\Sigma_Y}|}{2} - log\left|\left(\frac{\boldsymbol{\Sigma_X}^{-1} + \boldsymbol{\Sigma_Y}^{-1}}{2}\right)^{-1}\right|\right] \tag{2.35}$$

b) Kullback-Leibler;

$$d_{W_{KL}}(\boldsymbol{\Sigma_X}, \boldsymbol{\Sigma_Y}) = L\left[\frac{Tr(\boldsymbol{\Sigma_X}^{-1}\boldsymbol{\Sigma_Y} + \boldsymbol{\Sigma_Y}^{-1}\boldsymbol{\Sigma_X})}{2} - q\right] \tag{2.36}$$

c) Hellinger;

$$d_{W_H}(\boldsymbol{\Sigma_X}, \boldsymbol{\Sigma_Y}) = 1 - \left[\frac{|2^{-1}(\boldsymbol{\Sigma_X}^{-1} + \boldsymbol{\Sigma_Y}^{-1})^{-1}|}{\sqrt{|\boldsymbol{\Sigma_X}||\boldsymbol{\Sigma_Y}|}}\right]^{L} \tag{2.37}$$

d) Rényi of order β;

$$d_{W_R^\beta}(\boldsymbol{\Sigma_X}, \boldsymbol{\Sigma_Y}) = \frac{log2}{1-\beta} + \frac{1}{\beta-1} log\{[|\boldsymbol{\Sigma_X}|^{-\beta}|\boldsymbol{\Sigma_Y}|^{\beta-1}|(\beta\boldsymbol{\Sigma_X}^{-1} + (1-\beta)\boldsymbol{\Sigma_Y}^{-1})^{-1}|]^L + [|\boldsymbol{\Sigma_Y}|^{-\beta}|\boldsymbol{\Sigma_X}|^{\beta-1}|(\beta\boldsymbol{\Sigma_Y}^{-1} + (1-\beta)\boldsymbol{\Sigma_X}^{-1})^{-1}|]^L\} \tag{2.38}$$

e) Chi-square;

$$d_{W_{\chi^2}}(\boldsymbol{\Sigma_X}, \boldsymbol{\Sigma_Y}) = \frac{1}{4}\left\{\left[\frac{|\boldsymbol{\Sigma_X}|}{|\boldsymbol{\Sigma_Y}|^2} abs(|(2\boldsymbol{\Sigma_Y}^{-1} - \boldsymbol{\Sigma_X}^{-1})^{-1}|)\right]^L + \left[\frac{|\boldsymbol{\Sigma_Y}|}{|\boldsymbol{\Sigma_X}|^2} abs(|(2\boldsymbol{\Sigma_X}^{-1} - \boldsymbol{\Sigma_Y}^{-1})^{-1}|)\right]^L - 2\right\} \tag{2.39}$$

where q is the matrix order, $|.|$ represents the matrix determinant, $Tr(.)$ means the matrix trace, $(.)^{-1}$ indicates the inverse of a matrix, and abs$(\cdot)$ denotes the absolute value.

According to Frery et al. (2013), the number of looks L alters the data set distribution in a non-linear way, which can be perceived by the stochastic distances. As lower the number of looks, more sensitive are the stochastic distances to smaller differences between classes, leading to a noisier classification result. On the other side, as the number of looks becomes higher, more the data can be modeled as a normally distributed data.

2.6 Covariance matrices intrinsic mean

The PolSAR images can be represented by a set of covariance matrices $Z = \{\boldsymbol{Z_1}, \ldots, \boldsymbol{Z_n}\}$, with N samples and L number of looks. The covariance matrix $\boldsymbol{Z_n}$ is a Hermitian positive semidefinite matrix, which implies that the eigenvalues of $\boldsymbol{Z_n}$ are reals and non negatives, and the eigenvectors are orthogonal (LEE; POTTIER, 2009). According to Horn e Johnson (2012), a positive semidefinite matrix is positive definite if, and only if, it is nonsingular, i. e., if it has inverse. The stochastic distances defined in Section 2.5 requires nonsingular covariance matrix $\boldsymbol{Z_n}$ as input. Therefore only Symmetric Positive Definite (SPD) matrices are considered in this work.

During the PolSAR images classification or segmentation tasks it is needed to estimate a covariance matrix that represents a group or a cluster. In many cases, when the space of covariance matrices is viewed as an extrinsic convex cone in $\Re^n$,

the estimation can be done by using the sample covariance matrix, as expressed in Equation (2.40).

$$\boldsymbol{\Sigma} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{Z_n} \tag{2.40}$$

However, usually the cone of complex Hermitian positive definite matrices $\boldsymbol{C^n}$ has a natural Riemannian metric (LIM et al., 2019). Therefore, the regular sample covariance matrix is biased, resulting in an inefficient estimator. Indeed, according to Lin (2019), the Euclidean metrics are not adequate for applications with SPD matrices for two reasons:

a) The arithmetic mean of SPD matrices is not invariant under inversion;

b) The arithmetic mean of SPD matrices suffers from swelling effect, i.e., the average covariance matrix determinant is larger than any of the original determinants.

Various metrics have been introduced in the literature to handle SPD matrices, such as the Riemannian metric named Log-Euclidean Metric (LEM), introduced by Arsigny et al. (2006). In this metric, a covariance matrix $\boldsymbol{Z_n}$ forms a differentiable manifold $\mathcal{M}$, and at each point $\boldsymbol{Z_n}$ (i.e. each covariance matrix) of the manifold $\mathcal{M}$, a scalar product can be defined in the associated tangent space $\mathcal{T}_{\mathcal{Z}}\mathcal{M}$.

The tangent space $\mathcal{T}_{\mathcal{Z}}\mathcal{M}$ is Euclidean and locally homomorphic to the Riemannian manifold, therefore the metrics can be approximated by Euclidean metrics computations in the tangent space (BARACHANT et al., 2013). Figure 2.10 shows the transformation between $\mathcal{M}$ and $\mathcal{T}_{\mathcal{Z}}\mathcal{M}$. The logarithmic map $log_Z(.)$ projects the matrix $\mathbf{Z}_n$ into the tangent space, and the exponential map $exp_Z(.)$ projects the tangent space element $\mathbf{C}_n$ back to the manifold $\mathcal{M}$.

In LEM, the expectation is determined by the intrinsic mean, also known as geometric mean, which is defined as the point with minimal sum of squared distances:

$$\mu = argmin \sum_{n} \delta_R^2 dist(\boldsymbol{Z_1}, \boldsymbol{Z_2}) \tag{2.41}$$

where δ_R^2 denotes an optional weight for each sample, and $dist(\boldsymbol{Z_1}, \boldsymbol{Z_2})$ is denoted as:

$$dist(\boldsymbol{Z_1}, \boldsymbol{Z_2}) = |\log(\boldsymbol{Z}_1^{-1}, \boldsymbol{Z_2})| = \Big[\sum_i \log^2 \lambda_i\Big]^{1/2} \tag{2.42}$$

where λ_i are the real eigenvalues of $(\boldsymbol{Z}_1^{-1}, \boldsymbol{Z_2})$

Figure 2.10 - Representation of manifold $\mathcal{M}$ and the corresponding local tangent space $\mathcal{T}_{\mathcal{Z}}\mathcal{M}$ at $\mathbf{Z}$.

SOURCE: Barachant et al. (2013).

In this, work the PolSAR covariance matrices algebraic manipulation relies on Riemann geometry. The geometric mean can be computed interactively by the algorithm proposed by Moakher (2005):

Algorithm 1 Intrinsic Mean of SPD matrices.

Input: $\boldsymbol{Z}$ the set of N covariance matrices
Output: μ the estimated geometric mean

1: **procedure**
2: Initialize $\mu^1 = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{Z_n})$
3: **repeat**:
4: $\boldsymbol{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}\log_{\mu^t}(\boldsymbol{Z_n})$
5: $\mu^{t+1} = \exp_{\mu^t}(\boldsymbol{\Sigma})$
6: **until**: $error < \epsilon$
7: **return** μ^{t+1}

3 CLUSTERING

Image classification is one of the most important tools for image interpretation and analysis. It consists of assigning the same label to a set of data that share common properties, with the goal of reducing the amount of information and simplifying the data interpretation. This topic, in regards to the mathematical and algorithmic aspect, has been explored since 1967, when Birkhoff (1967) introduced the modal distributive lattices notion, while, at almost same time, Sneath et al. (1973) investigated the mathematical models of hierarchical classifications. Since then, a large number of algorithms for classification have been developed.

The classification methods can be roughly categorized as supervised or unsupervised. Supervised classification requires prior information (training samples), while in unsupervised classification, the prior information is not available. Of the several unsupervised classification methods, perhaps the most common is based on clustering algorithms.

Clusters can be defined as high density regions separated by low density regions (JAIN, 2010). Therefore, clustering techniques aim to maximize the inner cluster homogeneity and the inter clusters heterogeneity, based on the data sets natural evidence of division.

Clustering the data set $X = \{\vec{x}_1, ..., \vec{x}_M\}$, which usually can be represented in a N-dimensional Euclidean space $\mathbb{R}^N$, is the processes of division or partitioning this data set into K groups based on a similarity metric, such that (PEDRINI; SCHWARTZ, 2008):

(a) $C_k \neq \varnothing$, $k \in \{1, ..., K\}$;

(b) $C_k \cap C_z = \varnothing$, $k, z \in \{1, ..., K\}$;

(c) $\bigcup_{k=1}^{K} C_k = X$;

where C_k is the group k.

The basic steps of cluster analysis are presented in Figure 3.1, and they can be summarized into four main steps: feature extraction, clustering algorithm, cluster validation, and interpretation of results. The goal of feature extraction is to find significant feature samples that can represent data set patterns. The clustering algorithm step refers to the choice of appropriated algorithm and the suitable similarity

metric selection. The results of clustering algorithm can be verified by using the proper evaluation tool, and in the interpretation step, experts of a given application area must label the clusters or data partition. These four steps help the user to find meaningful insights from the original data (XU; WUNSCH, 2005).

Figure 3.1 - Clustering steps.

SOURCE: Adapted from Xu e Wunsch (2005).

Clustering technique has a broader field of application, for instance, it is widely used for market analysis, helping on the costumer groups and purchasing patterns characterization (TSAI; CHIU, 2004). It is crucial on Earth observation area, helping on the land cover and land use identification (CHAKRABORTY, 2019; MÜLLER-HANSEN et al., 2017). It also helps recognition of fraudulent or criminal activity (CALVO et al., 2017), and on diagnostic of neurological diseases (ALASHWAL et al., 2019), among many others application.

3.1 User's dilemma

Clustering a data set, from which no information is known, could be a difficult task, even harder when considering noise and interference presence. Therefore, seeking out an appropriated clustering technique is fundamental to reach an optimum result. Despite the large number of clustering algorithms, there is no optimal method; each clustering technique has it owns advantages and drawbacks. A single algorithm cannot appropriately handle the different type of cluster shape, singularities, and nature of various data types. Although the clustering techniques are characterized

as unsupervised, the user must define the proper algorithm, metric and some other input parameters. Jain e Dubes (1988) highlighted the fundamental challenges on clustering technique choosing as the identification of:

(a) Data set type (numerical real, numerical complex, categorical);

(b) Data set normalization need;

(c) Similarity measure;

(d) Number of clusters;

(e) Cluster shape;

(f) Outliers, and how to deal with them;

(g) The initial centroid location choice;

(h) The clusters evaluation.

The data type definition and whether the data set needs normalization helps with the feature selection, which is an important preprocessing step used to enhance the underlying clustering quality. Data features are the redundant attributes which are likely to belong to same cluster. Popular feature selection techniques include information gain, chi squares tests, and Fisher score, among others (AGGARWAL, 2014).

The appropriate similarity measure definition is still a challenge on clustering technique. It is necessary to exercise caution when choosing the similarity measure. For example, an improper metric can make false good separability that does well in capturing the data variability, but actually makes no sense, because the data has been wrongly clustered. According to Hastie et al. (2009), an appropriate similarity measure is far more important in obtaining success with clustering than the choice of clustering algorithm itself.

The number of cluster definition is one of most difficult tasks. For instance, if the number of clusters are bigger than needed the cluster quality may improve, however it could be useless for data comprehension and decision makers. Usually this parameter is determined by the user, based on prior information. However, there are algorithms used to automatically determine the number of clusters. A simple

and famous solution consists of analysing the dendrogram provided by hierarchical clustering, and additional examples includes the elbow and silhouette methods (YELLOWBRICK, 2019). We can also apply some statistical analyses to derive the number of cluster, such as gap statistic (TIBSHIRANI et al., 2001).

Clustering techniques may have restriction on the shape of clusters, for instance, algorithms that uses Euclidean distance just accept convex and hyperspherical shaped cluster. Algorithms that use Mahalanobis distance deal better with data set which tends to have a hyperellipsoidal shape. Usually, partitioning and model-based algorithms are limited to shapes that consider the data set as represented in a flat manifold, i. e., a manifold that locally looks like an Euclidean space. Some hierarchical (depending on method and dissimilarity metric), and density-based algorithms can be applied as non-flat geometry clustering, i. e., they can be used in situations where the clusters have a shape represented in a non-flat manifold, consequently the standard Euclidean distance is not the proper metric.

The proper statistical model definition helps to find out the proper cluster shape, the appropriated similarity measure, and to define the most suitable algorithm. It also can be used to derive information about outliers, helping on its identification and filtering.

3.2 Similarity metrics

The main goal of clustering is to divide the data set into subsets, based on the similarity between samples, regardless the type of clustering technique. Distances are the most predominately metric to quantify the similarity between two samples.

Let $X = \{\vec{x}_1, ..., \vec{x}_M\}$ be the analysed data, and the distance between two samples of X is given by $d(\vec{x_i}, \vec{x_j})$. In order to be considered valid, the distance shall satisfy the following properties (DEZA; DEZA, 2009):

(a) $d(\vec{x}_i, \vec{x}_j) \geq 0, \forall x_i, x_j \in X$;

(b) $d(\vec{x}_i, \vec{x}_j) = 0 \Rightarrow \vec{x}_i = \vec{x}_j$;

(c) $d(\vec{x}_i, \vec{x}_j) = d(\vec{x}_j, \vec{x}_i)$;

(d) $d(\vec{x}_i, \vec{x}_j) \leq d(\vec{x}_i, \vec{x}_z) + d(\vec{x}_z, \vec{x}_j)$;

There are a number of distance metrics, among them the Minkowski distance is a metric in normed vector space, i. e., a vector space over the real or complex numbers,

on which a norm is defined. The Minkowski distance between two samples $\vec{x}_i$ and $\vec{x}_j$ defined as:

$$d(\vec{x}_i, \vec{x}_j) = \left(\sum_{i=1}^{n} |\vec{x}_i - \vec{x}_j|^p \right)^{1/p} ; \forall p \geq 1, p \in \mathbb{Z}^+ \tag{3.1}$$

where $\mathbb{Z}^+$ is the set of positive integer numbers.

If $p = 1$, the Minkowski distance receives the name of Manhattan distance. This distance is measured along axes at right angles. In short, the Manhattan distance generally works better if the samples are arranged in the form of a grid and the problem prioritize the distance between the points only along with the grids, but not the geometric distance.

If $p = 2$, then the Minkowski distance is named Euclidean distance. The Euclidean distance gives the shortest, or minimum, distance between two points. In other words, the Euclidean distance is the square root of the sum of squared differences between $\vec{x_i}$ and $\vec{x_j}$. This distance is more suitable for data sample that could be grouped in a hyperspherical shaped cluster, and thus make it only appropriate for data measured on the same scale. Therefore, when Euclidean distance is used as a similarity metric, highly correlated samples should be filtered, otherwise it may lead to the distortion on the final classification.

For data sets represented in a multivariate space, and that could be grouped in a cluster with elongated shape, a better approach would be to apply the Mahalanobis distance. Note that for uncorrelated samples, the Mahalanobis distance is equal to the Euclidean distance. Let $\vec{x}_i = (x_1, ..., x_m)$ be an observation of the analysed data X, with mean $\vec{\mu} = (\mu_1, ..., \mu_m)$, and covariance matrix Σ, the Mahalanobis distance is defined as:

$$d(\vec{x}_i) = \sqrt{(\vec{x}_i, \vec{\mu})^T \, \Sigma^{-1} \, (\vec{x}_i, \vec{\mu})} \tag{3.2}$$

Another well known clustering metric in statistic field is the Bhattacharyya distance, which measures the similarity of two probability distributions, and is defined as:

$$d(p, q) = -\log(\sum_{\mathcal{A}} \sqrt{p(x)q(x)}) \tag{3.3}$$

where p and q are probability distributions in the same sample space $\mathcal{A}$.

Each distance has its own peculiarities, advantages, and disadvantages and the choice of similarity metric is a critical step in clustering since it influences the shape of the clusters. A good distance metric helps in improving the clustering and classification result performance.

3.3 Clustering taxonomy

The clustering algorithms can be categorized into four classes: partitioning based, hierarchical, density-based, and model-based clustering algorithms.

There are a large number of clustering techniques in literature, and they can be broadly categorized into four classes: partitional, hierarchical, density-based, and model-based, as shown in Figure 3.2. The classes and example of the algorithms are discussed in the following subsections.

Figure 3.2 - Clustering Taxonomy.

SOURCE: Adapted from Fahad et al. (2014).

3.3.1 Partitioning based clustering algorithms

Partitioning algorithms divide the data set into K partitions, where each partition represents a cluster. In this case, each cluster is represented by its center or cluster centroid (BANDYOPADHYAY; SAHA, 2012). The clusters in partitioning algorithms shall fulfil the following requirements: a) each cluster must have at least one sample; and b) each sample must belong to exactly one group.

One of most popular partitioning algorithm is the K-means algorithm, which is an interactive clustering technique that divides the data set X into K clusters based on the minimal Euclidean distance between a sample $\vec{x_i}$ and a centroid c_k. The K-means

association equation is represented as:

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} u_{i,k} d(\vec{x_i}, c_k) \tag{3.4}$$

where $u_{i,k} = 1$ if x_i belongs to cluster c_k, or $u_{i,k} = 0$ otherwise, and $d(.)$ represents the distance metric.

In K-means algorithm first iteration, the centroids are randomly chosen, then the initial partitions are formed using the distance criteria (Equation (3.4)). Subsequently, the cluster centroids are recomputed by the rule for updating given by:

$$c_k = \frac{\sum_{i=1}^{N} u_{i,k} \vec{x}_i}{\sum_{i=1}^{N} u_{i,k}} \tag{3.5}$$

The algorithm terminates when a number of iterations is achieved or when a defined error is lower than a threshold. The K-means algorithm is very easy to implement, and in many cases gives satisfactory results. However, it has a number of limitations, for instance, as shown in Carvalho et al. (2019), this algorithms is affected by the choice of initial centroid and may converge to values that are not optimum. Moreover, once this algorithm uses the Euclidean distance, it assumes that the clusters have a hyperspherical shape, being susceptible to outliers.

The K-means algorithm is computationally difficult and the optimal solution is NP-hard but under specific heuristic, the problem can be translated into a NP-problem and the algorithm can converge quickly to a local optimum. There are a variety of heuristic algorithms to find the optimal solution to the K-means clustering, for instance the Lloyd's algorithm (BOCK, 2008) assumes that the number of samples N, the data dimension D, the number of cluster K, and the number of interations I are known, therefore the complexity is $O(NDKI)$.

3.3.2 Hierarchical clustering algorithms

The hierarchical algorithms aim to produce a sequence of clusters that are represented graphically by a dendrogram or valued tree. The dendrogram is a n-tree, where the individual elements are named leaves, the joins of two branches are named nodes, the parent node of all the others is named root, and the height of dendrogram is named levels, which are determined by the similarity between the samples (GAN et al., 2007), as showed in Figure 3.3. Each dendrogram level (L_i, $i \in \{0, ..., Q\}$) is formed with a junction (or bipartition) of groups at lower levels. An important

feature of these algorithms is that it is not necessary to determine a specific number of clusters, but this value is determined by the dendrogram level.

Figure 3.3 - Hierarchical clustering dendrogram.

SOURCE: Author's production.

Hierarchical clustering methods can be agglomerative (bottom-up) or divisive (top-down) (BANDYOPADHYAY; SAHA, 2012). An agglomerative clustering starts with one object for each cluster and recursively merges two or more of clusters, therefore, the decisions are made based on local standards. A divisive clustering starts with the dataset as one cluster and recursively splits into a number smaller groups, consequently making global decisions.

Hierarchical clustering outputs are more informative than a set of flat clusters, which makes the analysis easier, especially when the number of classes is unknown. However, the major drawback of these methods is that once a step (merge or split) is performed, this cannot be undone, therefore, the hierarchical heuristic can contribute to the spread of errors. The algorithms DIANA, BIRCH, CURE, ROCK and Bisecting K-means are some of the well-known algorithms of this category.

The Bisecting K-means, proposed by Karypis et al. (2000), is a hierarchical variant of K-means algorithm. The idea is to split the dataset into two clusters in each iteration, using the K-means algorithm until K clusters are obtained. As a result, a binary tree is created and each tree node has two children that correspond to two sub-clusters, as shown in Figure 3.4.

Figure 3.4 - Tree generated by Bisecting K-Means.

SOURCE: Author's production.

3.3.3 Density-Based clustering algorithms

Density-based clustering techniques visualize clusters as areas of high density separated by areas of low density. These algorithms are closely related to the k Nearest Neighbors (kNN) algorithm, since it groups the samples in the a given point vicinity.

The most known density based algorithm is the Density-based spatial clustering of applications with noise (DBSCAN), proposed by Ester et al. (1996). This algorithm has two parameters: the minimum number of samples and the minimum distance. The data samples are classified as core points, (density-)reachable points and outliers. The core point is randomly picked from the dataset, then the DBSCAN forms an m dimensional shape around it and checks the points that fall within the shape based on a distance measurement, usually Euclidean distance. These are the reachable points and the points that are out of the shape are the outliers. Then DBSCAN counts how many samples are classified as reachable points, and if a minimum number of points is achieved, the cluster is expended interactively. A new cluster can be initialized by picking the core point in the group of samples classified as outlier. The process repeats until every points can be allocated in one cluster. The time complexity of DBSCAN is $O(n^2)$.

3.3.4 Model-Based clustering algorithms

Model-based clustering algorithms optimize the fit between the given data and some predefined mathematical model (FAHAD et al., 2014). Usually these methods assume that the samples are represented in statistical terms, such as mean, variance, and

covariance. These parameters are used to estimate degrees of statistical separability between classes. One of the biggest challenges of these algorithms is to determine the method that, given a probabilistic distribution of the data, allows an optimal degree of separation between the classes (JENSEN, 2005). A well-known algorithm is the Expectation Maximization (EM), which uses a mixture density model to perform the clustering.

The EM algorithm is an iterative procedure that uses the maximum a posteriori rule to compute the maximum likelihood of a mixture model distribution in the latent variables presence. Let X be the set of observable data, Y be the latent variables, $\boldsymbol{\theta}$ be vector of unknown parameter, and $L(\boldsymbol{\theta}; X, Y)$ be the likelihood function. The maximum likelihood of $\boldsymbol{\theta}$ can be determined by maximizing the marginal likelihood of X:

$$L(\boldsymbol{\theta}; X, Y) = \int f(X, Y|\boldsymbol{\theta})dY \tag{3.6}$$

The EM algorithm seeks to find the maximum likelihood by applying two steps:

a) E-step: estimate the values for the latent variables as the log likelihood expectation:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^t) = E_{Y|X,\boldsymbol{\theta}^t}[logL(\boldsymbol{\theta}; X, Y)] \tag{3.7}$$

b) M-step: do the model optimizing by maximizing $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^t)$:

$$\boldsymbol{\theta}^{t+1} : \frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^t)}{\partial \Psi} = 0 \tag{3.8}$$

3.4 Cluster validation

Cluster validation refers to the procedures used to evaluate the results of clustering in a quantitative way in order to find the clusters with a better fit in a given data set. According to Theodoridis e Koutroumbas (2010), the clustering validation statistics methods can be categorized into three classes: internal cluster validation, external cluster validation, and relative cluster validation.

3.4.1 Internal cluster validation

Internal cluster validation uses the internal information of the clustering process to quantify the clustering accuracy or quality Q. Usually, the internal validation aims

to evaluate two criteria (BERRY; LINOFF, 2004): compactness, and separation.

The compactness is an inner cluster metric and it measures how close the samples of cluster are from each other. The compactness assesses cluster homogeneity. Therefore, a common measure of compactness is the variance; the lower the variance the better the compactness. On the other hand, the separation is an inter cluster metric which evaluates how far the clusters are from each other. Compactness and separation are opposing trends, and while the compactness increases with the number of clusters, the separation decreases.

Given a data set $X = \{\vec{x}_1, ..., \vec{x}_M\}$ with dimension P, which is clustered into K groups $C = C_1, ...C_k$, having centroids $c = c_1, ..., c_k$, the simplest internal validation measure is the Root-Mean-Square Standard Deviation (RMSSTD), which is designed to evaluate either the compactness or separation:

$$Q_{RMSSTD} = \sqrt{\frac{\sum_k \sum_{x \in C_k} \|x - c_k|^2}{P \sum_k (n_k - 1)}} \tag{3.9}$$

where n_k is the number of data points in cluster C_k.

The Dunn Index (DUNN, 1973) and Silhouette Width (ROUSSEEUW, 1987) are examples of non-linear combinations of compactness and separation. The Dunn index is defined as:

$$Q_d = min\Big(min\Big(\frac{d(c_k, c_p)}{max(d(X_k))}\Big)\Big) \tag{3.10}$$

where $d(c_k, c_p)$ defines the intercluster distance, i. e. the distance between centroinds; $d(X_k)$ is the inner cluster distance, or the distance between the samples that belongs to a given cluster.

If the data set contains compact and well-separated clusters, the diameter of the clusters is expected to be small, and the distance between the clusters is expected to be large. Thus, the Dunn index should be maximized. The drawback of the Dunn index is the computational effort, which can be computationally expansive as the number of clusters and data points increase.

The silhouette analysis estimates the average distance between clusters and each sample. It is defined as:

$$Q_s = \frac{1}{K} \sum_k \Big(\frac{1}{n_k} \sum_{x \in C_k} \Big(\frac{b(x) - a(x)}{max(b(x), a(x))}\Big)\Big) \tag{3.11}$$

where $a(x) = \frac{1}{n_k - 1} \sum_{y \in C_i, y \neq x} d(x, y)$ and $b(x) = min_{p \neq k} \left(\frac{1}{n_p} \sum_{y \in C_p} d(x, y) \right)$.

The Silhouette value measures the degree of confidence in the clustering assignment of a particular sample. If a sample has Q_s close to one, it means that the sample was correctly clustered. Small values of Q_s, i. e., around 0, means that the sample may belong to a different cluster or that it can be be clustered into two or more clusters. Samples with a negative Q_s are likely placed in the wrong cluster.

3.4.2 External cluster validation

External cluster validation consists of comparing the results of a cluster analysis to an externally known result. In this kind of evaluation, the clustering labels must be already known, therefore, henceforth clusters will be referred as classes.

The confusion matrix $\boldsymbol{M}$, or error matrix, is a well-known means of reporting the clustering results by comparing it with the ground truth. The confusion matrix $\boldsymbol{M}$ is a $N \times N$ matrix, with the elements e_{ij}, where the rows refer to cluster outcomes results and columns refer to the ground truth. The $\boldsymbol{M}$ main diagonal represents the number of samples correctly classified, i.e., the outcome label is equal to the true label; while off-diagonal, i. e., elements e_{ij} with $i \neq j$, represent the incorrectly clustered samples. For instance, considering a binary classifier where the samples can be classified between 'positive' and 'negative' classes, the confusion matrix can be represented as shown in Figure 3.5.

Figure 3.5 - Confusion Matrix.

		Ground Truth		
		Positive	Negative	
Clustering Outcome	Positive	**True Positive (TP)**	**False Positive (FP)**	Precision $TP/(TP + FP)$
	Negative	**False Negative (FN)**	**True Negative (TN)**	
		Recall $TP/(TP + FN)$		Accuracy $Acc = \frac{Tr(\boldsymbol{M})}{\sum_i \sum_j e_{ij}}$

SOURCE: Author's production.

From the confusion matrix $\boldsymbol{M}$, a number of validation metrics can be derived, including the overall accuracy, the cluster precision, the cluster recall and the F1-measure. The accuracy (Acc) is computed as the number of all correct predictions divided by the total number of the dataset. The best accuracy result is equal to 1.0, whereas the worst is 0.0. The accuracy is defined as:

$$acc = \frac{Tr(\boldsymbol{M})}{\sum_{i=1} \sum_{j=1} e_{ij}} \tag{3.12}$$

where $Tr(.)$ is the matrix trace.

The recall is computed for each class, therefore it is the samples correctly identified as belonging to a given class proportion. The best recall is equal to 1.0, whereas the worst is 0.0. Considering we are analysing the class C_i, the recall is defined as:

$$r = \frac{e_{ii}}{\sum_{j=1} e_{ij}} \tag{3.13}$$

The precision is also computed for each class, and it refers to the proportion of samples correctly identified as belonging to class C_i, among all samples of which the classifier claims that they belong to class C_i. The best precision is equal to 1.0, whereas the worst is 0.0. Considering the analysis of class C_i, the precision is defined as:

$$p = \frac{e_{ii}}{\sum_{i=1} e_{ij}} \tag{3.14}$$

and the F1-Measure provides a single score that balances both the concerns of precision and recall in one number. The F-measure is defined as:

$$F1 = \frac{2 \times p \times r}{r + p} \tag{3.15}$$

3.4.3 Relative cluster validation

Relative cluster validation evaluates the clustering by varying different parameter values for the same algorithm, for instance, varying the number of clusters. Often an external or internal validation index is used for this function. The RMSSTD can be applied to evaluate the clusters, or if the ground truth is known, the results can be represented in a confusion matrix form. A good example of application for this method is the hierarchical algorithm, where the relative improvement between two consecutive partitions can be assessed by applying the information gain between

the partitions. The information gain is the expected reduction in entropy caused by the partitioning of a subset. Given a parent cluster C and a cluster child C_i, the information gain is defined as:

$$IG = H(S) - \sum_{i=1}^{n} \left(\frac{|C_i|}{|C|}\right) \times H(C_i) \tag{3.16}$$

where $H(.)$ is the entropy of a data set, and $|.|$ is the number of samples in a given data set.

4 NEW PROPOSED CLASSIFIER

4.1 Contextualization

Hierarchical clustering algorithms can be used to overcome the limitations of partitional clustering. Usually partitional clustering requires a predefined number of cluster and are strongly dependent on the initial centers. While hierarchical algorithms do not require input parameters, they are more flexible and easy to interpret. However, in practice, hierarchical algorithms are overtaken by partitional algorithms for images clustering applications, because hierarchical algorithms do not work well with missing data, or with large data sets.

Another latent difference between partitioning and hierarchical algorithms refers to scalability and flexibility. An algorithm is considered scalable if the computational cost roughly increases with the number of samples (ZAÏANE et al., 2002), and the clustering flexibility refers to the ability to discover clusters with different shapes or to work with different data type.

Partitional algorithms, as the K-means, can have complexity $O(NDKI)$, as presented in Section 3.3.1. If the K-means algorithm converges in a few number of iterations I, and the number of cluster K is small, then the time performance grows almost linearly as the amount of data increases, therefore it is considered scalable. On the other hand, hierarchical agglomerative clustering usually have complexity $O(N^3)$, while hierarchical divisive clustering with an exhaustive search, as the DIANA (STRUYF et al., 1997), can be $O(2^N)$, which make these algorithms unfeasible for large datasets. Additionally, large datasets, especially with agglomerative clustering, makes the tree too large and the nodes and leafs can overlap each other, which then makes the dendrogram difficult to visualize and analyze.

Although, in special situations, the K-means can be scalable, it cannot be used for flexible data. Since K-means uses the Euclidean distance, it is applicable only to numerical data that fit into the Euclidean space, and assumes that the shape of a given cluster is spherical. Additionally, K-means needs linear separability among clusters to correctly distinguish them. An algorithm using the Mahalanobis distance is more flexible because it has the data covariance information, which allows the user to detect spherical or ellipsoid cluster shapes. However, the Mahalanobis distance assumes that the standard deviations of the two classes are the same. Therefore, an even flexible algorithm would use the Bhattacharyya distance, which is the general case of Mahalanobis distance. Likewise, hierarchical algorithms can be more flexible

than the K-means, and can handle different types of data.

This work uses the partitional clustering scalability and the hierarchical clustering flexibility, allied to the statistical knowledge about the data, in order to provide an accurate classification result. Therefore, the algorithm proposed here is based on the Bisecting K-means, a hybrid algorithm of the divisive hierarchical clustering algorithm and K-means clustering. As discussed in Section 3.3.2, the Bisecting K-means algorithm splits one cluster into two sub clusters by using the K-means strategy, at each dendrogram level, until K clusters are obtained.

Since K-means is biased towards hyperspherical clusters, and handles the data set in Euclidean Space, the use of stochastic distances is proposed (Section 2.5) as a similarity metric for K-means strategy, in order to make the algorithm more flexible. This algorithm is named SC. The divisive hierarchical proposed algorithm is named BSC.

The focus of this work is on PolSAR images, therefore, the dataset is formed by a group of covariance matrices $Z = \{\boldsymbol{Z_1}, \ldots, \boldsymbol{Z_n}\}$, with N samples and L number of looks, where $\boldsymbol{Z_n}$ is given by Equation (2.21). The dataset $\boldsymbol{Z}$ follows the Wishart distribution, therefore the similarity metrics used by the BSC are the stochastic distances defined in Section 2.5.

An important feature of K-means, or SC (Stochastic Clustering), is its greedy behavior (JAIN, 2010), meaning that this algorithm makes locally optimal choices at each stage, with the goal of finding the global optimal solution to the entire partitioning problem. The drawback of this feature is the local minimum convergence; the algorithm is able to converge to the global optimum only when clusters are well separated. Additionally, as demonstrated in Carvalho et al. (2019), the K-means and SC can, frequently, be trapped in a local minimum, which results in wrong classification. One of the main reasons is the choice of initial centroids. Bad initial centroids are likely to result in non-optimized and wrong clustering estimations. Therefore, choosing good candidates for initial centroids implies the need for more accurate results and faster convergence, i. e., it reduces the amount of time spent in areas away from the global maximum. Therefore, in our approach, the two initial centroids can be determined by two different strategies: using the RPDDP algorithm or by the EM algorithm.

The new proposed classifier based on BSC algorithm (BSC Classifier) is divided into 7 modules, which are shown in Figure 4.1. The first step in BSC classifier is to

select the input PolSAR images. After that, the stochastic distances are presented, the user can select the suitable distance. The next steps are gathered as the BSC algorithm, and they are ordered from one to three. The BSC algorithm step number 1 consists in determine the suitable cluster to split, the step number 2 seeks the initial centroids or parameter determination, in here two options are presented, and the user shall select the strategy (RPDDP OR EM). The step number 3 is the dataset splitting. Once the BSC algorithm converges, the cluster labeling shall be done by the user. And, finally, the image classification is validated.

Figure 4.1 - BSC classifier modules.

SOURCE: Author's production.

The SC algorithm has time complexity $O(KNi)$, where K is the number of clusters, N is the number of samples and i number of iterations. Considering that the BSC divides the data into two clusters in each interaction, i. e., $K = 2$. The BSC does one bipartition per time therefore, considering C BSC iterations, this algorithm would have complexity $O(2CNi)$.

4.2 Bisecting stochastic clustering algorithm

The BSC algorithm, presented in Figure 4.2, is a combination between the SC algorithm, as described in Carvalho et al. (2019), and the hierarchical divisive algorithm. This algorithm aims to split one cluster into two sub clusters at each bisecting step by using the SC strategy. The BSC starts with a single cluster and the splitting

procedure is done until stop criteria is reached.

Figure 4.2 - BSC flowchart.

SOURCE: Author's production.

The BSC outputs are the clustered image and the dendrogram. There are three major steps in the bisection procedure: 1) the initial parameters determination; 2) the splitting procedure; 3) the choice of a suitable cluster to split. These steps will be further discussed in the following sections.

4.2.1 Initial parameters determination

The correctness of initial parameters choice makes the greedy algorithms convergence faster and the classification result more accurate. There is a large number of methods for initial parameters determination, including the Ball and Hall's, Simple Cluster Seeking, Maximin (CELEBI et al., 2013), and methods based on Principal Component

Analysis (PCA).

The BSC uses the stochastic distances as similarity metric. Due to that, the initial parameters are the expected covariance matrices $\boldsymbol{\Sigma_k}$, where $k = 1, ..., K$, with $K = 2$. Since PolSAR images are widely known for following the Wishart distribution, the initial parameter determination can be done by the hidden probabilistic parameter estimation. The EM algorithm is an iterative method to estimate of parameters in statistical models, therefore, the first way to determine the initial parameter is done by the EM for Wishart mixture model distribution.

The second way to determine the initial parameter is done by a new proposed approach named Riemannian Principal Direction Divisive Partitioning (RPDDP) which is based on the Principal Direction Divisive Partitioning (PDDP) algorithm. The original PDDP algorithm was proposed by Boley (1998) with the goal of bi-partitioning data samples based on an embedding in a high dimensional Euclidean space. However, since the PolSAR images are represented by the covariance matrices $\mathbf{Z}$, the original version of PDDP is not applicable, as long as these matrices do not form a Euclidean space (YING-HUA; CHONG-ZHAO, 2010). Indeed, symmetric positive definite matrices are better manipulated with Riemannian geometry (CONGEDO; BARACHANT, 2015), for this reason the the PDDP is handled by the Riemannian geometry.

4.2.1.1 Riemannian Principal Direction Divisive Partitioning (RPDDP)

The PDDP algorithm is a fast and scalable hierarchical divisive clustering algorithm. The basic idea is to recursively split the dataset into sub-clusters based on the Principal Component Analysis (PCA) routing; the output of this algorithm is the clustered data and a binary tree.

In our approach, the input data set is the group of covariance matrices $\boldsymbol{Z}$, and the covariance matrix $\boldsymbol{\Sigma}$ is defined as the geometric mean of the dataset Z, given by Equation (2.41). The matrix $\boldsymbol{\Sigma}$ can be decomposed as $\boldsymbol{\Sigma} w_j = \lambda_j w_j$, where λ_j are the eigenvalues associated to eigenvectors w_j. The principal direction is the eigenvector with the largest eigenvalue with which it is associated.

In order to perform the samples projection onto the principal direction, only the main diagonal of the $\boldsymbol{Z_n}$ is considered, i. e., only the information related to image brightness is taken into account at this phase. The $diag(\boldsymbol{Z_n})$ projection is given by the Equation (4.1).

$$\sigma v_n = w_j^T(diag(\boldsymbol{Z_n}) - diag(\boldsymbol{\Sigma})) \tag{4.1}$$

The values v_n are separated into two groups accordingly to their sign, as well the correspondent covariance matrix $\boldsymbol{Z_n}$. The geometric mean of the covariance matrices $\boldsymbol{\Sigma_k}$ is computed to each group, and the initial parameters are determined, as presented in Figure 4.3.

Figure 4.3 - RPDDP flowchart.

SOURCE: Author's production.

4.2.1.2 Expectation-Maximization (EM)

The EM for Wishart mixture model procedure used in this work was described in Carvalho et al. (2019). Assuming $Z = \{\boldsymbol{Z_1}, \ldots, \boldsymbol{Z_n}\}$ as the set of observed complex covariance matrices, the Wishart mixture model can be expressed by (HIDOT; SAINT-JEAN, 2010):

$$f(\boldsymbol{Z_n}; \Psi) = \sum_{k=1}^{K} \pi_k f_k(\boldsymbol{Z_n}; \theta_k) \tag{4.2}$$

where K is the number of Wisharts within the mixture. In this approach, there are two Wisharts per time, or $K = 2$, and $\Psi = \{\pi_1, ..., \pi_K, \theta_1, ..., \theta_K\}$ is the un-

known mixture model parameter vector, containing the weighting factor per Wishart π_k, which must sum up to one, and the parameter vector θ_k. All functions $f_k(.)$ correspond to same density model, i. e., the Wishart distribution, for this reason $\theta_k = \{\boldsymbol{\Sigma_k}, L\}$, with the expected covariance matrix $\boldsymbol{\Sigma_1}$ and the number of Looks L. Also, the algorithm is executed over all $\boldsymbol{Z}$ samples, therefore the complete-data log likelihood is formulated as:

$$log\mathcal{L}(\Psi) = \sum_{n=1}^{N}\sum_{k=1}^{K} u_{n,k} \log\Big(\pi_k f_k(\boldsymbol{Z_n}; \boldsymbol{\Sigma_k}, L)\Big) \tag{4.3}$$

where $u_{n,k} = 1$ if the sample n produces a measurement k, or $u_{n,k} = 0$ otherwise.

The goal of EM algorithm is to find the maximum likelihood estimation through the iterative maximization of the conditional expectation. At every iteration the EM algorithm consists of two steps:

a) In the expectation step, the log-likelihood of the observed data $\boldsymbol{Z_n}$, given the estimated parameter Ψ^t, is calculated as:

$$Q(\Psi, \Psi^t) = \sum_{n=1}^{N}\sum_{k=1}^{K} u_{n,k}\Big[\log(\pi_k^t) - L\log|\boldsymbol{\Sigma_k^t}| + (L-q)\log|\boldsymbol{Z_n}| - Tr(\boldsymbol{\Sigma_k^t}\boldsymbol{Z_n})\Big] \tag{4.4}$$

b) In the maximization step, the new parameter Ψ_{t+1} is estimated. Since the parameter Ψ is composed of π_k and Σ_k, the parameter optimization is done by setting the respective partial derivative to zero, meaning that the function local maximum has been found:

$$\Psi^{t+1} : \frac{\partial Q(\Psi, \Psi^t)}{\partial \Psi} = 0 \tag{4.5}$$

The optimization, with respect to π_k^{t+1}, can be summarized as:

$$\pi_k^{t+1} = \frac{1}{N}\sum_{n=1}^{N} u_{n,k} \tag{4.6}$$

and the new estimation for Σ_k^{t+1} is given by the geometric mean presented in Equation (2.41). After EM convergence, the initial parameter is given by the estimated Σ.

4.2.2 Splitting procedure: Stochastic Clustering

The partition algorithm presented here is based on K-means procedure, and was described in Carvalho et al. (2019). The K-means algorithm divides the dataset into K clusters based on Euclidean distance, and each cluster can be summarized by its mean. However, as previously discussed, the inputs of BSC are a group of covariance matrices $Z = \{\boldsymbol{Z_1}, \ldots, \boldsymbol{Z_n}\}$, and this set does not form an Euclidean space, nor can the group be summarized as the arithmetic mean.

Since the similarity metric used to split the dataset Z is the stochastic, the splitting procedure is called Stochastic Clustering. In Section 2.5, it is stated that Nascimento (2012) derived five stochastic distances between Wishart distributions: the Bhattacharyya, Kullback-Leibler, Hellinger, Rényi of order β, and Chi-square. However, as testified by Carvalho et al. (2019), only Bhattacharyya, Kullback-Leibler, and Hellinger distances achieved successful results when used as similarity metrics on the SC algorithm. The Rényi of order β and Chi-square presented poor accuracy classification results, due to their numerical instabilities. Therefore, the last two distances are not indicated to perform PolSAR classification, and for that reason, they were excluded from the final version of BSC.

The SC procedure follows a simple way to cluster the dataset Z into two cluster at each bisecting iteration. The main idea is to define two initial parameter, the next step is to associate each sample from Z to the more likely parameter. Once no sample is pending, the parameters are recalculated and the samples are reassigned. The algorithm repeats until convergence. The association function is given by:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} u_{n,k} d(\boldsymbol{Z_n}, \boldsymbol{\Sigma_k}) \tag{4.7}$$

where $K = 2$, $u_{n,k} = 1$ if x_n belongs to cluster c_k, or $u_{n,k} = 0$ otherwise, and $d(.)$ represents the stochastic distance. The rule for updating the cluster parameter is given by the geometric mean, represented by Equation (2.41).

4.2.3 Determination of new nodes and clusters

The BSC is built as a top-down procedure: it starts with all samples in an unique cluster (named parent), that are split into two new subclusters (named children), which later will be assigned as parents and divided successively into new subclusters. Once the number of subclusters and, therefore, the number of nodes, is higher than

one, a rule to determine which cluster shall be split is needed.

The BSC uses the information gain as node decision rule. The information gain is based on the decrease in entropy after the dataset splitting, i. e., it represents the amount of entropy removed between clusters. Mathematically, it can be represented as:

$$I(X,Y) = E(X) - E(X|Y) \tag{4.8}$$

where $I(X,Y)$ is the entropy gain between cluster X (the parent cluster) and Y (the child cluster). The entropy of the parent cluster $E(X)$ is subtracted by the entropy of the subcluster $E(X|Y)$, resulting in a decrease in entropy.

The entropy is a fundamental concept to compute the cluster homogeneity. There are several families of entropy; this work presents the $(h-\phi)$-family, proposed by Salicru et al. (1993), to derive the entropy for positive definite Hermitian matrices. The $\boldsymbol{Z}$ follows the Wishart distribution $f_Z(\boldsymbol{Z}, \boldsymbol{\Sigma}, L)$. The Shannon Entropy obtained by using the $(h-\phi)$-entropy class is defined as (FRERY et al., 2012):

$$\mathrm{H}_S(\boldsymbol{\theta}) = \tfrac{q(q-1)}{2}\ln(\pi) - q^2\ln(L) + q\ln|\Sigma| + qL + (q-L)\psi_q^{(0)}(L) + \textstyle\sum_{k=0}^{q-1}\ln\Gamma(L-k) \tag{4.9}$$

where q is the order of $\boldsymbol{\Sigma}$. The clustering procedure shall be done in one set of PolSAR image per time, therefore the number of looks L is assumed to be equal, as well the covariance matrices order q. Then, the number of looks L and the order q are considered constant, the entropy can be summarized as a function of Σ, and can be written as $H(\Sigma_i)$, where i is the cluster index.

The information gain is computed to all nodes in a given level; these nodes are named candidates. For each cluster associated to a candidate node, the initial parameter $\boldsymbol{\Sigma}$ is determined (by RPDDP or EM precedure). The estimated initial parameter is used to compute the entropy gain of future children, as shown in Figure 4.4. The candidate node with the highest information gain is the winner and will be split in the next bisecting step.

Figure 4.4 - Information Gain computation.

SOURCE: Author's production.

4.2.4 BSC Dendrogram

The dendrogram is a binary tree structure used to represent the hierarchical relationship between the subclusters. It can be defined as a nodes collection, starting with the root node and ending with leaf nodes (see Figure 3.3). Each node is a data structure consisting of: node ID, parent ID, cluster entropy, information gain, and the splitting initial parameters Σ_i, as show in Figure 4.5. Associated to the dendrogram is the ID image, which has the same number of rows and columns of the input image, and the pixel position (x, y) has as value the node ID. In the end, the image pixels will be classified according to ID number with which it is associated.

The dendrogram can show information about the splitting history, but it can also tell about the subclusters or classes that are more similar. For instance, looking at the leaves of a dendrogram, the classes that are closer, are similar (it can be, for example, the same type of plantation, but in different periods), and they can even be clustered again after analysis. While the classes that a far away from each other are unrelated (for example, forest and urban area).

Figure 4.5 - Dendrogram Structure and ID image.

SOURCE: Author's production.

5 METHODOLOGY AND DATA DESCRIPTION

In this chapter the work methodology used to obtain the results described in Chapter 6 is presented. Simulated and real PolSAR images are used in order to evaluate the proposed algorithm.

As presented in Figure 4.1, the last module in the BSC classifier is the classification validation. In this work the classification results are evaluated by a quantitative and qualitative analysis. Therefore, in this chapter the dataset and the methods used to analyse these outputs are presented. The quantitative analysis evaluates the confusion matrices and accuracy. While, the qualitative analysis assay the cluster information gain, the dendrogram, and the scattering mechanism.

5.1 Data set description

This section describes the PolSAR images analysed by the proposed algorithms. The first set of image, composed by four real PolSAR images, is described in Section 5.1.1. Section 5.1.2 describes the simulated images and the simulation procedure.

5.1.1 Real PolSAR images

In order to bring variety to the proposed algorithms validation, four PolSAR images were analysed. The images are from three different scenarios: the Brazilian Caatinga, the Amazon forest and urban area. Hence, our ambition is to analyse the BSC algorithm with various scattering mechanisms.

Two images are provided by the Spaceborne Imaging Radar-C/X-SAR (SIR-C/X-band), which is a joint project between National Aeronautics and Space Administration (NASA), the German Space Agency (DARA) and the Italian Space Agency (ASI). The SIR-C/X-SAR system acquires images at three wavelengths: L band, C band, and X band. The L and C band can be tuned to select among several polarization options, and the SIR-C is capable of recording full quad-polarized images (FREEMAN et al., 2019). The images used in this work are in L and C bands, and full quad-polarized.

The other two images are generated by the PALSAR (Phased Array type L-band Synthetic Aperture Radar), which is an active microwave and L band frequency sensor. The PALSAR is one of three instruments on the Advanced Land Observing Satellite (ALOS), which is a the Japan Aerospace Exploration Agency (JAXA) mission.

5.1.1.1 SIR-C images

The SIR-C instrument includes the L band, which has frequency of 1.25 GHz and 23 cm of wavelength, and the C band, which has 5.304 GHz and 6 cm of wavelength. Both frequencies have four polarizations (hh, hv, vh, and vv). The main information about them are described in Table 5.1.

Table 5.1 - SIR-C images parameters.

Year of acquisition	1994
Image Size	407 x 370 pixels
Frequency	L band (1.25GHz) / C band (5.304GHz)
Polarization	hh, hv, vv
Product Type	Complex Multilook
Number of nominal looks	4.79

SOURCE: Correia (1998).

Figures 5.1(a) and 5.1(b) present the color composition of SIR-C images in L and C bands, respectively. The scene refers to an agricultural region in Bebedouro, located in the São Francisco river sub-middle region, about 40 km northeast of Petrolina, Pernambuco, Brazil. The Bebedouro region is characterized by a vegetation named Caatinga, which is a common type of desert vegetation in the interior of Brazilian northeastern.

The Bebedouro area, according to information derived from a study field, has six main classes: caatinga, tillage, bare soil, soy, corn and river. However there are regions with different ages for soy and corn, resulting in nine different classes, as shown in Figure 5.1(c): Corn 1, Corn 2, Soy 1, Soy 2, soy3, Tillage, Bare Soil, Caatinga and River (CORREIA, 1998).

The input image of every algorithm analysed in this work is the PolSAR image covariance matrix representation, therefore, every image pixel is a covariance matrix, and the dataset is represented as a group of covariance matrices $Z = \{\boldsymbol{Z_1}, \ldots, \boldsymbol{Z_n}\}$ with N samples and L number of looks. Samples from the regions marked as classes in Figure 5.1(c) were collected, and, from these samples, nine covariance matrices were estimated. The covariance estimation was performed in two ways: 1. using the arithmetic mean, and this kind of estimated covariance matrix will be named

AvCM (average covariance matrix); 2. using the intrinsic mean described in Section 2.6, this kind of estimated covariance matrix is named, henceforth, InCM (intrinsic covariance matrix).

Figure 5.1 - SIR-C image of Bebedouro, with the components hh (R), hv (G) and vv (B), in C and L band.

(a) L band PolSAR image.

(b) C band PolSAR image.

(c) Classes.

SOURCE: Correia (1998).

Since the Bebedouro images have nine classes, nine AvCM and nine InCM were

estimated. The determinant and trace of these matrices, from the L band image, are presented in Table 5.2. As described in Section 2.6, the determinant values computed from AvCM tend to suffer from swelling effect, therefore these determinant values are bigger than the InCM ones. In the same way, the AvCM traces values a higher than the InCM traces values.

Table 5.2 - Bebedouro SIRC-C L band determinant and trace of the estimated covariance matrices.

Class	**AvCM Determinant**	**AvCM Trace**	**InCM Determinant**	**InCM Trace**
Corn 1	1.05×10^{-4}	0.24	6.06×10^{-5}	0.21
Corn 2	2.58×10^{-5}	0.13	1.62×10^{-5}	0.11
Soy 1	1.82×10^{-5}	0.12	1.01×10^{-5}	0.09
Soy 2	2.26×10^{-5}	0.11	1.21×10^{-5}	0.08
Soy 3	4.99×10^{-5}	0.14	2.80×10^{-5}	0.12
Tillage	1.54×10^{-6}	0.05	9.25×10^{-7}	0.04
Bare Soil	1.94×10^{-8}	0.02	1.24×10^{-8}	0.01
Caatinga	8.07×10^{-4}	0.31	5.14×10^{-4}	0.27
River	1.20×10^{-8}	0.02	7.98×10^{-9}	0.02

SOURCE: Author's production.

The PolSAR image covariance matrix determinant is related to the image brightness. Therefore, by analysing the values presented in Table 5.2 it is possible to have a key about what classes are similar to each other. For instance, the classes Corn 2 and Soy 2 have the closest determinant values, meaning that, visually, these classes are similar. The Caatinga has the highest determinant, consequently this is the brightest area in the PolSAR image, as can be seen in Figure 5.1(a). The class with the lowest determinant and less bright in the picture is the Bare Soil, followed by the River and Tillage.

In order to have a clearer idea about the classes relationship, the Hellinger stochastic distance between the estimated covariance matrices were computed and a distance table was derived. The distance values are represented by gray-scale in the distance table. Two distance tables were derived, one considering the AvCMs, shown in Figure 5.2(a), and the second using the InCMs, presented in Figure 5.2(b).

Figure 5.2 - Hellinger distances between all classes within SIR-C image in L band.

	Corn 1	Corn 2	Soy 1	Soy 2	Soy 3	Tillage	Bare soil	Caatinga	River
Corn 1	0.00	0.58	0.61	0.56	0.72	2.28	7.34	1.08	8.06
Corn 2		0.00	0.19	0.20	0.79	1.24	5.60	1.67	6.16
Soy 1			0.00	0.15	0.73	1.07	5.23	1.96	5.82
Soy 2				0.00	0.33	1.07	5.46	1.58	6.18
Soy 3					0.00	1.72	6.51	1.18	7.44
Tillage						0.00	2.38	4.38	3.21
Bare soil							0.00	10.10	0.57
Caatinga								0.00	10.83
River									0.00

(a) AvCM distance table.

	Corn 1	Corn 2	Soy 1	Soy 2	Soy 3	Tillage	Bare soil	Caatinga	River
Corn 1	0.00	0.58	0.64	0.58	0.72	2.24	7.22	1.14	7.90
Corn 2		0.00	0.20	0.20	0.81	1.28	5.58	1.69	6.10
Soy 1			0.00	0.16	0.82	1.06	5.07	2.10	5.60
Soy 2				0.00	0.39	1.01	5.27	1.73	5.93
Soy 3					0.00	1.64	6.37	1.28	7.27
Tillage						0.00	2.35	4.45	3.20
Bare soil							0.00	10.11	0.58
Caatinga								0.00	10.78
River									0.00

(b) InCM distance table.

SOURCE: Autor's production.

The distance table shades vary from white to dark gray, where a distance equals to zero is represented by the white color, and the highest distance is represented by the darkest gray shade. In both distance tables, the classes Soy 1 and Soy 2 are the closest, with 0.15 in AvCM table and 0.16 in InCM table. It is interesting to note that while these classes have the smallest stochastic distance between then, their determinants are not the most similar, Soy 2 determinant is more similar to Corn 2 determinant, as shown in Table 5.2. The second closest classes are Soy 1 and Corn 2, with 0.19 in AvCM table and 0.20 in InCM table; followed by Soy 2 and Corn 2, with 0.20 distance in both AvCM and InCM table. Therefore a confusion between

Soy 1, Soy 2 and Corn 2 is expected

The River and Bare Soil classes are similar, having distances 0.57 in AvCM table and 0.58 in InCM table, however these are the classes with the highest distances when compared with Corn 1, Corn 2, Soy 1, Soy 2, and Soy 3. The Caatinga class also has a great distance from the other classes, meaning that this classes should be well clustered. Comparing Figure 5.2(a) and Figure 5.2(b), the gray level does not change much between them, i. e., the classes with highest distances and lowest distance are the same in both distances tables. However, the closest classes have a slightly better/higher stochastic distance values in InCM distance table. For instance the classes Corn 1 and Corn 2 are distant 0.56 using AvCM and 0.58 using InCM.

Using the same procedure as for L band, Table 5.3 presents the determinant and traces of the C band PolSAR estimated covariance matrices. This table has the AvCM and InCM determinant and traces values, and, as seen in L band, the AvCM determinants and traces are bigger than the InCM ones.

Table 5.3 - Determinant and trace of the estimated covariance matrices from classes of C band image.

Class	**AvCM Determinant**	**AvCM Trace**	**InCM Determinant**	**InCM Trace**
Corn 1	4.53×10^{-4}	0.30	2.84×10^{-4}	0.27
Corn 2	3.79×10^{-4}	0.26	2.32×10^{-4}	0.22
Soy 1	7.90×10^{-4}	0.37	4.76×10^{-4}	0.33
Soy 2	3.89×10^{-4}	0.28	2.59×10^{-4}	0.25
Soy 3	3.16×10^{-4}	0.27	2.10×10^{-4}	0.24
Tillage	6.44×10^{-4}	0.33	3.94×10^{-4}	0.28
Bare Soil	7.82×10^{-07}	0.07	4.77×10^{-07}	0.06
Caatinga	1.38×10^{-3}	0.42	8.85×10^{-4}	0.37
River	5.12×10^{-8}	0.02	3.30×10^{-8}	0.02

SOURCE: Author's production.

In C band image, contrasting the L band PolSAR image, there are more confusion between the classes, i. e., the determinants values are similar, except for the River and Bare Soil. All classes have bigger determinant values in comparison with L band.

Therefore the C band image is brighter than the L band image, and the classes Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, Tillage, and Caatinga, visually, have low contrast, as a result it is almost not possible to discriminate between these classes. Consequently the C band image is expected to have three big classes: Vegetation (Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, Tillage, and Caatinga), River and Bare Soil.

The confusion between the vegetation classes can be explained by the nature of the C band. The L band has a much bigger wavelength (24 cm) than the C band (5.6 cm), what implies that L band wave interacts with canopy and vegetation branches, it can penetrate the vegetation, and it can interact with the ground. Therefore, the L band sensor can measure the vertical structure of vegetation, and provide information about the three majors scattering mechanism (surface, double-bounce and volumetric) in vegetation areas. The C band interacts much more with the vegetation canopy, leaves and branch, therefore, with this frequency is more difficult to have information about the vertical structure of the vegetation. This can explain why all Vegetation classes are very similar in C band image.

The similarity between the classes in C band image can be also checked in the distance table. As in L band, two distance tables, using the Hellinger stochastic distance, were derived. The first distance table uses the AvCM is presented in Figure 5.3(a), and the second, using InCM, is presented in Figure 5.3(b).

As above cited, it is possible to group the classes into three major classes: Vegetation, Bare Soil, and River. The distances between the Vegetation classes are smaller than 0.32, what is a low value in comparison with L band image distance tables. For instance the distance between Tillage and Caatinga is 4.45 in L band image (Figure 5.2(b)), while in C band the distance between these classes is only 0.28 (Figure 5.3(b)). Also, the distance between Tillage and Corn 1 is 2.24 in L band image (Figure 5.2(b)), and only 0.08 in C band (Figure 5.3(b)).

About the difference between the AvCM and InCM distance tables, as for L band, the gray level does not alter in both images. However, the InCM has a slightly better overall separation between the closest classes. For instance, the distance between Corn 2 and Soy 1 in AvCM distance table is 0.19, while in InCM distance table, the distance value is 0.21.

Figure 5.3 - Hellinger distances between all classes within SIR-C image in C band.

	Corn 1	Corn 2	Soy 1	Soy 2	Soy 3	Tillage	Bare soil	Caatinga	River
Corn 1	0.00	0.08	0.11	0.11	0.16	0.07	4.68	0.26	8.00
Corn 2		0.00	0.19	0.06	0.12	0.11	4.57	0.28	7.77
Soy 1			0.00	0.13	0.15	0.21	5.21	0.09	8.67
Soy 2				0.00	0.05	0.20	4.46	0.24	7.73
Soy 3					0.00	0.31	4.16	0.29	7.45
Tillage						0.00	5.22	0.25	8.53
Bare soil							0.00	5.91	1.29
Caatinga								0.00	9.44
River									0.00

(a) AvCM distance table.

	Corn 1	Corn 2	Soy 1	Soy 2	Soy 3	Tillage	Bare soil	Caatinga	River
Corn 1	0.00	0.10	0.14	0.13	0.18	0.08	4.74	0.28	7.98
Corn 2		0.00	0.21	0.07	0.15	0.12	4.62	0.31	7.72
Soy 1			0.00	0.14	0.14	0.24	5.21	0.11	8.58
Soy 2				0.00	0.06	0.20	4.58	0.22	7.77
Soy 3					0.00	0.31	4.27	0.29	7.49
Tillage						0.00	5.26	0.28	8.47
Bare soil							0.00	5.98	1.28
Caatinga								0.00	9.44
River									0.00

(b) InCM distance table.

SOURCE: Author's production.

5.1.1.2 ALOS/PALSAR images

The ALOS/PALSAR has three observation modes: the fine mode, the scanSAR mode, and the polarimetric mode. The fine resolution is the conventional mode, it offers a spatial resolution of 7 m. The scanSAR mode acquires a 250 to 350 km width SAR image, with a coarse resolution. The last mode is the polarimetric, which has a resolution of 30 m, but offers a complete polarization scheme (hh, hv, vh, and vv).

ALOS/PALSAR data are available in different levels of processing, according to range and azimuth compression to the acquired data. The ALOS/PALSAR images

level 1.0 are a slant range data as they are recorded and they require a processing program before utilizing the data. The images in level 1.1 are single-look slant-range imagery generated by applying range and azimuth compression to the level 1.0 data. The images in level 1.5 are amplitude data, after range and azimuth compression and multi-look process, and are also projected to ground range.

The ALOS/PALSAR images used in this work are level 1.1, in polarimetric mode. They are a dataset from the Tapajós National Forest and the well-known San Francisco polarimetric dataset. Table 5.4 presents the main information about both Tapajós and San Francisco images, as the year of acquisition, the image size, the sensor frequency, incidence angle, the product type and image number of looks.

Table 5.4 - ALOS/PALSAR images parameters.

Parameter	**Tapajós**	**San Francisco**
Year of acquisition	2007	2015
Image Size	401 x 576 pixels	374 x 626
Frequency	L band ($1.27GHz$)	L band ($1.27GHz$)
Polarization	hh, hv, vv	hh, hv, vv
Incidence angle	23.7^o	23.7^o
Product Type	Complex Single Look	Complex Single Look
Number of nominal looks	1	1

SOURCE: Author's Production.

The Tapajós National forest PolSAR image color composition image is presented in Figure 5.4(a). The Tapajós imaged area is located in Belterra, State of Pará, Brazil. This is considered an important conservation unit in the Brazilian Amazon Forest. The unit is bounded by the Tapajós River and the BR-163 Santarém–Cuiabá road. The forest coverage includes dense rainforest, alluvial rainforest, open tropical forest, and secondary forest along the borders and access routes.

The Tapajós PolSAR image has five identified classes: Primary Forest, Pasture, Bare Soil, and three types of agriculture (Agriculture 1, Agriculture 2 and Agriculture 3). The classes represented in Figure 5.4(b) were identified by a fieldwork campaign conducted by INPE.

Figure 5.4 - Tapajós National Forest ALOS/PALSAR image with the components hh (R), hv (G) and vv (B).

(a) PolSAR image.

(b) PolSAR image with classes.

SOURCE: Carvalho et al. (2019).

Table 5.5 - Determinant and trace of the average covariance matrices from Tapajós ALOS/PALSAR image classes.

Class	**AvCM Determinant**	**AvCM Trace**	**InCM Determinant**	**InCM Trace**
Bare Soil	3.07×10^{30}	5.39×10^{10}	1.64×10^{30}	4.50×10^{10}
Agriculture 1	2.82×10^{29}	2.64×10^{10}	1.48×10^{29}	2.19×10^{10}
Agriculture 2	1.66×10^{33}	1.26×10^{12}	6.30×10^{32}	9.62×10^{11}
Agriculture 3	5.08×10^{32}	8.73×10^{11}	2.12×10^{32}	7.22×10^{11}
Primary Forest	1.75×10^{32}	1.77×10^{11}	9.19×10^{31}	1.44×10^{11}
Pasture	5.54×10^{31}	1.46×10^{11}	2.81×10^{31}	1.23×10^{11}

SOURCE: Author's production.

As for Bebedouro SIR-C images, a number of covariance samples from the areas marked as classes in Figure 5.4(b) were selected, and the average covariance matrices were estimated as the AvCM and InCM, from where the determinant and matrix trace, shown in Table 5.5, were computed. The AvCM determinant and trace values are bigger than the InCM ones, however, for both AvCM and InCM, these values are bigger and diverse, what may indicate that the samples shall be well clustered.

The Hellinger stochastic distance between the six classes are presented in Figure 5.5(a) and 5.5(b). In these figures, the distance values are characterized by a gray-scale, where lower values are white and the highest values are dark gray. The distances between the same classes are, off course, zero; for example, the distance between Bare Soil against Bare Soil is zero, therefore those cells are white. On the other hand, the most distant classes are the Agriculture 1 and Agriculture 2, having distance 7.69 in AvCM distance table and 7.29 using InCM estimation, and these are the darkest cells.

In general, the AvCM and the InCM distance tables showed that the Tapajós classes have a good distance values between them, in comparison with the SIR-C L band lowest distances (Figures 5.2(a) and 5.2(b)). Both Figures 5.5(a) and 5.5(b) have similar gray-level, meaning that the classes with bigger distance values or lower distance values are the same in both distance tables. However the InCM overtakes the AvCM when the classes are closer. For instance the distances between Agriculture 3 and Agriculture 2 is 0.39 in AvCM distance table, and 0.42 in InCM distance table.

Figure 5.5 - Hellinger distances between all classes within Tapajós ALOS/PALSAR image.

	Bare soil	Agriculture 1	Agriculture 2	Agriculture 3	Primary forest	Pasture
Bare soil	0.00	0.80	5.09	3.81	2.05	1.06
Agriculture 1		0.00	7.69	6.23	4.53	3.20
Agriculture 2			0.00	0.39	3.30	2.98
Agriculture 3				0.00	2.69	2.13
Primary forest					0.00	0.34
Pasture						0.00

(a) AvCM distance table.

	Bare soil	Agriculture 1	Agriculture 2	Agriculture 3	Primary forest	Pasture
Bare soil	0.00	0.90	4.75	3.70	2.05	1.02
Agriculture 1		0.00	7.29	6.02	4.53	3.15
Agriculture 2			0.00	0.42	3.24	2.70
Agriculture 3				0.00	2.87	2.07
Primary forest					0.00	0.41
Pasture						0.00

(b) InCM distance table.
SOURCE: Autor's production.

The San Francisco PolSAR image, presented at Figure 5.6(a), is from the San Francisco Bay area, California, United States. This image is widely used for PolSAR image processing validation due to it diversity in terms of scattering mechanism.

The San Francisco PolSAR image is dominated by the San Francisco city, by the waterway entrance to San Francisco Bay from the Pacific Ocean, called the Golden Gate, and it has also the Golden Gate bridge and vegetation zones. This image is rich in term of urban scattering, volumetric and smooth scattering. Liu et al. (2019) identified six classes in San Francisco PolSAR image, videlicet, Mountain, Water, Vegetation, High Density Urban, Low Density Urban, and Developed Urban, as

presented in Figure 5.6(b).

Figure 5.6 - San Francisco ALOS/PALSAR image, located in the state of California, with the components hh (R), hv (G) and vv (B).

(a) PolSAR image. (b) PolSAR image with classes.

SOURCE: Liu et al. (2019).

Covariance matrices samples from the areas marked as classes in Figure 5.6(b) were selected, and from them, six AvCM and InCM were estimated in order to compute the determinants and trace values presented in Table 5.6. As happened in all above cited images, the determinants values are bigger with AvCM than with InCM.

The Hellinger stochastic distance between the six classes are presented in Figure 5.7(a) and 5.7(b). The classes Vegetation and Mountain are the closest classes, having a distance value of 0.17 in AvCM distance table, and 0.16 in InCM distance table, therefore some confusion between these classes is expected. The Low Density Urban class is also close to Vegetation, having a distance of 0.45 in AvCM distance table, and 0.46 in InCM distance table. The other classes have higher distance values between them.

Table 5.6 - Determinant and trace of the average covariance matrices from San Francisco ALOS/PALSAR image classes.

Class	**AvCM Determinant**	**AvCM Trace**	**InCM Determinant**	**InCM Trace**
Mountain	1.02×10^{32}	1.91×10^{11}	1.94×10^{31}	9.83×10^{10}
Low Density Urban	5.68×10^{32}	2.65×10^{11}	2.17×10^{32}	1.92×10^{11}
High Density Urban	5.68×10^{34}	2.65×10^{12}	1.59×10^{34}	1.78×10^{12}
Water	6.63×10^{29}	1.22×10^{11}	3.64×10^{29}	1.11×10^{11}
Vegetation	9.94×10^{31}	1.51×10^{11}	3.91×10^{31}	1.13×10^{11}
Develop Urban	6.81×10^{32}	3.98×10^{11}	2.74×10^{32}	2.98×10^{11}

SOURCE: Author's production.

Figure 5.7 - Hellinger distances between all classes within San Francisco ALOS/PALSAR image.

	Mountain	Low Density Urban	High Density Urban	Water	Vegetation	Develop Urban
Mountain	0.00	0.66	5.08	3.77	0.17	1.07
Low Density Urban		0.00	4.13	6.18	0.45	1.01
High Density Urban			0.00	11.31	5.26	2.58
Water				0.00	4.45	6.01
Vegetation					0.00	1.08
Develop Urban						0.00

(a) AvCM distance table.

	Mountain	Low Density Urban	High Density Urban	Water	Vegetation	Develop Urban
Mountain	0.00	0.88	5.53	3.39	0.16	1.38
Low Density Urban		0.00	3.98	6.01	0.46	1.01
High Density Urban			0.00	10.46	4.89	2.29
Water				0.00	4.30	5.82
Vegetation					0.00	1.13
Develop Urban						0.00

(b) InCM distance table.

SOURCE: Author's production.

5.1.2 PolSAR simulated image

As discussed in Chapter 2, a radar sensor emits electromagnetic signals which interact with W elementary scatterers within a resolution cell, each scatterer reflects the signal back to the sensor and the received signal S is the result of a coherent sum of W waves reflected by the elementary scatterers. The received signal $S = s_a + js_b$ is, then, transformed into an image pixel value. Assuming the model proposed by Arsenault e April (1976), the speckle formation is modeled as a random walk process in complex domain, and the real and imaginary part of the returned signal, i. e. s_a and s_b, are assumed to be identically Gaussian distributed with zero mean and a variance $\sigma^2/2$.

The PolSAR image pixel can be represented as the complex vector $\vec{S_q}^T = [s_1 + js_4, s_2 + js_5, s_3 + js_6]$, where $s_1 + js_4$ represents the S_{hh} component of the scattering Matrix $\boldsymbol{S}$ (equation (2.10)), $s_2 + js_5$ represents the S_{hv} component, and $s_3 + js_6$ represents the S_{vv} component. $\vec{S_q}$ follows a circularly symmetric multivariate complex Gaussian distribution, denoted by $\vec{S_q} \sim C\mathcal{N}(0, \boldsymbol{\Sigma_q})$. The vector $\vec{S_q}$ can be represented in real domain as the vector $\vec{x_{2q}}^T = [s_1, s_2, s_3, s_4, s_5, s_6]$, such that $\vec{x_{2q}} \sim \mathcal{N}(0, \boldsymbol{\Sigma_{2q}})$, where $\boldsymbol{\Sigma_{2q}}$ is described as (SILVA, 2013):

$$\boldsymbol{\Sigma_{2q}} = \frac{1}{2}\begin{bmatrix} \Re(\boldsymbol{\Sigma_q}) & -\Im(\boldsymbol{\Sigma_q}) \\ \Im(\boldsymbol{\Sigma_q}) & \Re(\boldsymbol{\Sigma_q}) \end{bmatrix} = \frac{1}{2}\boldsymbol{\Sigma_{2q}}' \tag{5.1}$$

where $\Re(.)$ and $\Im(.)$ represents the real and imaginary part of a complex number, and q is the $\boldsymbol{\Sigma_q}$ order. Considering that $S_{hv} = S_{vh}$, the matrix Σ_q is described by equation (2.21). Therefore the matrix Σ_{2q} can be expanded as:

$$\boldsymbol{\Sigma_{2q}}' = \begin{bmatrix} \Re(|S_{hh}|^2) & -\Im(|S_{hh}|^2) & \Re(\sqrt{2}S_{hh}.S_{hv}^*) & -\Im(\sqrt{2}S_{hh}.S_{hv}^*) & \Re(S_{hh}.S_{vv}^*) & -\Im(S_{hh}.S_{vv}^*) \\ \Im(|S_{hh}|^2) & \Re(|S_{hh}|^2) & \Im(\sqrt{2}S_{hh}.S_{hv}^*) & \Re(\sqrt{2}S_{hh}.S_{hv}^*) & \Im(S_{hh}.S_{vv}^*) & \Re(S_{hh}.S_{vv}^*) \\ \Re(\sqrt{2}S_{hv}.S_{hh}^*) & -\Im(\sqrt{2}S_{hv}.S_{hh}^*) & \Re(2|S_{hv}|^2) & -\Im(2|S_{hv}|^2) & \Re(\sqrt{2}S_{hv}.S_{vv}^*) & -\Im(\sqrt{2}S_{hv}.S_{vv}^*) \\ \Im(\sqrt{2}S_{hv}.S_{hh}^*) & \Re(\sqrt{2}S_{hv}.S_{hh}^*) & \Im(2|S_{hv}|^2) & \Re(2|S_{hv}|^2) & \Im(\sqrt{2}S_{hv}.S_{vv}^*) & \Re(\sqrt{2}S_{hv}.S_{vv}^*) \\ \Re(S_{vv}.S_{hh}^*) & -\Im(S_{vv}.S_{hh}^*) & \Re(\sqrt{2}S_{vv}.S_{hv}^*) & -\Im(\sqrt{2}S_{vv}.S_{hv}^*) & \Re(|S_{vv}|^2) & -\Im(|S_{vv}|^2) \\ \Im(S_{vv}.S_{hh}^*) & \Re(S_{vv}.S_{hh}^*) & \Im(\sqrt{2}S_{vv}.S_{hv}^*) & \Re(\sqrt{2}S_{vv}.S_{hv}^*) & \Im(|S_{vv}|^2) & \Re(|S_{vv}|^2) \end{bmatrix} \tag{5.2}$$

The pixel simulation is done by sampling the vector $\vec{x_{2q}}$ using the covariance $\boldsymbol{\Sigma_{2q}}$. The first q elements are the real part of $\vec{S_q}$, while the last q elements are the imaginary part of $\vec{S_q}$. This process is executed in order to create a single look image PolSAR image. The multilook covariance PolSAR image is achieved by:

$$\mathbf{\Sigma} = \frac{1}{L}\sum_{i=1}^{L} \vec{S_q}.\vec{S_q}^{\,T} \tag{5.3}$$

The covariance matrices $\mathbf{\Sigma_q}$ used as samples in the simulation process were estimated from Bebedouro SIR-C L band image. Six classes from the nine described in Section 5.1.1.1 were selected: Corn 2, Soy 1, Soy 3, Bare Soil, Caatinga, and River. The covariance matrices $\mathbf{\Sigma_q}$ were estimated by the arithmetic mean, in order to follow the procedure proposed by Silva (2013). The covariance matrices used to create the simulated data are described from equation (5.4) to (5.9). Since the matrices are symmetric, only the upper triangular is shown in the following equations.

$$\mathbf{\Sigma_{Corn2}} = \begin{bmatrix} 4.53\times10^{-2} & 3.08\times10^{-3} - j6.01\times10^{-4} & 7.30\times10^{-3} - j1.39\times10^{-4} \\ & 1.08\times10^{-2} & 4.10\times10^{-4} - j1.79\times10^{-4} \\ & & 4.98\times10^{-2} \end{bmatrix} \tag{5.4}$$

$$\mathbf{\Sigma_{Soy1}} = \begin{bmatrix} 1.82\times10^{-2} & 1.86\times10^{-4} + j7.78\times10^{-4} & 1.26\times10^{-4} + j4.37\times10^{-3} \\ & 3.43\times10^{-3} & 3.94\times10^{-4} - j5.83\times10^{-6} \\ & & 4.05\times10^{-2} \end{bmatrix} \tag{5.5}$$

$$\mathbf{\Sigma_{Soy3}} = \begin{bmatrix} 9.26\times10^{-2} & -9.72\times10^{-3} + j7.73\times10^{-3} & 5.44\times10^{-4} - j1.01\times10^{-2} \\ & 1.65\times10^{-2} & -5.16\times10^{-3} + j2.98\times10^{-3} \\ & & 4.41\times10^{-2} \end{bmatrix} \tag{5.6}$$

$$\mathbf{\Sigma_{BareSoil}} = \begin{bmatrix} 1.20\times10^{-2} & 8.11\times10^{-5} + j1.69\times10^{-4} & 8.65\times10^{-3} - j1.61\times10^{3} \\ & 8.12\times10^{-4} & -2.91\times10^{-5} - j2.79\times10^{-5} \\ & & 1.15\times10^{-2} \end{bmatrix} \tag{5.7}$$

$$\mathbf{\Sigma_{Caatinga}} = \begin{bmatrix} 1.25\times10^{-1} & -5.44\times10^{-3} + j5.65\times10^{-5} & 7.22\times10^{-3} - j1.54\times10^{-2} \\ & 4.59\times10^{-2} & -4.12\times10^{-3} - j5.71\times10^{-3} \\ & & 1.40\times10^{-1} \end{bmatrix} \tag{5.8}$$

$$\boldsymbol{\Sigma}_{River} = \begin{bmatrix} 3.40 \times 10^{-3} & 3.38 \times 10^{-5} + j7.75 \times 10^{-5} & 4.32 \times 10^{-3} - j5.96 \times 10^{-4} \\ & 3.80 \times 10^{-4} & -5.58 \times 10^{-5} - j1.35 \times 10^{-4} \\ & & 1.31 \times 10^{-2} \end{bmatrix} \quad (5.9)$$

The simulated images size is 240 by 240 pixels, and each class were simulated as a square segment. The number of segments within an image and the classes location are random. An example of simulated image is shown in Figure 5.8(a), this image has 64 segments and six classes. The Corn 2 class has eleven segments, the Soy 1 has eight segments, the Soy 3 has thirteen segment, the Bare Soil has nine segments, the Caatinga class has nine, and the River has fourteen segments. The truth map image (Figure 5.8(b)) is generated for each simulated image, and this image will be later used to compute the classification result accuracy.

Figure 5.8 - Simulated PolSAR image with the components hh (R), hv (G) and vv (B).

(a) Simulate image. (b) Simulated truth image.

SOURCE: Author's production.

According to (FRERY et al., 2013) the number of looks alters the data set distribution in a non-linear way, which can be perceived by the stochastic distances. As lower the number of looks, more sensitive are the distances to smaller differences between classes, leading to a noisier classification result. Because of that, the PolSAR images were simulated with a number of looks equals to five.

Since the images were simulated under the circularly symmetry assumption, the real and imaginary parts are assumed to be equally Gaussian distributed with zero mean and a variance $^{\sigma^2}/_2$. Therefore, the PolSAR multilook image in amplitude follows the square root Gamma distribution, as discussed in Section 2.4.1. Figure 5.9(a) shows the square root Gamma distribution of S_{hh} PolSAR simulated image, presented in Figure 5.8(a), classes. In the same way, Figure 5.9(b) presents the square root Gamma distribution of S_{hv} PolSAR image classes. Finally, the Figure 5.9(c) presents the square root Gamma distribution of S_{vv} PolSAR image classes. The classes distribution are represented by curves, where the Corn 2 is represented by the purple curve, the Soy 1 by the salmon curve, the Soy 3 by the maroon curve, the Bare Soil is represented by the red curve, the Caatinga is represented by the green curve, and the River is represented by the blue curve.

As presented by Table 5.2, the Caatinga is the class with highest determinant, therefore the brighter one. This behaviour can be checked also in Figures 5.9(a), 5.9(b), and 5.9(c). The Caatinga curves are the ones with highest mean, and with the highest variance. Differently, the River is the class with the lowest determinant, and its curves are the ones with lowest mean value and less spreaded, i. e., with low variance values. As discussed in Figure 5.2, the Caatinga and River are the most distant classes, with 10.83 of distance value, what can be perceived by the curves as well.

Further, in Figure 5.2 the closest classes are Corn 2 and Soy 1, with distance value of 0.19, and in Figures 5.9 this behavior is also confirmed, the Corn 2 curve (the purple one) is even covered by Soy 1 curve (the salmon one), in S_{hh} and S_{hv} figures. However in S_{vv} curves (Figure 5.9(c)), the Soy 1 curve is more similar to Soy 3 curve, meaning that the vertical polarimetric answer between Soy 1 and Soy 3 are probably closer than Soy 1 and Corn2.

In Figure 5.2, the Bare Soil and River are also close classes, with a distance of 0.57. In Figures 5.9, the Bare Soil and River curves are more similar in th S_{vv} band, and more dissimilar in S_{hh} band.

Figure 5.9 - Simulated PolSAR image square root Gamma distribution.

(a) S_{hh} square root Gamma distribution.

(b) S_{hv} square root Gamma distribution.

(c) S_{vv} square root Gamma distribution.

SOURCE: Author's production.

5.2 Experiments description

In order to evaluate the BSC classification result, four different experiments were conducted. The first experiment address the algorithm initialization problem in the SC algorithm. The second experiment explore the pros and cons of using the Rie-

mannian geometry to estimate covariance matrices. The third experiment has the goal of quantify the proposed algorithms accuracy, therefore it is focused on the quantitative analysis. In this experiment a Monte Carlo simulation over a set of one hundred simulated PolSAR images was conducted. The fourth experiment was conducted using a set of real PolSAR images in different frequency bands and from different scenarios.

5.2.1 Experiment I

In Carvalho et al. (2019) an analysis about the algorithm initialization problem in the SC algorithm were addressed. The SC algorithm, described in Section 4.2.2, identifies K clusters and allocates every data sample to the closest cluster, by using as similarity metric the stochastic distances.

The SC, and consequently the BSC, are greedy algorithms (JAIN, 2010), i.e., these algorithms make the optimal choice at each step with the goal of finding the global optimal solution to the entire problem. Due to that, these algorithms are able to converge to the global optimum only when clusters are well separated. In (CARVALHO et al., 2019), a Monte Carlo simulation over 100 simulated PolSAR images was performed. The simulation exposed the algorithms dependency upon the correctness of the initial parameters determination. Randomly determined initial parameters may terminate at a local minimum, meaning that the algorithm may run a lot of iterations trapped in segments away from the global minimum, resulting in incorrect clustering results.

In order to investigate the classification behaviour according to the initial parameter determination, the PolSAR image with six classes, showed in Figure 5.8(a) was used as the input for SC algorithm using the Bhattacharyya, Hellinger, and Kullback-Leibler distances as similarity metrics. The SC algorithm requires the initial parameter determination, therefore, six scenarios to force the bad and good initial parameters choosing were defined:

a) S01: All six initial parameters were selected from the same class;

b) S02: The six initial parameters are distributed over three classes;

c) S03: The six initial parameters were picked from the borders of two classes;

d) S04: Three initial parameters were selected in three different class, and the other three comes from the borders of two classes;

e) S05: All initial parameters comes from outliers;

f) S06: One initial parameters were picked per class.

Good initial parameters are defined as the ones that allows a faster and assertive convergence. On the other hand, bad initial parameters can lead to increased numbers of required clustering iterations to reach convergence, a greater overall run-time, and a less-efficient algorithm overall accuracy. The scenarios S01, S02, S03, S04, and S05 are considered bad scenarios, while S06 is the best possible scenario.

Figure 5.10 shows the flowchart of the initial parameter determination evaluation schema. In this experiment the three variants of SC algorithm, according to the stochastic distance, are used. The first SC variant uses the Bhattacharyya distance as similarity metric, and it is named SC-B. The second variant is the SC with Hellinger distance, named SC-H, and the last variant is the SC using Kullback-Leibler distance, named SC-KL. For each scenario, the initial parameter are determined and the SC three variants share the same initial parameter. In the end the accuracy of every classification image shall be computed.

Figure 5.10 - Initial parameter determination analysis flowchart.

SOURCE: Author's production.

5.2.2 Experiment II

This section describes the experiment II, which is responsible for compare the covariance arithmetic mean versus the covariance intrinsic mean. Here, the algorithm proposed in Section 4.2.1.1 is addressed, and two variants of this algorithm is defined: 1) PDDP using the AvCM, named only PDDP; 2) PDDP using InCM, named

RPDDP. The idea is to split the PolSAR images into two big clusters and then analyse the intrinsic mean performance.

The input images are the covariance representation of PolSAR images, therefore the samples are a set of covariance matrices $Z = \{\boldsymbol{Z_1}, \ldots, \boldsymbol{Z_n}\}$ with N samples and L number of looks. The average covariance matrix $\boldsymbol{\Sigma_{AvCM}}$ is computed by the arithmetic mean, while the $\boldsymbol{\Sigma_{InCM}}$ is estimated by the intrinsic mean. From the average covariance matrix, the eigenvalues and eigenvectors are estimated, and the samples are grouped into two classes, according with the Equation 4.1 sign result. This procedure is done for both $\boldsymbol{\Sigma_{AvCM}}$ and $\boldsymbol{\Sigma_{InCM}}$.

The PDDP and RPDDP algorithms variants are evaluated using the four real images described in Section 5.1.1. As evaluation criteria, the stochastic distance between classes and the Root-Mean-Square Standard Deviation (RMSSTD) error (Equation 3.9) are computed. The goal of these evaluation criteria is to judge the cluster quality by assuming that an optimum clustering means compact clusters (i. e., cluster with low RMSSTD error), and well-separation from other clusters (i. e., bigger stochastic distances between clusters). Figure 5.11 shows the Experiment II flowchart.

Figure 5.11 - Σ_{AvCM} versus Σ_{InCM} comparisson flowchart .

SOURCE: Author's production.

5.2.3 Experiment III

In this experiment, a set of 100 simulated PolSAR images shall be classified by the SC and BSC using the stochastic distances of Bhattacharyya, Hellinger, and Kullback-Leibler distances. Therefore, in this work, the PolSAR images were classified using the techniques presented in Table 5.7.

Table 5.7 - Algorithms used for the PolSAR images classification analysis.

Characteristic	**Parameter Selection**	**Stochastic Distance**	**Algorithm initials**
Partitional	Random	Bhattacharyya	SC-B
		Kullback-Leibler	SC-KL
		Hellinger	SC-H
Hierarchical	RPDDP	Bhattacharyya	BSC-R-B
		Kullback-Leibler	BSC-R-KL
		Hellinger	BSC-R-H
	EM	Bhattacharyya	BSC-EM-B
		Kullback-Leibler	BSC-EM-KL
		Hellinger	BSC-EM-H

SOURCE: Author's production.

As described in Figure 4.1, before the data splitting, the BSC algorithm must determine the initial parameter. The same procedure is required to the SC algorithm. As described in Table 5.7, the SC algorithm, described in Section 4.2.2, randomly selects the initial parameters, while in the BSC algorithm, described in Section 4.2, the parameters are estimated in two ways: or using the RPDDP, or using the EM algorithm.

After the K initial parameter determination the SC algorithm allocates every data sample the closest cluster, represented by its parameter. After the allocation, K new parameters are estimated and the allocation procedure starts again. These steps will be performed until the stop criteria is reached.

The BSC algorithm, is a divisive procedure that bi-partitions the data set using the SC algorithm at each dendrogram level. Therefore for every new created node, two initial parameter must be determined.

In order to quantitatively evaluate the BSC and SC algorithms, a Monte Carlo simulation over a set of one hundred simulated PolSAR images is conducted. Figure 5.12 presents the Monte Carlo simulation flowchart. At each Monte Carlo simulation iteration, one of the simulated PolSAR image is selected, and from this image the initial parameters are defined, or randomly for the SC algorithm, or by RPDDP or EM for the BSC algorithm. The simulated PolSAR image is tested by the nine above cited algorithms, therefore, in the end, 900 classified images are generated.

Figure 5.12 - Monte Carlo Simulation flowchart.

SOURCE: Author's production.

In this experiment only the quantitative analysis is addressed. The classified image is compared against its correspondent truth image and the confusion matrices are built (see Section 6.4.1). From the confusion matrix the accuracy values are computed, and with this values, the classification result with better accuracy and performance can be determined.

5.2.4 Experiments IV

The second experiment has the goal of ratifying the Monte Carlo simulation results. In this experiment four real PolSAR images, described in Section 5.1.1 are analysed.

This experiment evaluates the classification in both quantitative and qualitative ways. However, since the Monte Carlo simulation showed that the Bhattacharyya, Kullback-Leibler, and Hellinger stochastic distances achieved similar accuracy results for SC and BSC algorithms, only the Hellinger distance was selected to perform the second experiment analysis. In this case the images were classified by the:

a) SC using Hellinger distance (SC-H);

b) BSC using Hellinger distance + RPDDP (BSC-R-H);

c) BSC using Hellinger distance + EM (BSC-EM-H).

5.3 Quantitative analysis

The confusion matrix, as explained in Section 3.4.2, is an external cluster validation method used to quantitatively characterise the image classification accuracy. It consists in a table that shows the correspondence between the classification result and the truth image. Therefore, to compute the confusion matrix it is needed to know the ground truth data.

The confusion matrix rows represent the truth classes, and the columns refer to cluster outcomes results. The elements e_{ij}, with $i = j$, represents the number of samples correctly classified; while the elements e_{ij}, with $i \neq j$, represent the incorrectly classified samples. Figure 5.13 illustrates a confusion matrix, the left picture represents the truth image with its labels, and the up center picture represents the classified image. The classified image pixels are compared against the truth image pixels, and the confusion matrix elements show the number of pixels for all the possible correlations between the truth and the classified image.

In order to make the analysis clearer and easier to understand, the confusion matrix elements values are transformed into percentage values. In this way, the e_{ij} value varies from 0 to 100. Figure 5.14 shows how the confusion matrix shall be presented hereafter. The balls diameter varies according to the e_{ij} value, the bigger the value, bigger the diameter. The e_{ij} values are displayed in the confusion matrix chart only if they are higher than 10. The confusion matrix columns values should sum up to

100. The ideal confusion matrix would have 100 value on its main diagonal, meaning that all samples were correctly classified.

Figure 5.13 - Confusion matrix.

Truth	a	b	c	d	e	f
A	8379	907	381	0	0	6
B	4049	2625	214	0	3	2
C	245	198	6717	0	26	0
D	66	514	60	6604	0	227
E	101	37	148	0	5980	0
F	75	482	90	230	0	10754

SOURCE: Author's production.

Figure 5.14 - Confusion matrix representation.

SOURCE: Author's production.

The Figure 5.14 title shows the overall accuracy result. The accuracy is computed according to Equation (3.12). The classification accuracy is the ratio of correct predictions over the total predictions made, therefore it is a simple and good indicative about the correctness of a classification.

Although the accuracy is a very useful metric, it has two major two drawbacks: the values out of the diagonal are not used, and the classes with a low number of sample are lower weighted in the calculation. Therefore it is important to analyse the confusion matrix together with the accuracy result.

5.4 Qualitative analysis

The qualitative analysis seeks for features that are observed and can not be measured with a numerical result in order to validate the BSC algorithm classified image. In this work the qualitative analysis uses the cluster dendrogram, and the scattering mechanism. In Section 5.4.1 the dendrogram analysis is explained. Section 5.4.2 presents the Plan $H - \alpha$, which is a product of the incoherent target decomposition based on eigenvalues analysis proposed by Cloude e Pottier (1996), whose goal is to describe the polarimetric data in terms of the scattering mechanism. The idea is to use the scattering mechanism information to help the validation/characterization of unknown clustered classes.

5.4.1 Dendrogram analysis

The BSC algorithm is a divisive hierarchical clustering algorithm that splits the data set into two sub-clusters at each iteration. The dataset splitting history is saved on a structure named dendrogram, which is a binary tree used to visually represent the hierarchical relationship between the sub-clusters. The dendrogram contains nodes and branches, where each node represents one sub-cluster, and stores information as the cluster ID, entropy, and information gain between the actual node and its parent node. Every node, as described in Figure 3.3, is connected by branches that represent the hierarchical relationships between the sub-clusters.

The information gain is used to measure the entropy decrease within sub-clusters while the image or dataset is splitted. Ideally, the entropy should decrease after a cluster splitting, and one cluster is considered completely homogeneous if the entropy is zero. In that way, the dendrogram grows while seeking for groups that returns the highest information gain.

In order to elucidate how the dendrogram analysis should be performed, Figure 5.15

presents a BSC algorithm dendrogram. As presented in Figure 5.14, the truth image has the classes A, B, C, D, E, and F. Therefore the first assumption is that the input image has at least these six classes, thereat the first node has classes A, B, C, D, E, F. Since clusters IDs are sorted in crescent way and respecting the splitting sequence, the first node has ID 1. This node is splitted into sub-cluster ID 2 and sub-cluster ID 3.

Figure 5.15 - Simulated PolSAR image dendrogram, generated by the BSC-R-H algorithm.

SOURCE: Author's production.

As explained in Section 4.2.3, except for the root node (ID 1), it is needed to choose which node shall be splitted at each BSC iteration. The selected node is determined by the information gain value between the current nodes in a given level and their possible children. The possible children parameters are determined by the initial parameter selector algorithm (RPDDP or EM, in our approach). Once a particular node is selected, the SC algorithm, using the already determined initial parameter, finds the new sub-clusters. This procedure is internal to the algorithm, therefore, to perceive which is the next node in the splitting queue, the user must follow the ID order. The information gain presented (I) in Figure 5.15 is computed between the current node and its parent, therefore it represents the change in entropy inside a sub-cluster after the parent node cluster splitting. For instance, node ID 2 has $I = 6.88$ and node ID 3 has $I = 13.76$.

In Figure 5.15, the next splitting occurs to cluster ID 2, where the classes A, B, C, D, and E were splitted into A, B, and C (ID 5), and E (ID 4). Sub-cluster ID 5 is splitted into A and B (ID 6) and C 3 (ID 7). Then cluster ID 3 is finally splitted into D (ID 8) and F (ID 9), and successively. The BSC algorithm, in the example, created two new classes, meaning that the classified image has 8 classes named: a, b, c, d, e, f, new1 and new2.

The information gain tends to decrease while the dendrogram grows, however it is not rule. Sub-clusters which are too different from their parent node cluster can have a higher information gain value, for instance, the node ID 4 has a bigger information gain than node ID 2.

As already discussed in Section 4.2.4, the dendrogram leaves carry an important information about the clusters: the dissimilarity between classes. As far one class is from another, the more different they are in relation to each-other. Lets pick the f and e class, for example, these are unrelated classes, meaning that f could be, for example, a river classes while e could be a vegetation class.

5.4.2 Scattering mechanism analysis

In this work, the method based on an eigenvalue analysis proposed by Cloude and Pottier, which is presented in Section 2.3.2.2, was chosen for the cluster splitting qualitative analysis. The Cloude-Pottier decomposition characterize the PolSAR images scattering mechanisms given their Entropy (H) and Alpha Angle (α) parameters.

The entropy and angle alpha can be represented in a graphical space named Plan $H - \alpha$, as described in Figure 2.8. The goal of using Plan $H - \alpha$ is to analyse if the BSC splitting makes sense, and if the samples with different scattering mechanism are separated.

Figure 5.16 extrapolate one dendrogram node and its connections into a picture representation. This figure has two levels: the parent level, identified as Level 1, and the children level, identified as Level 2. On Level 1, there are three pictures, and the middle picture represents the input image cluster, the picture on its right side represents the Plan $H - \alpha$ of the entire data set. The Plan $H - \alpha$ is represented by a heat map, were the red spots are the regions with high sample concentration.

The cluster ID 1 is splitted into two subclusters, ID2, represented by the orange color, and ID 3 represented by the green color. In level 2 the cluster split and the

correspondent Plan $H - \alpha$ of ID 2 and ID 3 are presented.

Figure 5.16 - Plan $H - \alpha$ of image first splitting.

SOURCE: Author's production.

The first picture in Level 1 summarizes the scattering mechanism representation. The little balls in the plan represents the average entropy and α angle for each node, for instance, the node ID 1 is represented by a gray ball, and node ID 2 is represented by the orange ball, while the node ID 3 is represented by the green ball.

6 RESULTS

This chapter presents the BSC classification results. In Section 6.1 the initial parameter determination importance is outlined. Section 6.2 describes the pros and cons of using the Riemann geometry to compute the data average when using SPD matrices as feature. After that, the simulated PolSAR images classification results is outlined in Section 6.3, were the results are quantitatively evaluate, by using confusion matrices in order to delivery the classification accuracy values. Finally, the experiment using the real PolSAR images classification is presented in Section 6.4, in this section the dendrogram information is explored together with the Plan $H-\alpha$ information, in order to provide a qualitative analysis of the classifications results.

6.1 Experiment I - initial parameter determination

In Section 5.2.1 the experiment methodology is described. In this experiment, the SC algorithm greedy behavior is explored. Six scenarios are defined and their classification accuracy result was calculated by using the truth image associated to the simulated PolSAR image, as explained in Section 5.3.

Figure 6.1 and Table 6.1 present the accuracy results for each classification algorithm per scenario. The scenarios S05, where the initial parameter where selected from outliers samples, has the worst classification result. This scenario achieved only 17.26 % of overall accuracy. The second worst scenario is the S01, where all initial parameters are picked from the same class, this scenario achied 35.55 % of overall accuracy with SC-B and SC-H, and 54.49 % with SC-KL.

The scenarios S02, S03, and S04 are intermediate scenarios, with an average accuracy of approximately 60 % for all SC variants. However, when the parameters have a good fit, i.e., when the parameters are taken from scenario S06 for instance, the classification results achieves high accuracy for all distances, accomplishing around 82 % of overall accuracy.

Figures 6.2 to 6.7 enlighten the accuracy results showed in Figure 6.1 and Table 6.1. Each figure has six sub-figures, where the sub-figures on the left (a), (c), and (e) show the PolSAR image classification using the SC-B, SC-KL, and SC-H, respectively. And the sub-figures (b), (d), and (f) present the correspondents confusion matrices. As discussed in Section 5.3, the confusion matrix rows represents the ground truth, given by the truth image, and the columns represent the algorithm outcomes. The confusion matrix main diagonal elements e_{ij} represent the number

of samples corrected classified; while off-diagonal elements e_{ij}, with represent the misclassified samples. The e_{ij} value varies from 0% to 100%.

Figure 6.1 - Accuracy of the simulated PolSAR image classification using the SC algorithm with different initial parameters.

SOURCE: Author's production.

Table 6.1 - Initial parameter scenarios accuracy [%].

Algorithm	**S01**	**S02**	**S03**	**S04**	**S05**	**S06**
SC-B	35.55	59.16	62.32	58.05	17.26	82.09
SC-KL	54.49	61.52	62.51	58.19	17.26	81.92
SH-H	33.55	59.16	62.32	58.05	17.26	82.09

SOURCE: Author's production.

Figure 6.2 describes the classification results of SC algorithm using the initial parameters taken from the S01. Figure 6.2(a) shows the SC-B classification result and Figure 6.2(b) shows the related confusion matrix. Using the Figure 5.8 as pattern, it is possible to see that the SC-B algorithm in S01 was able to correctly identify the

Corn 2 class, however the Soy 1 class was also classified as Corn 2, therefore in the column related to Corn 2 the values are spread, mainly, between Corn2 and Soy1. The Corn 2-Corn 2 cell in Figure 6.2(b), for instance, has an accuracy of only 51 %.

As happened with Corn 2 and Soy 1, the Bare Soil were correctly classified, but its accuracy dropped because the River class was classified as Bare Soil. Remember that the PolSAR simulated image has nine Bare Soil segments and fourteen River segments, for this reason the River-Bare Soil cell has a high percentage value (61 %) than Bare Soil-Bare Soil cell (39 %).

The Caatinga class, looking only at Figure 6.2(b), had a good result, achieving 99 % of accuracy. However this accuracy is not so related to reality, because only few pixels that truly belongs to Caatinga class were actually classified as Caatinga class. Indeed, a great amount of samples that are from Caatinga class were classified as Soy 3, instead. Therefore the cell Soy 3-Caatinga had a high percentage value (92 %).

Figure 6.2(c) shows the SC-KL classification result and Figure 6.2(d) presents the related confusion matrix. In this classification, the Caatinga were truly classified as Caatinga class, while the Soy 3 were correctly classified as Soy 3. The confusion between Corn 2 and Soy 1, and Bare Soil and River is present in the SC-KL classification also. Figure 6.2(e) presents the SC-H classification result and Figure 6.2(f) shows the related confusion matrix. The SC-B and SC-H had an analogous results, what is excepted due to the close relationship between the Hellinger and Bhattacharyya distance (CARVALHO et al., 2019). Note that the SC-B and SC-H alike results are a pattern, that will be repeated in every scenario.

It is important to recall Figure 5.2, where the stochastic distances between classes are presented. By that figure, Corn 2 and Soy 1 are the most similar classes, and they are also close related to Soy 3. In the same way River and Bare Soil are quite close and similar. Therefore a confusion between these classes are expected.

In scenario S02 the six initial parameters are picked over three different classes. The Figure 6.3(a) shows the SC-B classification result and Figure 6.3(b) shows the related confusion matrix. In this scenario, the SC-B was able to correctly identified the Caatinga and Soy 3. Also, it was fairly able to separate the Corn 2 and Soy 1 classes. As in S01, the river were completely classified as Bare Soil. However, some samples from Caatinga class were wrongly classified as River.

Figure 6.3(c) presents the SC-KL classification result and Figure 6.3(d) shows the related confusion matrix. SC-KL classification result, in S02, is similar to SC-B, excepted that in SC-KL classification less Caatinga samples were classified as River. Figure 6.3(e) shows the SC-H classification result and Figure 6.3(f) shows the related confusion matrix. The SC-H result is very similar to SC-B. As already discussed, these two algorithm tend to have similar answers.

The classification results for scenario S03, where the initial parameters were picked from the borders of two classes, were very similar to S02. Figure 6.4(a) shows the SC-B classification result and Figure 6.4(b) shows the related confusion matrix. Figure 6.4(c) shows the SC-KL classification result and Figure 6.4(d) shows the related confusion matrix. Figure 6.4(e) shows the SC-H classification result and Figure 6.4(f) shows the related confusion matrix. Again, Corn 2 and Soy 1 are mixed, and the River still classified as Bare Soil. However the Caatinga class is much less nosier in S03.

In S04, three initial parameters were selected from three different class, and the other three comes from the borders of two classes. Figures 6.5(a), 6.5(c), and 6.5(e) shows the SC-B, SC-KL, and SC-H classification result, respectively, and Figures 6.5(b), 6.5(d), and 6.5(f) shows the SC-B, SC-KL, and SC-H confusion matrices, respectively. In the S04 scenario, almost 40% of the samples classified as Corn 2, were actually Soy 1, in all SC variants. The Soy 3 achieved a great classification, while the Caatinga was fairly well classified, but some Caatinga samples still classified as River. The River still classified as Bare Soil.

The worst classification result came with S05 (Figure 6.6), where all initial parameters come from outliers. In this scenario the image was classified as one single class in all SC variants, because the algorithm got stuck in a local minimum.

As expected, the best result came with S06, which is shown in Figure 6.7. In this scenario one initial parameter was picked per class, therefore this is the best possible situation. In this case, the algorithm found the global minimums in a quick and effective way. The Soy 3, Bare Soil, Caatinga and River achieved a great accuracy value, over 90 %. The Corn 2 and Soy 1 were mixed, however this an expected behavior, due to these class similarity. Given the results of the initial parameter choice scenarios, it is clear that the good candidates estimation for the SC initial parameter determination is essential for improving the clustering quality and performance, allowing the algorithm to have a faster and accurate convergence.

Figure 6.2 - Classification result S01.

(a) SC-B.

(b) Confusion matrix -SC-B.

(c) SC-KL.

(d) Confusion matrix - SC-KL.

(e) SC-H.

(f) Confusion matrix - SC-H.

SOURCE: Author's production.

Figure 6.3 - Classification result S02.

(a) SC-B.

(b) Confusion matrix -SC-B.

(c) SC-KL.

(d) Confusion matrix - SC-KL.

(e) SC-H.

(f) Confusion matrix - SC-H.

SOURCE: Author's production.

Figure 6.4 - Classification result S03.

(a) SC-B.

(b) Confusion matrix -SC-B.

(c) SC-KL.

(d) Confusion matrix - SC-KL.

(e) SC-H.

(f) Confusion matrix - SC-H.

SOURCE: Author's production.

Figure 6.5 - Classification result S04.

(a) SC-B.

(b) Confusion matrix -SC-B.

(c) SC-KL.

(d) Confusion matrix - SC-KL.

(e) SC-H.

(f) Confusion matrix - SC-H.

SOURCE: Author's production.

Figure 6.6 - Classification result S05.

(a) SC-B.

(b) Confusion matrix -SC-B.

(c) SC-KL.

(d) Confusion matrix - SC-KL.

(e) SC-H.

(f) Confusion matrix - SC-H.

SOURCE: Author's production.

Figure 6.7 - Classification result S06.

(a) SC-B.

(b) Confusion matrix -SC-B.

(c) SC-KL.

(d) Confusion matrix - SC-KL.

(e) SC-H.

(f) Confusion matrix - SC-H.

SOURCE: Author's production.

6.2 Experiment II - covariance arithmetic mean versus covariance intrinsic mean

The goal of this section is to explore the pros and cons of using the Riemannian geometry to estimate covariance matrices. The methodology and the evaluation criteria of this experiment was described in Section 5.2.2.

The use of intrinsic mean is indicated to estimate positive-definite matrices, its calculation involves the use of logarithm and exponential operator. Therefore, the pro of intrinsic mean is that it can estimate a more assertive and less biased average covariance matrix, and the con is the time consuming; since the algorithm should run until convergence, the intrinsic mean is computationally expensive in comparison with arithmetic mean.

As already discussed in Section 5.2.2, the PDDP and RPDDP algorithms aims to create two clusters. These clusters are presented in this section in a form of a segmented image, where the clusters are represented by the red segments, and a clustered image. The clustered image has two classes: the Black and the Gray.

The first analysed image is the Bebedouro SIR-C in C band PolSAR image. Figures 6.8(a) presents the RPDDP segmented image and 6.8(c) presents the RPDDP clustered image with the two classes Black and Gray. Figures 6.8(b) and 6.8(d), shows the PDDP segmented image and the clustered image respectively. The Black and Gray clusters generated by both RPDDP and PDDP algorithms are very similar, however the RPDDP segmented and clustered images are slightly less nosier than PDDP output images, i. e., RPDDP images have less segments than the PDDP images. On the other hand, the region marked as Bare Soil in Figure 5.1(c) is better clustered by the PDDP algorithm.

Table 6.2 summarizes the results of the evaluation criteria. The first row in the table refers to the Hellinger stochastic distance between the Black and Gray clusters generated by PDDP and RPDDP. In the second row the RMSSTD error of the Black cluster class is presented, and the third row presents the RMSSTD error of the Gray cluster. By Table 6.2, it is possible to conclude that the RPDDP, for this particular image, generates a more homogeneous clusters, in comparison with the PDDP algorithm. The Black and Gray cluster have a lower RMSSTD error with RPDDP algorithm, than with PDDP. The clusters also have a better separation inter cluster, i. e., a bigger distance error, with RPDDP algorithm.

Figure 6.8 - RPDDP and PDDP algorithm applied on Bebedouro SIR-C C band image.

(a) RPDDP segmented image.

(b) PDDP segmented image.

(c) RPDDP clustered image.

(d) PDDP clustered image.

SOURCE: Author's production.

Table 6.2 - Bebedouro SIR-C C band image PDDP and RPDDP analysis.

Evaluation criteria	**PDDP**	**RPDDP**
Hellinger distance	5.5402	6.9937
RMSSTD error - Black class	7.3551e-16	4.6496e-16
RMSSTD error - Gray class	4.5048e-16	2.6194e-16

SOURCE: Author's production.

Following the same procedure, Figure 6.9 shows the result of RPDDP and PDDP algorithms applied on the Bebedouro SIR-C in L band PolSAR image. Figures 6.9(a) and 6.9(b) presents the RPDDP and PDDP segmented images, respectevely. While Figures 6.9(c) and 6.9(d) show the RPDDP and PDDP clustered images.

The RPDDP algorithms clustered the Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, Caatinga, and Tillage classes in the Gray cluster, and the Bare Soil and River classes in the Black Cluster. The PDDP algorithm clustered the Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, and Caatinga classes in the Gray cluster, and the Tillage, Bare Soil, and River in the Black Cluster. Therefore, the major difference between RPDDP and PDDP clustering is the Tillage clustering decision, whether it is more similar to the vegetation classes, as in RPDDP, or more similar to the surface mechanisms classes (River and Bare Soil), as in PDDP algorithm.

The tillage practices disturb soils in different ways, it can use manual or mechanical agitation of various types, such as digging, stirring, and overturning, leaving amounts of crop residues on the soil surface. Unfortunately there is not much information about the level of disturbance in the soil on the region classified as Tillage in SIR-C images. Therefore, it is needed a deeper investigation on the Tillage scattering mechanisms to decide whether this class would fit better with the vegetation mechanisms or with the surface mechanism.

Table 6.3 summarizes the evaluation criteria results for Bebedour SIR-C in L band PolSAR image. Although there is a doubt about the Tillage class clustering, the RPDDP clusters seams to be more homogeneous than the PDDP, having a bigger stochastic distance between clusters and lower RMSSTD error inner clusters.

Figure 6.9 - RPDDP and PDDP algorithm applied on Bebedouro SIR-C L band image.

(a) RPDDP segmented image.

(b) PDDP segmented image.

(c) RPDDP clustered image.

(d) PDDP clustered image.

SOURCE: Author's production.

Table 6.3 - Bebedouro SIR-C L band image PDDP and RPDDP analysis.

Evaluation criteria	**PDDP**	**RPDDP**
Hellinger distance	3.1455	5.3413
RMSSTD error - Black class	1.0589e-15	3.3128e-16
RMSSTD error - Gray class	5.8445e-15	5.4605e-16

SOURCE: Author's production.

In the Tapajós ALOS/PALSAR image analysis, the clusters difference between RPDDP and PDDP algorithm is more evident. The RPDDP clustered the Primary Forest, Agriculture 2, and Agriculture 3 in the Gray class, while the Agriculture 1, Bare Soil, and Pasture were clustered in the Black class. The PDDP clustered part of Primary Forest samples, Agriculture 2, and Agriculture 3 in the Gray class, while the other part of Primary Forest samples, Agriculture 1, Bare Soil, and Pasture were clustered in the Black class, as showed in Figure 6.10.

In the PDDP algorithm clustering, the major clusters confusion came with the forest samples. Visually the forest samples are brighter and it is present in more quantity than the non-forest samples. Since the arithmetic mean is the sum of a samples collection divided by the amount of these samples, the AvCM reflects a systematic sample size bias, meaning that the forest class has a bigger influence in the average covariance matrix, what make the forest class be splitted.

Due to the confusion generated by forest mechanism in the PDDP algorithm clustering, the PDDP segmented image, shown in Figure 6.10(b), has much more segments than the RPDDP segmented image, presented by Figure 6.10(a). In the same way, in the RPDDP clustered image, presented in Figure 6.10(c), the clusters borders are clearer than in PDDP clustered image, shown in Figure 6.10(d). Therefore, the RPDDP algorithm could fairly well separate the forest and non forest classes, while the PDDP struggle to do the same task.

Table 6.4 confirms the visual information. The RMSSTD error inner clusters, for Black or Gray cluster, are lower in RPDDP algorithm. While The stochastic distance between the cluster is bigger in RPDDP approach. For this reason, the clusters generated by RPDDP are more homogeneous and they have a better separability, than the PDDP clusters.

Figure 6.10 - RPDDP and PDDP algorithm applied on Tapajós ALOS/PALSAR image.

(a) RPDDP segmented image.

(b) PDDP segmented image.

(c) RPDDP clustered image.

(d) PDDP clustered image.

SOURCE: Author's production.

Table 6.4 - Tapajós ALOS/PALSAR image PDDP and RPDDP analysis.

Evaluation criteria	**PDDP**	**RPDDP**
Hellinger distance	0.28899	0.65475
RMSSTD error - Black class	3.5107e-10	3.8243e-11
RMSSTD error - Gray class	3.9766e-11	8.4702e-12

SOURCE: Author's production.

In ALOS/PALSAR San Francisco image the difference between the segmented images is more subtle. The RPDDP and PDDP algorithms results are very similar. The RPDDP clustered the Mountain, Low Density Urban, High Density Urban, Vegetation, and Develop Urban in the Gray class, while the Water were clustered in the Black class; the PDDP had a similar result to RPDDP clusters, excepted for the region labeled as Mountain, where part of the samples were clustered as water, and also some Vegetation samples were clustered in the water cluster.

The Mountain region in the image presents darker areas and brighter patches on the mountain slopes, which face in the radar illumination direction. In PDDP clustering, the mountains brighter patches were clustered with the urban region, which is characterized by a stronger and, therefore, brighter answer due to the double bounce mechanism presence in these regions. While the darker Mountain regions were clustered with the Water region. On the other hand, the RPDDP was able to fairly separate the water and non-water regions.

Therefore, the RPDDP segmented image, presented in 6.11(a) has less segments than the PDDP segmented image, presented in 6.11(b). What can be perceived also in Figures 6.11(c) and 6.11(d).

Table 6.5 provides the evaluation criteria results. As happened for the previous images, i. e., the RPDDP clustering has a bigger stochastic distance value between the Black and Gray clusters, than the distances between the PDDP Black and Gray clusters. The RMSSTD error for the Gray class is smaller with the RPDDP clustering. However, the RMSSTD error, for the Black cluster, is lower with the PDDP clustering than in RPDDP clustering.

Figure 6.11 - RPDDP and PDDP algorithm applied on ALOS/PALSAR San Francisco image.

(a) RPDDP segmented image.

(b) PDDP segmented image.

(c) RPDDP clustered image.

(d) PDDP clustered image.

SOURCE: Author's production.

Table 6.5 - San Francisco ALOS/PALSAR image PDDP and RPDDP analysis.

Evaluation criteria	**PDDP**	**RPDDP**
Hellinger distance	6.2626	6.7132
RMSSTD error - Black class	6.935e-10	1.1928e-09
RMSSTD error - Gray class	4.0379e-10	2.9415e-10

SOURCE: Author's production.

6.3 Experiment III - simulated PolSAR image classification

Section 5.2.3 describes the Monte Carlo simulation over 100 simulated PolSAR methodology. Nine different approaches are tested: SC-B, SC-H, SC-KL, BSC-R-B, BSC-R-H, BSC-R-KL, BSC-EM-B, BSC-EM-H, and BSC-EM-KL.

The output of Monte Carlo simulation is presented in Figure 6.12, which synthesizes simulated images classification accuracy. In this graph, on each blue box, the bottom and top edges indicate the 25th and 75th percentiles of the overall accuracy outcomes, the dotted line indicates the maximum and minimum accuracy result (excluding the outliers), the central red mark indicates the median accuracy, and the outliers are plotted individually using the '+' symbol in red.

The SC-B, SC-KL, and SC-H have a balanced accuracy results, with the first quartile and third quartile between 60% and 90%, which is in agreement with the resulted presented in Carvalho et al. (2019). The BSC-EM-B, BSC-EM-KL, and BSC-EM-H had a good result, having the first quartile and third quartile between 90% and 95%, however this approach had more outliers results than the others approaches. The BSC-R-B, BSC-R-KL, and BSC-R-H had the best accuracy results, having the first quartile and third quartile between 95% and 98%, this approach also had outliers accuracy.

Tables 6.6 to 6.7 presents the execution time for each algorithm analysed in this work. All computations were performed in a computer with Intel Core i7 2.4GHz processor, 8GB of RAM, and Matlab 2018. The SC algorithm had the best performance in terms of computational effort, what is expected, since this algorithm is a partitional method. The BSC variants had a much bigger computational time in comparison with SC algorithm. The BSC-EM variants were the most expensive technique, since two interactive algorithms are executed at each bisecting step: the EM, used to

estimate the initial parameters, and the SC used to do the bipartition.

Figure 6.12 - Monte Carlo simulation accuracy box plot.

SOURCE: Author's production.

Table 6.6 - Algorithms execution time (s), considering 5 iterations for SC.

	SC-B	**SC-H**	**SC-KL**
Time [s]	32.93	34.47	34.07

SOURCE: Author's production.

Table 6.7 - Algorithms execution time (s), considering 15 levels with 5 iterations each for BSC algorithm.

	BSC-EM-B	**BSC-EM-H**	**BSC-EM-KL**
Time [s]	163.82	162.95	164.24

SOURCE: Author's production.

Table 6.8 - Algorithms execution time (s), considering 15 levels with 5 iterations each for BSC algorithm.

	BSC-R-B	**BSC-R-H**	**BSC-R-KL**
Time [s]	115.12	114.56	125.32

SOURCE: Author's production.

In order to elucidate the this experiment, the simulated PolSAR image presented in Figure 5.8 were classified by the SC-B, SC-H, SC-KL, BSC-EM-B, BSC-EM-H, BSC-EM-KL, BSC-R-B, BSC-R-H, and BSC-R-KL. Figures 6.13, 6.14 and 6.15 present the classification results, and the correspondents confusion matrices and overall accuracy results.

The SC algorithm, using randomly chosen initial parameters presented the worst result. The SC-B classified image is presented in Figure 6.13(a), the SC-KL output is shown in Figure 6.13(c), and the SC-H result is presented in 6.13(e). In the three variants, SC-B, SC-KL, and SC-H, only 5 of the six classes were intensified. By the confusion matrices presented by Figures 6.13(b), 6.13(d), and 6.13(f), can be checked that the River was classified as Bare Soil, and there are a substantial mix between Corn 2 and Soy 1 classes. The accuracy results were very similar for the three variants. The SC-B had 61.66 % of overall classification accuracy, the SC-KL had 61.66 % of overall accuracy, and the SC-H achieved 61.59 % of overall accuracy.

The BSC-EM classification results are presented in Figures 6.14(a), 6.14(c), and 6.14(e), while the respective confusion matrices are presented in Figures 6.14(b), 6.14(d), and 6.14(f). This method was able to identify the six classes within the image. The River, Bare Soil, Caatinga, and Soy 3 segments were correctly classified in the three BSC-EM variants (BSC-EM-B, BSC-EM-KL, and BSC-EM-H), while the Corn 2 and Soy 1 were mixed. In Table 5.2 the smallest distance intra classes is between Corn 2 and Soy 1, having a Hellinger distance of 0.19, therefore the mixture between these classes is expected, specially when the segments of both classes share borders, as presented in Figure 5.8. The BSC-EM algorithm achieved a good accuracy result. The BSC-EM-B had 85.23 % of overall classification accuracy, the BSC-EM-KL had 81.68 % of overall accuracy, and the BSC-EM-H also achieved 83.87 % of overall accuracy.

The classification result of the BSC-R-B, BSC-R-KL, and BSC-R-H achied the best

overall accuracy in comparassion in the SC and the BSC-EM variants, as show in Figure 6.15. The BSC-R-B achieved 85.96 % of overall classification accuracy, the BSC-R-KL had 85.47 % of overall accuracy, and the BSC-R-H achieved 85.05 % of overall accuracy. The three BSC-R variants were able to correctly classify the River, Caatinga, Bare Soil, and Soy 3. As in BSC-EM, the classes Corn 2 and Soy 1 were mixed. However, it is interesting to notice that even when the initial parameter were selected in the best possible scenario S06 (Figure 6.7), the confusion between Corn 2 and Soy 1 was big.

The SC and BSC algorithms explored the Bhattacharyya Kullback-Leibler and Hellinger stochastic distances. The three distances achieved similar results. The Kullback-Leibler divergence and, therefore, the distance is usually associated to the relative entropy and information gain characterization. In Figure 6.12, this distance achieved a slightly worse accuracy result in average, in comparison with Bhattacharyya and Hellinger distances. The Bhattacharyya distance is widely used to evaluate class separability, it is an efficient tool for image segmentation and classification, and this distance achieved one of the best results in 6.12. The Hellinger distance had a very similar accuracy result to the Bhattacharyya distance, what can be explained by the close relationship between Bhattacharyya and Hellinger distance.

Figure 6.13 - Simulated PolSAR image classification using SC algorithm and randomly chosen initial parameter.

(a) SC-B.

(b) Confusion matrix - SC-B.

(c) SC-KL.

(d) Confusion matrix - SC-KL.

(e) SC-H.

(f) Confusion matrix - SC-H.

SOURCE: Author's production.

Figure 6.14 - Simulated PolSAR image classification using EM and BSC algorithm.

(a) BSC-EM-B.

(b) Confusion matrix - BSC-EM-B.

(c) BSC-EM-KL.

(d) Confusion matrix - BSC-EM-KL.

(e) BSC-EM-H.

(f) Confusion matrix - BSC-EM-H.

SOURCE: Author's production.

Figure 6.15 - Simulated PolSAR image classification using RPDDP and BSC algorithm.

(a) BSC-R-B.

(b) Confusion matrix - BSC-R-B.

(c) BSC-R-KL.

(d) Confusion matrix - BSC-R-KL.

(e) BSC-R-H.

(f) Confusion matrix - BSC-R-H.

SOURCE: Author's production.

6.4 Experiment IV - real PolSAR image classification

In Section 6.3 the SC-B, SC-H, SC-KL, BSC-R-B, BSC-R-H, BSC-R-KL, BSC-EM-B, BSC-EM-H, and BSC-EM-KL were evaluated in a quantitative way, using simulated PolSAR images, in order to derive the confusion matrices and the accuracy values.

As discussed in Section 6.3, the three analysed stochastic distances achieved similarly accuracy results, and according to Frery et al. (2013) the Hellinger distance is the best option to handle Wishart distributed data. Therefore in order to avoid analyzing repeated or very similar results, the Hellinger distance were chosen to perform the real PolSAR classification analysis. In next sections the algorithm BSC-R-H will be analysed in qualitative and quantitative way. The SC-H, and BSC-EM-H classification results will be analysed in a quantitative way only.

6.4.1 Real PolSAR image qualitative analysis

In this section, the qualitative analysis using the BSC dendrogram and Plan $H - \alpha$ is explored. The classification results of the BSC-R-H algorithm for the images Bebedouro SIR-C in C band, Bebedouro SIR-C in L band, Tapajós ALOS/PALSAR and San Francisco ALOS/PALSAR are addressed.

6.4.1.1 Bebedouro SIR-C in C band image qualitative analysis

The first analysed image is the Bebedouro SIR-C in C band, showed in Figure 5.1(b). As presented in Figure 5.1(c), the Bebedouro region has none identified classes: Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, Tillage, Bare Soil, Caatinga, and River.

Figure 6.16 shows the Plan $H - \alpha$ scattering distribution of all Bebedouro region classes. As discussed in Section 5.1.1.1, Bebedouro SIR-C in C band should have a lot of confusion inside the Vegetation group, due to the backscattering pattern generated by the C band frequency. In this band, the radar wave interacts much more with the vegetation canopy, leaves and branches, what means that the scattering mechanism are mainly volumetric, dipole, or anisotropic, and the entropy is moderate.

In Figure 6.16 the classes Corn 1, Corn 2, Soy 2, Soy 3, Tillage, and Caatinga occupy basically the same regions, z5 and z6. Soy 1 also occupies the z5 and z6 regions, however its samples are more spread. As discussed in Section 5.4.2, z5 zone is related with anisotropic particles, where the α angle scattering mechanisms is the volumetric, and the entropy is moderated. The z6 zone gathers random surface

elements, were the α angle mechanism is the surface scattering and the entropy is also moderated.

Figure 6.16 - Bebedouro SIR-C in C band Plan $H - \alpha$ of all identified classes.

SOURCE: Author's production.

The Bare Soil classes samples are mainly located in region z9. This region is characterized by the Bragg surface scattering. The α angle is represented by the surface

mechanism and the entropy is low. The River samples are spread between z9 and z6, therefore the samples are a mix of Bragg surface and random surface elements.

Figure 6.17 shows the Bebedouro SIR-C in C band classified image. On the right side of the classified image, the outcome legend is shown.

Figure 6.17 - BSC-R-H Bebedouro SIR-C in C band classification result.

SOURCE: Author's production.

Figure 5.1(c) shows that the Bebedouro SIR-C image has nine identified classes: Corn 1, Corn 2, Soy 1, Soy 2, soy3, Tillage, Bare Soil, Caatinga and River. However the BSC-R-H was able to identify mainly three classes: River, Bare Soil, and Vegetation, which includes the classes Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, Tillage, and Caatinga. The Vegetation class was splitted in many unknown classes, the clusters are actually named as New 1, New 2, New 3, New 4, New 5, New 6, and New 7.

The clustering history can be followed in Figure 6.18, which presents the Bebedouro SIR-C in C band clustering steps. The dendrogram grows in a top bottom way until the stop criteria is reached, as discussed in Section 4.2.

Every node in the dendrogram has a cluster associated to it, therefore hereafter the

clusters will be addressed by the node ID. For instance, the cluster associated to node ID: 1 will be named 'cluster ID1'. Besides that, an important clue presented by the dendrogram is the information gain (I) between the parent node and its two children. Using the I information, it is possible to follow the entropy decrease between the bipartions.

Figure 6.18 - Bebedouro SIR-C in C band BSC-R-H dendrogram.

SOURCE: Author's production.

In Figure 6.18, the cluster ID1 is splitted into cluster ID2 and the cluster ID3. The cluster ID2 gathers the Vegetation and Tillage samples, while the cluster ID3 groups the Bare Soil and River classes. The biggest information gain comes with cluster ID3, which has $I = 20.12$. This high value is associated to cluster ID3 because the Bebedouro SIR-C in C band image has much more Vegetation samples than River or Bare Soil. Therefore cluster ID3 diverges much more from cluster ID1, than node ID2, whose information gain is $I = 6.28$.

The difference between cluster ID2 and cluster ID3 can be checked in Figure 6.19,

which represents the cluster ID1 splitting. This figure has two levels: the parent level, identified as Level 1, and the children level, identified as Level 2. In the Level 1, the middle picture represents the Bebedouro SIR-C in C band PolSAR image, and on this picture side there are the Plans $H - \alpha$ representation. The upper right Plan $H - \alpha$ plots the cluster ID1 samples, while the Plan $H - \alpha$ on the upper left image represents the average entropy and α angles of cluster ID1, cluster ID2, and cluster ID3. On Level 2, the cluster ID2 and cluster ID3, and their respective Plans $H - \alpha$ are presented.

The Plan $H - \alpha$ is represented by heat map, were the red spots are the regions with high sample concentration. The cluster ID1 has a high amount of vegetation samples, in comparison with the River and Bare soil samples. As already discussed, the vegetation class (Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, and Caatinga) are mainly concentrated in z5 and z9 on the Plan $H - \alpha$, therefore the cluster ID1 samples concentration in z5 and z6 is expected. In the cluster ID2 the samples are even more concentrated on the z5 and z6 Plan $H - \alpha$ regions. On the other hand, the Bare Soil and River samples are spread between z5, z6, and z9. The Plan $H - \alpha$ scheme presented in Figure 6.19 can be generated for every node, allowing to the user to do a deeper analysis on the clusters splitting and about their scattering mechanism.

As discussed in Section 5.4.1, in order to perceive which is the next node in the splitting queue, the user must follow the node ID order. Therefore in Figure 6.18, the next splitted cluster is cluster ID3, which is divided between River (cluster ID4) and Bare Soil (cluster ID5). It is interesting to note that the Bare Soil information gain ($I = 10.15$) is much higher than the River (1.80). This happens because the cluster ID3 has much more River samples than Bare Soil, as shown in Figure 6.19, therefore the cluster ID4 (River) is more similar to its parent, the cluster ID3. Therefore, the entropy difference between these two cluster is smaller than between cluster ID3 and cluster ID5.

The next splitted cluster is the cluster ID2, where all vegetation classes are grouped. On this side of the tree, the splittings are hard to follow, since cluster borders are not well defined. For this reason the cluster are label as 'New'.

Figure 6.20 summarizes the Hellinger distances between the classes clustered by BSC-R-H. As presented by Figure 6.17 and Figure 6.18, the BSC-R-H identified 9 classes, which are: New 1, New 2, New 3, New 4, New 5, New 6, New 7, River and Bare Soil.

Figure 6.19 - Bebedouro SIR-C image in C band BSC-R-H dendrogram. Node ID 1 and its children.

SOURCE: Author's production.

The farthest classes are River and New 1, with a Hellinger distance of 10.46. Analysing Figure 6.18, one can see that these clusters (cluster ID12 and cluster ID4) are also far way from each other. This happens because the dendrogram leaf location gives a clue about the dissimilarity between classes. As far one cluster is from another, the more dissimilar its samples are from the other cluster samples.

On the other hand, the cluster New 2 and New 3 have the smallest distance value, i. e., 0.14, and, in Figure 6.18, they share the same parent cluster. The second smaller distance, is between classes New 5 and New 6, which are neighbors clusters (cluster ID17 and cluster ID8), even though these clusters are from different parents. The clusters on the left branch are similar between them, therefore the distances are low.

Figure 6.20 - Hellinger distances between the classes clustered by BSC-R-H using Bebedouro SIR-C image in C band as input image.

	New 1	New 2	New 3	New 4	New 5	New 6	New 7	River	Bare Soil
New 1	0.00	0.32	0.33	0.74	0.47	1.03	2.71	10.46	6.83
New 2		0.00	0.14	0.30	0.26	0.68	1.77	9.39	5.72
New 3			0.00	0.23	0.30	0.54	1.64	9.03	5.43
New 4				0.00	0.23	0.22	0.94	7.52	4.23
New 5					0.00	0.21	1.24	8.08	4.69
New 6						0.00	0.69	6.58	3.52
New 7							0.00	4.66	1.68
River								0.00	1.34
Bare Soil									0.00

SOURCE: Author's production.

Inside the Vegetation branches, the classes New 7 and New 1 have a distance of 2.71, which is the highest within these groups. This distance can be also checked in Figure 6.18, where the leaf cluster ID12 and cluster ID9 are far from each other. The class New 7 is, actually, close to River and Bare Soil, and this class, by the distance table, is the vegetation group class with the smaller distance to River (4.66) and Bare Soil (1.68). The River and Bare Soil in dendrogram are neighbor class, and the River is closest class to Bare Soil with a distance of 1.34.

6.4.1.2 Bebedouro SIR-C in L band image qualitative analysis

The second analysed image is the Bebedouro SIR-C in L band, presented in Figure 5.1(a). As for Bebedouro SIR-C C band image, the first step of the qualitative analysis is to inspect the entropy and α angles of each class presented in Figure 5.1(c), videlicet, Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, Caatinga, Tillage, Bare Soil, and River. Figure 6.21 shows the Plan $H - \alpha$ of classes.

Different from C band, the L band waves penetrates the vegetation, and interact with the ground, vegetation stem, branches and leaves. This behavior allows the L band sensor to measure the vegetation vertical structure, consequently, the vegetation area can have information about the three majors scattering mechanism: surface, double-bounce and volumetric.

Due to that, a latent difference between Bebedouro SIR-C in C band and in L band is the Caatinga scattering mechanisms location. While in C band (Figure 6.16), Caatinga is located between z5 and z6 zones, in L band, the Caatinga class is mainly located in z2. The z2 zone is where high entropy volumetric scattering arises. Example of objects with this kind of entropy and α angles are the forest canopies.

Another difference between C and L band scattering mechanism is with the River samples. In C band the River samples were splitted between z6 and z9. However, in L band, the River samples are heavily located in z9, where mechanism with low entropy and surface scatterer type are located. The Bare Soil samples scattering mechanism are spread between z9 and z6 for both SIR-C in C and L band.

The Tillage scattering mechanisms samples location also differs in C and L bands. In C band the samples are spread between z5 and z6, but very concentrated. In L band. this class samples are spread between zones z5, z6,and z9.

The Corn 1, Corn 2, Soy 1, Soy 2, and Soy 3 samples are mainly grouped at z5, which is a moderate entropy region with a dominant dipole type of scattering mechanism. This zone would include scattering from vegetated surfaces with anisotropic scatterers.

Figure 6.22 shows the Bebedouro SIR-C in L band (Figure 5.1(a)) classified image. The BSC-R-H were able to correctly identify the Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, Tillage, Caatinga, River, and Bare Soil classes described in Figure 5.1(c). In addition, the algorithm created two new classes, named 'New 1' and 'New 2'.

Figure 6.21 - Bebedouro SIR-C in L band Plan $H - \alpha$ of all identified classes.

SOURCE: Author's production.

The clustering history is described in Figure 6.23. In the first level, the cluster ID1 groups all Bebedouro SIR-C in L band image samples, and, therefore, it gathers the classes Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, Tillage, Caatinga, Bare Soil, and River. The cluster ID1 is splitted into cluster ID2 and cluster ID3. The cluster ID2 groups the Corn 1, Corn 2, Soy 1, Soy 2, Soy3, and Caatinga, and the information gain is $I = 7.29$. The cluster ID3 groups the classes Bare Soil, River, and Tillage, and has

an information gain of $I = 14.14$.

Figure 6.22 - BSC-R-H Bebedouro SIR-C in L band classification result.

SOURCE: Author's production.

Note that the information gain value for cluster ID3 in C band are higher than in L band. This happens because the cluster ID3 in C band groups Bare Soil and River, while cluster ID3 in L band groups, besides Bare Soil and River, the Tillage class, therefore the cluster ID3 in L band is more similar to its parent cluster, than the cluster ID3 in C band.

Figure 6.24 extrapolates the cluster ID1 splitting. The upper right picture shows the Plan $H - \alpha$ of the Bebedouro SIR-C in L band, in this Plan the samples, although spread over z1, z2, z4, z5, z6, z7, z8, and z9, are highly concentrated in z5 zone, because the amount of vegetation samples is high on this image. In z9 there is a moderated concentration of samples, since the River class also has a good amount of samples.

Figure 6.23 - Bebedouro SIR-C image in L band BSC-R-H dendrogram.

SOURCE: Author's production.

Figure 6.24 - Bebedouro SIR-C image in L band BSC-R-H dendrogram. Node ID 1 and its children.

SOURCE: Author's production.

On level 2, the cluster ID2 Plan $H-\alpha$ is similar to its parent Plan $H-\alpha$, except for the z9 concentration, which disappeared. This similarity can be checked also by the information gain value ($I = 7.29$), which is lower, in comparison with cluster ID3 information gain ($I = 14.14$). On the other hand, the cluster ID3 Plan $H-\alpha$ has a high concentration on z9 region, because the River samples are more representative in this cluster.

Getting back to Figure 6.23, the next splitted cluster is the cluster ID3, which is divided between cluster ID4 and cluster ID5. The cluster ID4 groups the Tillage samples. The cluster ID5 groups the Bare Soil and River samples, which by the distance table, presented in Figure 5.2(b), are closer classes, with a distance equal to 0.58. While the distance from River to Tillage is 3.20, and the from Bare soil to Tillage the distance is 2.35.

The cluster ID2 is the next in the queue. It is divided into cluster ID6 and cluster ID7. Cluster ID6 groups the Corn1 and Caatinga, while cluster ID7 groups Corn 2, Soy 1, Soy 2 and Soy 3. Again, by the distance table, presented in Figure 5.2(b), the Corn 1 is the closest class to Caatinga, with a distance of 1.14. And because the amount of Caatinga samples is higher in cluster ID2 and in cluster ID6, the information gain ($I = 2.62$) between these cluster is lower, than between cluster ID2 and ID7 ($I = 4.82$).

The splitting procedure continues until the stop criteria is reached. The last splitted cluster is the cluster ID17, which is divided between Soy 2 and Soy 3. These classes have one of the lowest distance in Figure 5.2(b), being only 0.33. In addition, their Plan $H-\alpha$, presented in Figure 6.21 are very similar, and, finally, these classes are spatially neighbors, meaning that their borders are a mix of the two classes.

Figure 6.25 presents the Hellinger distances between the classes New 1, Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, Tillage, Bare Soil, Caatinga, New 2, and River. The closest classes are the Soy 3 and Corn 2, which distance value is 0.26. These two classes are very close in the Figure 6.23, as well. The Soy 3 is also very similar to New 2, with a distance of 0.27, and these classes also share the same parent, as presented in Figure 6.23. The most dissimilar classes, according to Figure 6.25, are the Caatinga and River, which clusters (cluster ID13 and cluster ID10) are far way in BSC-R-H dendrogram. The second unknown class, the New 1, is close to Corn 1. Although New 1 and Corn 1 are from different parents, they are very close in the BSC-R-H dendrogram.

Figure 6.25 - Hellinger distances between the classes clustered by BSC-R-H using Bebedouro PolSAR image in L band as input image.

	New 1	Corn 1	Corn 2	Soy 1	Soy 2	Soy 3	Tillage	Bare Soil	Caatinga	New 2	River
New 1	0.00	0.55	2.16	2.50	1.45	1.43	4.62	8.59	0.77	1.00	10.45
Corn 1		0.00	0.97	1.36	0.55	0.62	2.77	6.45	1.40	0.53	8.26
Corn 2			0.00	0.55	0.50	0.26	1.22	4.56	2.42	0.48	6.44
Soy 1				0.00	0.34	0.96	0.84	3.50	3.22	0.78	4.97
Soy 2					0.00	0.59	1.48	4.58	2.25	0.43	6.17
Soy 3						0.00	2.17	5.95	1.52	0.27	7.88
Tillage							0.00	1.66	5.63	2.33	3.25
Bare Soil								0.00	9.86	5.94	0.79
Caatinga									0.00	1.11	11.84
New 2										0.00	7.82
River											0.00

SOURCE: Author's production.

6.4.1.3 Bebedouro SIR-C in C+L band image qualitative analysis

In this section, the multifrequency PolSAR classification is addressed. The Bebedouro PolSAR images in C and L bands are combined in order to evaluate the BSC-R-H when using PolSAR image from different frequencies. As already discussed, the focus of this work is on PolSAR images, meaning that the dataset is formed by a group of covariance matrices $Z = \{\boldsymbol{Z_1}, \ldots, \boldsymbol{Z_n}\}$, with N samples and L number of looks, where $\boldsymbol{Z_n}$ is given by Equation (2.21).

In order to generate a multifrequency covariance matrix, the C band covariance matrices, named $\boldsymbol{Z_C}$, and the L band covariance matrices, named $\boldsymbol{Z_L}$ shall be combined in a unique matrix, as displayed in Figure 6.26. Considering $\boldsymbol{Z_L}$ and $\boldsymbol{Z_C}$ as a 3×3, merged covariance matrix will be 6×6.

Figure 6.27 shows the Bebedouro SIR-C multifrequency dataset classification result. The BSC-R-H identified the classes Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, New 1, New 2, New 3, Bare Soil, Caatinga, and River. The multifrequency had a better classification than the Bebedouro SIR-C in C band, but worse that Bebedouro SIR-C in L band.

The clustering history is described in Figure 6.28. In the first level, the cluster ID1, which groups all Bebedouro SIR-C in C and L band image samples, is divided between cluster ID2 and cluster ID3. The cluster ID2 groups the Corn 1, Corn 2,

Soy 1, Soy 2, Soy 3, Tillage, and Caatinga, while the cluster ID3 groups the Bare Soil and River. Note that the information gain is much higher for the multifrequency cluster, than in L or C band. The splitting procedure evolve until the stop criteria is reached.

Figure 6.26 - Multifrequency covariance matrix.

$$z_{cL} = \begin{bmatrix} Z_C & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & Z_L \end{bmatrix}$$

SOURCE: Author's production.

Figure 6.27 - BSC-R-H Bebedouro SIR-C multifrequency classification result.

SOURCE: Author's production.

Figure 6.28 - Bebedouro SIR-C multifrequency BSC-R-H dendrogram.

SOURCE: Author's production.

Figure 6.29 presents the Hellinger distances between the classes Corn 1, Corn 2, Soy 1, Soy 2, Soy 3, New 1, New 2, New 3, Bare Soil, Caatinga, and River. The closest classes are the New 1 and New 2, looking at Figure 6.27, these two class seams to be part of Caatinga region. New 3 is the closest to Corn 1, and by Figure 6.27 and Figure 6.28, these classes are correlationated. The farthest classes are Caatinga and River with a distance equal to 15.78.

Figure 6.29 - Hellinger distances between the classes clustered by BSC-R-H using Bebedouro PolSAR multifrequency.

	Corn 1	Corn 2	Soy 1	Soy 2	Soy 3	New 1	New 2	New 3	Bare Soil	Caatinga	River
Corn 1	0.00	0.84	3.24	0.84	0.87	0.53	1.04	0.71	5.96	1.44	11.17
Corn 2		0.00	2.07	0.71	0.40	0.38	0.50	1.96	5.51	1.87	11.20
Soy 1			0.00	2.83	2.05	3.10	3.66	5.61	1.74	6.28	6.22
Soy 2				0.00	0.49	0.48	0.92	2.05	5.44	1.66	10.94
Soy 3					0.00	0.42	0.68	2.25	4.73	2.01	10.06
New 1						0.00	0.28	1.47	6.10	0.99	11.65
New 2							0.00	1.66	7.13	0.78	12.93
New 3								0.00	9.18	1.32	14.77
Bare Soil									0.00	9.89	2.68
Caatinga										0.00	15.78
River											0.00

SOURCE: Author's production.

6.4.1.4 Tapajós ALOS/PALSAR image qualitative analysis

The next analysed image is the Tapajós ALOS/PALSAR, which is shown in Figure 5.4(a). As presented in Section 5.1.1.2, the ALOS/PALSAR is a L band sensor, that can operate in polarimetric mode, where the complete polarization scheme is offered. The Tapajós PolSAR image is the polarimetric representation of Tapajós forest. Within this image, besides the primary forest samples, there are agricultural, pasture, and different types of forests, among others samples. As presented by Figure 5.4(b), the Tapajós region has six identified classes: Bare Soil, Agriculture 1, Agriculture 2, Agriculture 3, Primary Forest, and Pasture.

As already discussed, the wavelength band strongly affects what type of objects the SAR is sensitive to. For instance, in forest areas, longer wavelength radar signals, such as L-band, penetrate through the forest canopy and interact with the

larger structures such as the trunks and larger trees branches. All these interactions means that the every polarization combinations (S_{hh}, S_{hv}, and S_{vv}) will have a strong answer, consequently the entropy is high in forest scenarios. Since the Tapajós ALOS/PALSAR is in L band and this image is formed mostly by forest samples, the Tapajós ALOS/PALSAR Plan $H - \alpha$ will have samples mainly in zones with high entropy, i. e., z1 and z2.

Figure 6.30 shows the Plans $H - \alpha$ for the six above cited classes. As expected, the Primary Forest scattering mechanism samples are highly concentrated in z2, which is a high entropy and volumetric scattering mechanism region. The Bare Soil, Agriculture 1, and Pasture samples are mainly located in z5 and z6. While the Agriculture 2 and Agriculture 3 scattering mechanism samples are in z6 and z9.

Figure 6.30 - Tapajós ALOS/PALSAR Plan $H - \alpha$ of all identified classes.

SOURCE: Author's production.

Figure 6.31 presents the classified image. On its right side the outcome legend is

presented. Even though the Tapajós region has six identified class, as presented in Figure 5.4(b), the BSC-R-H generated ten classes, as showed by the outcome legend. The classes are Bare Soil, Agriculture 1, Agriculture 2, Agriculture 3, Primary Forest, Pasture, New 1, New 2, New 3, and New 4.

Figure 6.31 - BSC-R-H Tapajós ALOS/PALSAR classification result.

SOURCE: Author's production.

Figure 6.32 shows the clustering dendrogram. In the first level, the cluster ID1 is divided into cluster ID2 and cluster ID3. The cluster ID2 groups the Primary Forest and Agriculture 2. It is interesting to note that Agriculture 2 is not the closest to Primary Forest, their distance is 3.24, according to distance Table presented in Figure 5.5(b), against the 2.69 distance values between Primary Forest and Agriculture 3, for instance. However the Agriculture 2 is the farthest classes from the others, for example, the distance between Agriculture 2 and Bare Soil is 4.75, and the distance between Agriculture 2 and Agriculture 1 is 7.29. The cluster ID3 gathers the remained classes, i. e., Bare Soil, Agriculture 1, Agriculture 3, and Pasture.

Figure 6.32 - Tapajós ALOS/PALSAR image BSC-R-H dendrogram.

SOURCE: Author's production.

Figure 6.33 extrapolates the cluster ID1 splitting. This clustering division separates mainly forest and non-forest samples. On Level 1, the cluster ID1 groups all samples available in Tapajós ALOS/PALSAR image. The upper right image shows the cluster ID1 Plan $H-\alpha$, although the samples are spread over all zones in this plan, they are mainly located in z1, z2, and z5. This is due to the high amount of forest samples in cluster ID1, which pushes the entropy and α angle frequency onto these classes.

On level 2, the Plan $H-\alpha$ associated to cluster ID2 has scattering mechanism samples largely located in z1 and z2, which usually are zones associated to forest scatterings. The cluster ID3 Plan $H-\alpha$ has samples spread principally between z5 and z6, which are zones associated to less developed and more sparse vegetation zones.

The cluster ID2 information gain is $I = 1.66$, while the cluster ID3 is $I = 5.94$. Since the amount of forest sample is bigger in cluster ID1, and the cluster ID2 groups the forest samples, this group is more similar to its parent (cluster ID1), therefore the information gain is lower. On the other hand, cluster ID2 groups non-forest samples, which are less representative in this image, in terms of samples amount, resulting in a more dissimilar image in comparison with its parent, therefore the information gain is higher between cluster ID3 and cluster ID1.

The next splitted cluster in Figure 6.32 is the cluster ID3, which is divided between cluster ID4 and cluster ID5. The cluster ID4 groups the Agriculture 3 and Pasture, and the cluster ID5 groups the Bare Soil and Agriculture 1. The Bare Soil and Agriculture 1 have similar Plan $H-\alpha$ samples spreading pattern (Figure 6.30), and the distance between them is only 0.9, which is one of the smallest distances in Figure 5.5(b). The Pasture and Agriculture 3 have a distance value equals to 2.07, and their scattering mechanisms are in different zones, while the Pasture are mainly in z6, the Agriculture 3 samples are spread over z9 region.

The next splitted cluster is the cluster ID5 which is divided between cluster ID7, that represents a new class, labeled New 4, and cluster ID 6. Then, the cluster ID6 is finally divided between Agriculture 1 and Bare Soil. After that the cluster ID4 is divided between cluster ID10 and cluster ID11. The cluster ID10 splits into the class Pasture and New 2 class. The splitting procedure continues until the stop criteria is reached. The last divided cluster is the ID12, which groups the Primary Forest and Agriculture 1.

Figure 6.34 presents the Hellinger distances between the classes Bare Soil, Agricul-

Figure 6.33 - Tapajós ALOS/PALSAR image BSC-R-H dendrogram. Node ID 1 and its children.

SOURCE: Author's production.

ture 1, Agriculture 2, Agriculture 3, Primary Forest, New 1, Pasture, New , New 3, and New4. The closest classes are the New 1 and Primary Forest, with a distance of 0.25. These two classes are very close in the Figure 6.32, and they share the same parent, in Figure 6.31, the class New 1, seams to be part of forest.

The class New 2 is close to Pasture, their distance is 0.42, and in the dendrogram, these class share the same parent. The class New 3 is is distant 0.40 from class Agriculture 1, which are close classes in the dendrogram, although they have different parents. New 4 class is very close to New 2, with a distance of 0.30, and also the Bare Soil class, with a distance of 0.40. Bare Soil and New 4 share the same parent, however, New 4 and New 2 are not that close, having as common cluster the cluster the cluster ID3.

Figure 6.34 - Hellinger distances between the classes clustered by BSC-R-H using Tapajós ALOS/PALSAR image as input image.

	Bare Soil	Agriculture 1	Agriculture 2	Agriculture 3	Primary Forest	New 1	Pasture	New 2	New 3	New 4
Bare Soil	0.00	0.42	6.69	2.68	3.53	2.19	1.39	0.96	0.53	0.40
Agriculture 1		0.00	7.49	3.34	4.56	3.09	2.19	1.50	0.40	0.76
Agriculture 2			0.00	1.86	3.24	3.88	3.74	3.95	8.47	5.35
Agriculture 3				0.00	1.98	1.86	1.13	1.03	3.92	1.91
Primary Forest					0.00	0.25	0.85	1.53	5.69	2.16
New 1						0.00	0.36	0.86	4.11	1.11
Pasture							0.00	0.42	2.97	0.56
New 2								0.00	2.32	0.30
New 3									0.00	1.44
New 4										0.00

SOURCE: Author's production.

6.4.1.5 San Francisco ALOS/PALSAR image qualitative analysis

The last analysed image is the San Francisco ALOS/PALSAR image. This image is divided into six classes (Mountain, Water, Vegetation, High Density Urban, Low Density Urban, and Developed Urban), as presented in Figure 5.6(b).

Figure 6.35 shows the Plan $H-\alpha$ of each above cited classes. The Mountain samples are spread between z2, z5, z6, z9. In San Francisco image, the mountain is one of the most complex regions. According to Xiang et al. (2016), the Mountain class has

urban areas, lakes, pasture, forest, grass land, cultivated crops, and many others scenarios, therefore the different scattering mechanisms and diverse entropy values in this region is understandable.

Figure 6.35 - San Francisco ALOS/PALSAR Plan $H - \alpha$ of all identified classes.

SOURCE: Author's production.

The Low Density Urban class has a moderate to high entropy, and it is located between z1, z2, z5, and z6. The High Density Urban class samples are concentrated in z5, and spread over z4, z5, z6, z7, z8, and z9. Note that the z1 and z2, which are zones known for having forest scattering mechanisms, have no samples from High Density Urban class. The water class is completely in z9, while the Vegetation is spread over z1, z2, z5 and z6. The Develop Urban is concentrated on z5.

Figure 6.31 shows the BSC-R-H classification result. The classified image has thirteen classes, eight more than what Figure 5.6(b) shows. Only the Water class was splitted into three classes (Water, New 1, and New 2), for instance. The classes High Density Urban, Low Density Urban, and Developed Urban will be addressed by their initial hereafter, i. e., HDU, LDU, and DU, respectively.

Figure 6.36 - BSC-R-H San Francisco ALOS/PALSAR classification result.

SOURCE: Author's production.

Figure 6.37 shows the San Francisco clustering dendrogram. The cluster ID1 groups all San Francisco ALOS/PALSAR samples, and it is divided between cluster ID2 and cluster ID3. This clustering split divided the San Francisco ALOS/PALSAR samples between water and non-water, for instance, cluster ID2 gathers the Mountain, LDU, HDU, Vegetation, and DU, whilhe cluster ID 3 groups the Water samples.

Figure 6.38 shows the Level 1 splitting procedure. On Level 1, the Parent image represents the complete San Francisco ALOS/PALSAR dataset, the upper right

image shows the Plan $H - \alpha$, in this figure the samples are highly concentrated in z9 region, due to the high amount of Water samples in the cluster ID1. On Level 2, the cluster ID2 image shows the non-water samples, and, on its left, the Plan $H - \alpha$ shows a high concentration of samples in z5 and z2. The cluster ID3 image shows the water samples, and its Plan $H - \alpha$ has a high concentration in z9, which is a zone characterized by low entropy surface mechanisms.

The cluster ID2 information gain is $I = 12.70$, meaning that this cluster is more dissimilar to its parent, therefore the reduction in entropy in high for this cluster. The cluster ID3 information gain is $I = 6.15$, which isn't such a low value, however is lower than the cluster ID2 information gain. Since the amount of water samples is superior in San Francisco ALOS/PALSAR dataset, the dissimilarity is lower between cluster ID3 and cluster ID1.

The next splitted cluster is the cluster ID2, which is divided between cluster ID4, which groups the three urban areas, i. e., LDU, HDU, DU, and cluster ID5, which groups the Mountain and Vegetation. The splitting procedure continues until the sop criteria is reached. Again, following the dendrogram growth, presented in Figure 6.37, one can see that the cluster ID3, correspondent to the Water class, is splitted into cluster ID12 and cluster ID13, and further the cluster ID13 is splitted between cluster ID14 and cluster ID15.

According to Figure 5.6(b), the water class shouldn't be divided, however as presented by 6.39, on level 6, the samples associated to cluster ID12 are mainly located in z9, in the Plan $H - \alpha$, while the cluster ID13 samples are mainly located at z6, this is a strong indicative that, indeed, cluster ID12 and cluster ID13 are different cluster.

Figure 6.40 presents the Hellinger distances between the classes Mountain, LDU, HDU, Water, New 1, New 2, Vegetation, New 3, DU, New 4, New 5, New 6, and New 7. The class New 1 and New 2 share the same parent and are very close with a distance of 0.10. Even though the New 1, New 2, and the Water are grouped by cluster ID3, the class New 1 is closer to New 3, having 0.12 of distance value, and LDU having a distance of 0.18. The same pattern is followed by New 2, which is closer to LDU, having 0.08 of distance. The New 4 is closer to DU class. The New 5 is closer to DU and New 3. New 6 is closer to LDU, and New 7 is closer to LDU.

Figure 6.37 - San Francisco PolSAR image BSC-R-H dendrogram.

SOURCE: Author's production.

Figure 6.38 - San Francisco PolSAR image BSC-R-H dendrogram - Level 1.

SOURCE: Author's production.

Figure 6.39 - San Francisco ALOS/PALSAR water splitting.

SOURCE: Author's production.

Figure 6.40 - Hellinger distances between the classes clustered by BSC-R-H using Tapajós PolSAR image as input image.

	Mountain	LDU	HDU	Water	New 1	New 2	Vegetation	New 3	DU	New 4	New 5	New 6	New 7
Mountain	0.00	0.26	2.31	0.04	0.47	0.19	5.00	0.41	3.03	4.63	2.25	0.29	0.27
LDU		0.00	1.57	0.20	0.18	0.08	3.99	0.34	2.00	3.52	1.54	0.08	0.16
HDU			0.00	2.25	1.96	1.98	1.43	2.05	0.75	1.66	1.53	1.77	1.81
Water				0.00	0.40	0.15	4.81	0.39	2.81	4.40	2.08	0.23	0.25
New 1					0.00	0.10	4.70	0.12	2.23	3.98	1.20	0.12	0.16
New 2						0.00	4.60	0.19	2.38	4.01	1.58	0.10	0.15
Vegetation							0.00	5.17	1.18	0.81	3.93	4.49	4.69
New 3								0.00	2.56	4.46	1.20	0.19	0.10
DU									0.00	0.52	1.30	2.24	2.36
New 4										0.00	2.93	3.92	4.12
New 5											0.00	1.34	1.42
New 6												0.00	0.09
New 7													0.00

SOURCE: Author's production.

6.4.2 Real PolSAR data quantitative analysis

This section address the quantitative analysis of the classified images. The confusion matrices and the accuracy results of Bebedouro SIR-C images in C and L band, Tapajós ALOS/PALSAR and San Francisco ALOS/PALSAR images using the SC-H, BSC-R-H, and BSC-EM-H algorithms are presented.

6.4.2.1 Bebedouro SIR-C in C band image quantitative analysis

The first analysed image is the Bebedouro SIR-C in C band. Figure 6.41(a) presents the classification result of SC-H algorithm. This algorithm selected the initial parameter randomly and the number of cluster were nine. The algorithm were fairly able to identify five classes, and the four remained received just few samples, being classified as noise. Figure 6.41(b) shows the confusion matrix and the classes samples distribution. The Corn 1 were spread over all image, and just a small parcel of the samples were correctly classified. The Soy 1 were correctly classified, however it grouped the Caatinga samples together in it. The Bare Soil and River were correctly classified. The overall accuracy of SC-H algorithm were 45.17 %

Figure 6.42(a) and Figure 6.42(b) present the classification result of BSC-EM-H algorithm and their confusion matrix, respectively. The BSC-EM-H algorithm gets the initial parameter using the EM algorithm. The classification overall accuracy of this algorithm were much worse than the SC-H, it achieved only 19.6 %. The BSC-

Figure 6.41 - Bebedouro SIR-C in C band image classification using the SC-H algorithm.

(a) Classified Image. (b) Confusion matrix.

SOURCE: Author's production.

Figure 6.42 - Bebedouro SIR-C in C band image classification using the BSC-EM-H algorithm.

(a) Classified Image. (b) Confusion matrix.

SOURCE: Author's production.

Figure 6.43 - Bebedouro SIR-C in C band image classification using the BSC-R-H algorithm.

(a) Classified Image. (b) Confusion matrix.

SOURCE: Author's production.

EM-H were able to correctly identify the Corn 1 class, and the Bare Soil, but all other class were miss-classified. The worst case is the River, which were completely classified as Bare Soil. The Soy 2, Corn 1, and Corn 2 classes were spread over the whole image, therefore the Caatinga was classified as a mix of Corn 2 and Soy 2.

Figure 6.43(a) and Figure 6.43(b) present the classification result of BSC-R-H algorithm and its confusion matrix, respectively. This algorithm achieved a better result in comparison with SC-H and BSC-EM-H. It had 48.54 % of overall accuracy. The BSC-R-H correctly classified the Bare Soil and River. There is a confusion between Corn 1 and Corn 2 classes. The Tillage region was classified as Corn 1, Corn 2 and Soy 1, while the Caatinga was classified as Tillage and Soy 3.

6.4.2.2 Bebedouro SIR-C in L band image quantitative Analysis

Figure 6.44(a) presents the classification result of SC-H algorithm of the Bebedouro SIR-C in L band image, and Figure 6.44(b) shows the confusion matrix. The initial parameters ware randomly chosen and the number of cluster were nine. This algorithm were able to correctly classify the Corn 1, Corn 2, Bare Soil, and River. The Tillage region were classified as Soy 1 . The Soy 3 region were classified as Soy 2, and the Caatinga is a mix of Caatinga and Soy 3. The overall accuracy is 66.16 %.

Figure 6.44 - Bebedouro SIR-C in L band image classification using the SC-H algorithm.

(a) Classified Image. (b) Confusion matrix.

SOURCE: Author's production.

Figure 6.45 - Bebedouro SIR-C in L band image classification using the BSC-EM-H algorithm.

(a) Classified Image. (b) Confusion matrix.

SOURCE: Author's production.

Figure 6.46 - Bebedouro SIR-C in L band image classification using the BSC-R-H algorithm.

(a) Classified Image.

(b) Confusion matrix.

SOURCE: Author's production.

Figure 6.45(a) and Figure 6.45(b) present the classification result of BSC-EM-H algorithm and its confusion matrix, respectively. The BSC-EM-H algorithm gets the initial parameter using the EM algorithm. This algorithm correctly classified the Corn 1, Corn 2, and Caatinga. It miss-classified the Bare Soil and River, which were classified as Tillage. And there is a substantial confusion between Soy 1, Soy 2, and Soy3 classes. The overall accuracy is 36.29 %.

Figure 6.46(a) and Figure 6.46(b) show the classification result of BSC-R-H algorithm and its confusion matrix, respectively. This algorithm had a better result in comparison with SC-H and BSC-EM-H. It achieved 89.52 % of overall accuracy. The BSC-R-H correctly classified the Corn 1 class, however this class is splitted into a new class (the orange), as described in Section 6.4.1.2. The Corn 2 samples were also correctly classified, while the Soy 1 and Soy 2 were mixed. There was another mixture between Soy 2 and Soy 3. The Tillage was fairly well classified, while the Bare Soil had few samples classified as River. The River and Caatinga achieved a good classification level.

6.4.2.3 Tapajós ALOS/PALSAR image quantitative analysis

Figure 6.47(a) presents the Tapajós ALOS/PALSAR classification result of SC-H algorithm and Figure 6.47(b) shows its confusion matrix. The initial parameters

ware randomly chosen and the number of cluster were six (Bare Soil, Agriculture 1, Agriculture 2, Agriculture 3, Primary Forest, and Pasture). The algorithm was able to identify three main classes represented by the colors green, yellow, and magenta. Only the Primary Forest samples were correctly classified. The Agriculture 3 region was classified as Agriculture 2, and all the other classes were classified as Bare Soil. The overall accuracy is 38.18 %.

Figure 6.47(a) presents the Tapajós ALOS/PALSAR classification result of BSC-EM-H algorithm and Figure 6.47(b) shows its confusion matrix. The initial parameters were selected by the EM algorithm. The algorithm correctly classified the Bare Soil, and some Agriculture 1 samples were classified as Bare Soil as well. However all samples which were classified as Agriculture 1 were actually Agriculture 1, therefore the cell Agriculture 1-Agriculture 1 in the confusion matrix achieved 100 %. The Agriculture 2 were correctly classified, even though some samples from Agriculture 3 were classified as Agriculture 2. The Agriculture 3, Primary Forest and Pasture were correctly classified. The overall accuracy is 85.98 %.

Figure 6.47 - Tapajós image classification using the SC-H algorithm.

(a) Classified Image. (b) Confusion matrix.

SOURCE: Author's production.

Figure 6.49 - Tapajós image classification using the BSC-R-H algorithm.

(a) Classified Image. (b) Confusion matrix.

SOURCE: Author's production.

Figure 6.48 - Tapajós image classification using the BSC-EM-H algorithm.

(a) Classified Image. (b) Confusion matrix.

SOURCE: Author's production.

Figure 6.49(a) presents the Tapajós ALOS/PALSAR classification result of BSC-EM-H algorithm and Figure 6.49(b) shows its confusion matrix. The BSC-RPPD-H had a slightly better overall accuracy than BSC-EM-H, achieving 90.77 %. The biggest difference between these classification is on Bare Soil classification, which had less Agriculture 1 samples classified as Bare Soil.

6.4.2.4 San Francisco ALOS/PALSAR image quantitative analysis

The last analysed PolSAR image is the San Francisco ALOS/PALSAR. Figure 6.50(a) presents the classification result of SC-H algorithm, this algorithm randomly selected the initial parameter and the number of cluster is six. The algorithm was able to correctly identify four classes, the remaining two classes got a very small number of samples. The samples classified as Mountain are from Moutain, LDU, Vegetation, and DU. The samples classified as LDU are from Mountain, LDU, Water, and Vegetation. The samples classified as HDU are from HDU and DU. Even though the Water got 100 % of accuracy in the confusion matrix, this were splitted in two: Water and Mountain. The overall accuracy is 71.6 %.

Figure 6.50 - San Francisco image classification using the SC-H algorithm.

(a) Classified Image. (b) Confusion matrix.

SOURCE: Author's production.

Figure 6.52 - San Francisco image classification using the BSC-R-H algorithm.

(a) Classified Image. (b) Confusion matrix.

SOURCE: Author's production.

Figure 6.51 - San Francisco image classification using the BSC-EM-H algorithm.

(a) Classified Image. (b) Confusion matrix.

SOURCE: Author's production.

Figure 6.51(a) presents the classification result of BS-EM-H algorithm and Figure 6.51(b) shows its confusion matrix. This images was fairly well classified. The biggest

confusion happened between Water and Mountain. Some Mountain samples were classified as Water. The overall accuracy was 79.49 %.

Finally, Figure 6.52(a) presents the classification result of BS-RPDDP-H algorithm and Figure 6.52(b) shows the confusion matrix and the classes samples distribution. This algorithm classified fairly well all classes. Note that the Water class were splitted into two more, and this splitting were already discussed in Section 6.4.1.5.

7 CONCLUSIONS

In this work, a new unsupervised algorithm to clustering/classify PolSAR images was proposed. The algorithm is named Bisecting Stochastic Clustering, and uses the proposed RPDDP and the Stochastic Clustering to build a hierarchical structure that stores the clustering splitting history. The algorithm input is an image and the output is a classified image and a set of tools that may help the analyst to better understating the clustering result.

The focus of this work relied in PolSAR data. This kind of data is widely used in the process of monitoring the environment, due to the advantages of sensors operating in microwave frequencies, such as independence from daylight and less influence from the weather. Therefore this work explored the characteristic of PolSAR images, as the scattering mechanisms, with the goal of achieve the best classification result.

The PolSAR data used in this work is assumed to follow the Wishart distribution, therefore the image regions are considered homogeneous and multi-look. It means that the real and imaginary parts of polarimetric bands (s_{hh}, s_{hv}, and s_{vv}) are assumed to follow a circularly symmetric multi-variate complex Gaussian distribution. Consequently the stochastic distances are computed between the Wishart parameters. In order to validate the algorithm metric, a set of PolSAR were simulated, where the main idea underlying this simulation is to guarantee that the cluster samples will follow the defined distribution and that the samples have less noise as possible.

Simulated images were used to validate the BSC classification. They were also used in order to investigate the initial parameter estimation influence on partitional algorithms. The BSC algorithm is a hybrid algorithm between the divisive hierarchical clustering algorithms and partitional clustering algorithms. The BSC partitional algorithm is the SC algorithm, therefore, this algorithm were used in order to investigate the greedy behavior of partitional algorithms.

Six scenarios were defined in order to investigate the classification behaviour according to the initial parameter determination. Except for scenario S06, all others scenarios had an accuracy result bellow 65 %, what implies that, besides many other facts, the correctness of the initial parameters determination is essential to accomplish a good and consistent classification result. The use of assertive initial parameter, as presented in scenario S06 significantly improved the final classification result, and the accuracy was around 20 % better.

Specially when the clusters are well separated, the SC performance depends completely on the goodness of the initialization. For instance, lets consider S02, where two out the of six parameters were picked from the Bare Soil class. One of these initial parameters should be associated to Bare Soil and the second should be associated to River, however both seeds stood trapped in local minimum and the SC wasn't able to discovered the River class, and the segments which belongs to River were classified as Bare Soil.

Although there are many initialization algorithms in the literature, as the Ball and Hall's method, Simple Cluster Seeking method, Maximin method, in general, the SC initialization problem is not easy to solve. Therefore, our proposal has the goal of divide the problem, and works on steps by using the hierarchical algorithm. In this way two clusters could be found per iteration, and the RPDDP would be used in order to define the two initial parameters.

The RPDDP is a modification of PDDP algorithm in order to handle covariance matrices. The PolSAR image can be defined by a group of covariance matrices Z, which can be better represented in a Riemannian space. In order to investigated the difference between the average covariance matrix and the intrinsic mean covariance matrix, an experiment using the PDDP against the RPDDP algorithms were conducted.

Four PolSAR images from different sensors and frequency band are used in this experiment: The Bebedouro SIR-C in C band, the Bebedouro SIR-C in L band, the Tapajós ALOS/PALSAR, and the San Francsico ALOS/PALSAR. As evaluation criteria, the stochastic distance between classes and the Root-Mean-Square Standard Deviation (RMSSTD) error were computed. The goal of these evaluation criteria is to judge the cluster quality by assuming that an optimum clustering means compact and well separated clusters, i. e., a cluster with low RMSSTD and high stochastic distance inter cluster shall be considered optimum.

The RPDDP had a bigger stochastic distance and lower RMSSTD error with the Bebedouro SIR-C in C and L bands, and with the Tapajós ALOS/PALSAR. The San Francisco ALOS/PALSAR clusters had a bigger distance with RPDDP, but one of the clusters had a smaller RMSSTD error with PDDP. In general the RPDDP presented a better result; the cluster were less nosier and had a better definition of the cluster borders, contrasting the PDDP results.

The BSC classification result were evaluated in a quantitative and qualitative way.

While the quantitative analysis evaluates the confusion matrices and accuracy results, the qualitative analysis deals with the cluster information gain, the BSC dendrogram, and the PolSAR scattering mechanisms.

Besides the initial parameter determination and RPDDP versus PDDP analysis, two more experiments were analysed, one using simulated images and the second using real PolSAR images. Therefore, the third experiment had the goal of quantify the sensitiveness of the BSC algorithm, for such, a Monte Carlo simulation over a set of one hundred simulated PolSAR images was conducted. Aiming to ratify the Monte Carlo simulation results, the second experiment used the Bebedouro SIR-C in C band, the Bebedouro SIR-C in L band, the Tapajós ALOS/PALSAR, and the San Francsico ALOS/PALSAR as input of BSC algorithm. While the first experiment addressed only a quantitative analysis, the second inspected the accuracy result, the BSC dendrogram information and the images scattering mechanisms.

In the third experiment, a Monte Carlo simulation delivered the accuracy results of the classification using the SC, BSC using the EM as estimator of initial parameter, and the BSC using the RPDDP in order to get the initial parameters. The results showed that the SC had an inconsistency on the accuracy value, whith this algorithm the classification accuracy varied from 60 % to 90 %. While the BSC-EM and the BSC-R had more constancy. The classification accuracy result varied from 90 % to 95 % with BSC-EM, and the classification accuracy result varied from 95 % to 98 % with BSC-R. Therefore, with simulated PolSAR images, the BSC-R had the best results.

In the forth experiment, the dendrogram and PolSAR scattering mechanisms were addressed. Using the dendrogram report is possible to follow the clustering bipartition steps, and the information gain can help on the decision if a data were well clustered or not. Also, the Plan $H - \alpha$ helps on the clusters characterization.

Alone, the dendrogram, information gain, and scattering mechanisms information, can provide knowledge and advice about what to do with a dataset. Together they form a powerful tool for segmentation, clustering, classification, and analysis of PolSAR data.

One of the advantages of the BSC algorithm is that there is no need for inputs such as the number of classes. The algorithm can run until the stop criteria is reached and generate a suitable number of classes. The dendrogram, information gain, and scattering mechanisms information helps on the unknown classes characterization.

For instance the BSC-R-H classified the San Francisco image and generated seven classes more than the predicted on the truth image. Three out of seven classes were created on Water region, meaning that the ocean may have different classes due to, for example, sea level, pollution, and so on.

In general the L bands images had superior classification results in comparison with the C band image. The real PolSAR image classification qualitative analysis ratified the results of experiment 3 and showed that the BSC-R had the best accuracy result. The L bands images had an average accuracy result of approximately 60 % with the SC-H algorithm, 70 % with the BSC-EM-H, and 90 % with the BSC-R.

In Section 1.3 two hypothesis has been defined. The first statement hypnotised that the use of the Rimmenian geometry to deal with PolSAR covariance matrices could be beneficial, specially to the determination of the initial parameters of the BSC classifier. As already discussed the the RPDDP proved to be a good algorithm to the determination of the initial parameter of PolSAR images data. The second statement hypnotised that it was possible to develop a robust unsupervised classifier using divisive hierarchical algorithms and stochastic distances. As presented in Chapter 6, the BSC using the RPDDP had a better overall accuracy with all tested image. Therefore the hypothesis has been proved correct.

7.1 Future works

The BSC algorithm has potential for many adaptations, tests, and experiment. In this context, some experiments and results can be improved or further extended. This section briefly describes some interesting research topics, which worth investigating. Here are these points:

a) Deeper analysis on PolSAR image covariances under the Riemannian geometry. Exploring the mathematical foundation of the Hermitian positive definite covariance matrices and how it forms a differentiable Riemannian manifold. Explore also the geodesic distances against stochastic distances;

b) As presented in Chapter 2, the SAR image can follow diverse PDFs according to the type of image and region. This work addressed only the Wishart distribution. Therefore, the study and development of stochastic distances between the $\mathcal{G}^0$ and $\mathcal{K}$ distributions, for instance, can be considered as well.

c) A short analysis of multi-frequency image classification was conducted on Chapter 6, therefore a deeper analysis on the potential of multi-frequency

data is required. Since this kind of image can improve the classification accuracy, as illustrated by many researchers (FERRO-FAMIL et al., 2001; FRERY et al., 2007; FREEMAN et al., 1994);

d) The use of multi-source images as input to BSC algorithm, which can be applied to hyperspectral data, or fused data from a SAR and camera.

e) From the point of view of application, this algorithm can help on the binary classes characterization, for instance forest and non forest, water and non water.

f) The BSC can be used to extract features, helping on the unsupervised learning development, which is a type of machine learning that seeks for patterns in a dataset without pre-existing labels.

g) The BSC was developed in Matlab, which is a proprietary, closed-source software, that needs a license, which is quite expensive. Therefore the algorithm translation to an open source language is needed.

REFERENCES

AGGARWAL, C. C. **Data classification: algorithms and applications**. [S.l.]: CRC Press, 2014. 35

ALASHWAL, H.; HALABY, M. E.; CROUSE, J. J.; ABDALLA, A.; MOUSTAFA, A. A. The application of unsupervised clustering methods to Alzheimer's disease. **Frontiers in Computational Neuroscience**, v. 13, 2019. 34

ARCGIS. **Oil spill monitoring from space**. 2019. Disponível em: <https://www.arcgis.com/apps/MapTour/index.html?appid=6ec47cd4a0984af987dbd4c31ee5db73/>. 1

ARSENAULT, H.; APRIL, G. Properties of speckle integrated with a finite aperture and logarithmically transformed. **JOSA**, v. 66, n. 11, p. 1160–1163, 1976. 23, 73

ARSIGNY, V.; FILLARD, P.; PENNEC, X.; AYACHE, N. Log-euclidean metrics for fast and simple calculus on diffusion tensors. **Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine**, v. 56, n. 2, p. 411–421, 2006. 31

AZIMI-SADJADI, M. R.; BANNOUR, S. Two-dimensional adaptive block Kalman filtering of sar imagery. **IEEE transactions on Geoscience and Remote Sensing**, v. 29, n. 5, p. 742–753, 1991. 13

BANDYOPADHYAY, S.; SAHA, S. **Unsupervised classification: similarity measures, classical and metaheuristic approaches, and applications**. [S.l.]: Springer Science & Business Media, 2012. 38, 40

BARACHANT, A.; BONNET, S.; CONGEDO, M.; JUTTEN, C. Classification of covariance matrices using a riemannian-based kernel for bci applications. **Neurocomputing**, v. 112, p. 172–178, 2013. 31, 32

BERRY, M. J.; LINOFF, G. S. **Data mining techniques: for marketing, sales, and customer relationship management**. [S.l.]: John Wiley & Sons, 2004. 43

BIRKHOFF, G. Lattice theory. In: AMERICAN MATHEMATICAL SOCIETY COLLOQUIUM PUBLICATIONS, 1967 (Ed.). **Proceedings...** [S.l.: s.n.], 1967. v. 25. 33

BOCK, H.-H. Origins and extensions of the k-means algorithm in cluster analysis. **Electronic Journal for History of Probability and Statistics**, v. 4, n. 2, p. 1–18, 2008. 39

BOLEY, D. Principal direction divisive partitioning. **Data mining and Knowledge Discovery**, v. 2, n. 4, p. 325–344, 1998. 51

BRAGA, B. C.; FREITAS, C. d. C.; SANT'ANNA, S. J. Multisource classification based on uncertainty maps. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2015 (Ed.). **Proceedings...** [S.l.], 2015. p. 1630–1633. 2, 23

CALVO, H.; GODOY-CALDERÓN, S.; MORENO-ARMENDÁRIZ, M. A.; MARTÍNEZ-HERNÁNDEZ, V. M. Forecasting, clustering and patrolling criminal activities. **Intelligent Data Analysis**, v. 21, n. 3, p. 697–720, 2017. 34

CANADA, N. R. **Polarization in radar systems**. 2014. Disponível em: <https://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/satellite-imagery-products/educational-resources/9567>. 15

CARVALHO, N. C. R. L.; BINS, L. S.; SANT'ANNA, S. J. S. Analysis of stochastic distances and wishart mixture models applied on polsar images. **Remote Sensing**, Multidisciplinary Digital Publishing Institute, v. 11, n. 24, p. 2994, 2019. 4, 39, 48, 49, 52, 54, 68, 78, 91, 107

CELEBI, M. E.; KINGRAVI, H. A.; VELA, P. A. A comparative study of efficient initialization methods for the k-means clustering algorithm. **Expert Systems With Applications**, v. 40, n. 1, p. 200–210, 2013. 5, 50

CERRA, D.; DATCU, M. Compression-based hierarchical clustering of SAR images. **Remote Sensing Letters**, v. 1, n. 3, p. 141–147, 2010. 4

CHAKRABORTY, D. Clustering techniques for land use land cover classification of remotely sensed images. In: RUSTAMOV, R. B. (Ed.). **Geographic Information systems in geospatial intelligence**. [S.l.]: IntechOpen, 2019. 34

CLOUDE, S. R.; POTTIER, E. A review of target decomposition theorems in radar polarimetry. **IEEE Transactions on Geoscience and Remote Sensing**, v. 34, n. 2, p. 498–518, 1996. 20, 85

_____. An entropy based classification scheme for land applications of polarimetric sar. **IEEE Transactions on Geoscience and Remote Sensing**, v. 35, n. 1, p. 68–78, 1997. 3, 23

CLOUDE, S. R.; POTTIER, E.; BOERNER, W. M. Unsupervised image classification using the entropy/alpha/anisotropy method in radar polarimetry. In: NASA-JPL, AIRSAR-02 WORKSHOP, 2002 (Ed.). **Proceedings...** [S.l.: s.n.], 2002. 21

CONGEDO, M.; BARACHANT, A. A special form of spd covariance matrix for interpretation and visualization of data manipulated with riemannian geometry. **AIP Conference Proceedings**, v. 1641, n. 1, p. 495–503, 2015. 51

CORREIA, A. H. **Projeto, densenvolvimento e avaliação de classificadores estatísticos pontuais e contextuais para imagens SAR polarimétricas**. Dissertao (Mestrado em Sensoriamento Remoto) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1998. 60, 61

DENG, X.; LÓPEZ-MARTÍNEZ, C.; CHEN, J.; HAN, P. Statistical modeling of polarimetric sar data: a survey and challenges. **Remote Sensing**, v. 9, n. 4, p. 348, 2017. 18

DEZA, M. M.; DEZA, E. **Encyclopedia of distances**. [S.l.]: Springer, 2009. 583 p. 36

DIERKING, W.; BUSCHE, T. Sea ice monitoring by l-band sar: an assessment based on literature and comparisons of jers-1 and ers-1 imagery. **IEEE Transactions on Geoscience and Remote Sensing**, v. 44, n. 4, p. 957–970, 2006. 1

DOULGERIS, A. P. An automatic $\backslash u$-distribution and markov random field segmentation algorithm for polsar images. **IEEE Transactions on Geoscience and Remote Sensing**, v. 53, n. 4, p. 1819–1827, 2014. 2, 23

DOULGERIS, A. P.; ELTOFT, T. Polsar image segmentation—advanced statistical modelling versus simple feature extraction. In: IEEE GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2014 (Ed.). **Procedings...** [S.l.], 2014. p. 1021–1024. 2, 23

DUNN, J. C. **A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters**. [S.l.]: Taylor & Francis, 1973. 43

ESTER, M. et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 1996 (Ed.). **Procedings...** [S.l.: s.n.], 1996. v. 96, n. 34, p. 226–231. 41

EUROPEAN SPACE AGENCY - ESA. **Radar course**. 2017. Disponível em: <https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/ers/instruments/sar/applications/radar-courses/course-2/>. 14

FAHAD, A.; ALSHATRI, N.; TARI, Z.; ALAMRI, A.; KHALIL, I.; ZOMAYA, A. Y.; FOUFOU, S.; BOURAS, A. A survey of clustering algorithms for big data: taxonomy and empirical analysis. **IEEE Transactions on Emerging Topics in Computing**, v. 2, n. 3, p. 267–279, 2014. 38, 41

FERRO-FAMIL, L.; POTTIER, E.; LEE, J.-S. Unsupervised classification of multifrequency and fully polarimetric sar images based on the h/a/alpha-wishart classifier. **IEEE Transactions on Geoscience and Remote Sensing**, v. 39, n. 11, p. 2332–2342, 2001. 157

FORMONT, P.; PASCAL, F.; VASILE, G.; OVARLEZ, J.-P.; FERRO-FAMIL, L. Statistical classification for heterogeneous polarimetric sar images. **IEEE Journal of Selected Topics in Signal Processing**, v. 5, n. 3, p. 567–576, 2010. 23

FREEMAN, A.; DURDEN, S. L. A three-component scattering model for polarimetric sar data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 36, n. 3, p. 963–973, 1998. 3, 20

FREEMAN, A.; VILLASENOR, J.; KLEIN, J.; HOOGEBOOM, P.; GROOT, J. On the use of multi-frequency and polarimetric radar backscatter features for classification of agricultural crops. **International Journal of Remote Sensing**, v. 15, n. 9, p. 1799–1812, 1994. 157

FREEMAN, A.; ZINK, M.; CARO, E.; MOREIRA, A.; VEILLEUX, L.; WERNER, M. The legacy of the sir-c/x-sar radar system: 25 years on. **Remote Sensing of Environment**, v. 231, p. 111255, 2019. 59

FRERY, A. C.; CINTRA, R. J.; NASCIMENTO, A. D. Entropy-based statistical analysis of polsar data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 51, n. 6, p. 3733–3743, 2012. 4, 55

FRERY, A. C.; CORREIA, A.; RENNÓ, C. D.; FREITAS, C.; JACOBO-BERLLES, J.; MEJAIL, M.; VASCONCELLOS, K.; SANT'ANNA, S. Models for synthetic aperture radar image analysis. **Resenhas (IME-USP)**, v. 4, n. 1, p. 45–77, 1999. 2

FRERY, A. C.; CORREIA, A. H.; FREITAS, C. d. C. Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information. **IEEE Transactions on Geoscience and Remote Sensing**, v. 45, n. 10, p. 3098–3109, 2007. 26, 157

FRERY, A. C.; MULLER, H.-J.; YANASSE, C. d. C. F.; SANT'ANNA, S. J. S. A model for extremely heterogeneous clutter. **IEEE Transactions on Geoscience and Remote Sensing**, v. 35, n. 3, p. 648–659, 1997. 26

FRERY, A. C.; NASCIMENTO, A. D.; CINTRA, R. J. Analytic expressions for stochastic distances between relaxed complex wishart distributions. **IEEE Transactions on Geoscience and Remote Sensing**, v. 52, n. 2, p. 1213–1226, 2013. 2, 27, 29, 30, 75, 114

FROST, V. S.; STILES, J. A.; SHANMUGAN, K. S.; HOLTZMAN, J. C. A model for radar images and its application to adaptive digital filtering of multiplicative noise. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, n. 2, p. 157–166, 1982. 13

GAN, G.; MA, C.; WU, J. **Data clustering: theory, algorithms, and applications**. [S.l.]: SIAM, 2007. 39

GAO, G. Statistical modeling of sar images: a survey. **Sensors**, v. 10, n. 1, p. 775–795, 2010. 23, 24, 25

GOMEZ, L.; ALVAREZ, L.; MAZORRA, L.; FRERY, A. C. Fully polsar image classification using machine learning techniques and reaction-diffusion systems. **Neurocomputing**, v. 255, p. 52–60, 2017. 5

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. [S.l.]: Springer Science & Business Media, 2009. 35

HELLMAN, M. **Classification of fully polarimetric SAR-data for cartographic applications**. Tese (PhD dissertation) — Institute for High Frequency Technology and Radar Systems - DLR, 2000. 13, 16

HENDERSON, F. M.; XIA, Z.-G. Sar applications in human settlement detection, population estimation and urban land use pattern analysis: a status report. **IEEE Transactions on Geoscience and Remote Sensing**, v. 35, n. 1, p. 79–85, 1997. 1

HIDOT, S.; SAINT-JEAN, C. An expectation–maximization algorithm for the wishart mixture model: application to movement clustering. **Pattern Recognition Letters**, Elsevier, v. 31, n. 14, p. 2318–2324, 2010. 52

HORN, R. A.; JOHNSON, C. R. **Matrix analysis**. [S.l.]: Cambridge University Press, 2012. 30

IMAGING, E. S. **Satellites: Earth's most advanced constellation**. 2019. Disponível em: <https://www.euspaceimaging.com/about/satellites/>. 1

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. COORDENAÇÃO GERAL DE OBSERVAÇÃO DA TERRA - INPE/OBT. **Observação da Terra**. 2016. Disponível em: <http://www.obt.inpe.br/projetos.php>. 2

JAIN, A. K. Data clustering: 50 years beyond k-means. **Pattern Recognition Letters**, v. 31, n. 8, p. 651–666, 2010. 33, 48, 78

JAIN, A. K.; DUBES, R. C. **Algorithms for clustering data**. [S.l.]: Prentice-Hall, 1988. 35

JENSEN, J. R. **Introductory digital image processing: a remote sensing perspective**. [S.l.]: Pearson Prentice Hall, 2005. 42

KARYPIS, M. S. G.; KUMAR, V.; STEINBACH, M. A comparison of document clustering techniques. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2000 (Ed.). **Proceedings...** [S.l.: s.n.], 2000. 40

KONG, J.; SWARTZ, A.; YUEH, H.; NOVAK, L.; SHIN, R. Identification of terrain cover using the optimum polarimetric classifier. **Journal of Electromagnetic Waves and Applications**, Taylor & Francis, v. 2, n. 2, p. 171–194, 1988. 3

KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. **The Annals of Mathematical Statistics**, v. 22, n. 1, p. 79–86, 1951. 4, 27

KURT, W. **Kullback-Leibler divergence explained**. 2017. Disponível em: <https://www.countbayesie.com/blog/2017/5/9/kullback-leibler-divergence-explained/>. 27

KUSSUL, N.; SHELESTOV, A.; SKAKUN, S. Flood monitoring from sar data. In: KOGAN F.; POWELL, A. F. O. (Ed.). **Use of satellite and in-situ data to improve sustainability**. [S.l.]: Springer, 2011. p. 19–29. 1

LEE, J.; POTTIER, E. **Polarimetric radar imaging: from basics to applications**. [S.l.]: CRC Press, 2009. 10, 12, 16, 18, 20, 21, 24, 30

LEE, J.-S. Digital image enhancement and noise filtering by use of local statistics. **IEEE Transactions on Pattern Analysis & Machine Intelligence**, n. 2, p. 165–168, 1980. 13

_____. Refined filtering of image noise using local statistics. **Computer Graphics and Image Processing**, v. 15, n. 4, p. 380–389, 1981. 13

_____. A simple speckle smoothing algorithm for synthetic aperture radar images. **IEEE Transactions on Systems, Man, and Cybernetics**, n. 1, p. 85–89, 1983. 13

LEE, J.-S.; GRUNES, M. R.; AINSWORTH, T. L.; DU, L.-J.; SCHULER, D. L.; CLOUDE, S. R. Unsupervised classification using polarimetric decomposition and the complex wishart classifier. **IEEE Transactions on Geoscience and Remote Sensing**, v. 37, n. 5, p. 2249–2258, 1999. 3

LEE, J.-S.; GRUNES, M. R.; KWOK, R. Classification of multi-look polarimetric sar imagery based on complex wishart distribution. **International Journal of Remote Sensing**, v. 15, n. 11, p. 2299–2311, 1994. 3

LENCINA, A.; VAVELIUK, P.; TEBALDI, M.; BOLOGNINI, N. Modulated speckle simulations based on the random-walk model. **Optics Letters**, v. 28, n. 19, p. 1748–1750, 2003. 24

LIM, L.-H.; SEPULCHRE, R.; YE, K. Geometric distance between positive definite matrices of different dimensions. **IEEE Transactions on Information Theory**, v. 65, n. 9, p. 5401–5405, 2019. 3, 31

LIN, Z. Riemannian geometry of symmetric positive definite matrices via cholesky decomposition. **SIAM Journal on Matrix Analysis and Applications**, v. 40, n. 4, p. 1353–1370, 2019. 31

LIU, H.; YANG, S.; GOU, S.; LIU, S.; JIAO, L. Terrain classification based on spatial multi-attribute graph using polarimetric sar data. **Applied Soft Computing**, v. 68, p. 24–38, 2018. 1

LIU, X.; JIAO, L.; LIU, F. Polsf: Polsar image dataset on san francisco. **arXiv preprint arXiv:1912.07259**, 2019. 70, 71

LÓPEZ-MARTÍNEZ, C.; FERRO-FAMIL, L.; POTTIER, E. Tutorial on sar polarimetry. 2005. 14, 24

MOAKHER, M. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. **SIAM Journal on Matrix Analysis and Applications**, v. 26, n. 3, p. 735–747, 2005. 32

MOREIRA, A.; PRATS-IRAOLA, P.; YOUNIS, M.; KRIEGER, G.; HAJNSEK, I.; PAPATHANASSIOU, K. P. A tutorial on synthetic aperture radar. **IEEE Geoscience and Remote Sensing Magazine**, v. 1, n. 1, p. 6–43, 2013. 11

MÜLLER-HANSEN, F.; CARDOSO, M. F.; DALLA-NORA, E. L.; DONGES, J. F.; HEITZIG, J.; KURTHS, J.; THONICKE, K. A matrix clustering method to explore patterns of land-cover transitions in satellite-derived maps of the brazilian amazon. **Nonlinear Processes in Geophysics**, v. 24, n. 1, p. 113, 2017. 34

NASCIMENTO, A. a. D. C. d. **Teoria estatística da informação para dados de Radar de Abertura Sintética Univariados e Polarimétricos**. Tese (Doutorado em Estatística) — Universidade Federal de Pernambuco - Recife, 2012. 2, 4, 14, 28, 29, 54

NEGRI, R. G.; FRERY, A. C.; SILVA, W. B.; MENDES, T. S.; DUTRA, L. V. Region-based classification of polsar data using radial basis kernel functions with stochastic distances. **International Journal of Digital Earth**, v. 12, n. 6, p. 699–719, 2019. 5, 23

PALUMBO, G. **Detroit: reducing the noise speckle**. 2017. Disponível em: <https://www.earthstartsbeating.com/2017/05/26/detroit-reducing-the-noise-speckle/>. 13

PEDRINI, H.; SCHWARTZ, W. **Análise de imagens digitais: princípios, algoritmos e aplicações**. [S.l.]: Thomson Pioneira, 2008. ISBN 9788522105953. 33

POTTIER, E. **PolSARpro Home**. 2019. Disponível em: <https://earth.esa.int/web/polsarpro/home>. 3

POTTIER, E.; CLOUDE, S. R. Application of the h/a/alpha polarimetric decomposition theorem for land classification. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Wideband Interferometric Sensing and Imaging Polarimetry**. [S.l.], 1997. v. 3120, p. 132–143. 3

PRZYBORSKI, P. **San Francisco from Landsat 7**. 2020. Disponível em: <https://earthobservatory.nasa.gov/images/451/san-francisco-from-landsat-7>. 174

REIGBER, A.; HELLWICH, O. Rat (radar tools): a free sar image analysis software package. In: EUROPEAN CONFERENCE ON SYNTHETIC APERTURE RADAR, . (Ed.). **Proceedings...** [S.l.: s.n.], 2004. v. 4, p. 997–1000. 3

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53–65, 1987. 43

SALDANHA, M. F. S. **Um segmentador multinível para imagens SAR polarimétricas baseado na distribuição Wishart**. Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2013. 2, 23

SALICRU, M.; MENENDEZ, M.; MORALES, D.; PARDO, L. Asymptotic distribution of (h, φ)-entropies. **Communications in Statistics-Theory and Methods**, v. 22, n. 7, p. 2015–2031, 1993. 28, 55

SAPUCCI, G. R.; NEGRI, R. G. Hierarchical clustering and stochastic distance for indirect semi-supervised remote sensing image classification. **SN Applied Sciences**, v. 1, n. 3, p. 272, 2019. 4

SEGHOUANE, A.-K.; AMARI, S.-I. The aic criterion and symmetrizing the kullback–leibler divergence. **IEEE Transactions on Neural Networks**, v. 18, n. 1, p. 97–106, 2007. 29

SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**, v. 27, n. 3, p. 379–423, 1948. 27

SILVA, W.; FREITAS, C.; SANT'ANNA, S.; FRERY, A. C. Polsar region classifier based on stochastic distances and hypothesis tests. In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, . (Ed.). **Proceedings...** [S.l.], 2012. p. 1473–1476. 4

SILVA, W. B.; FREITAS, C. C.; SANT'ANNA, S. J.; FRERY, A. C. Classification of segments in polsar imagery by minimum stochastic distances between wishart distributions. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 6, n. 3, p. 1263–1273, 2013. 5, 23

SILVA, W. B. d. **Classificação de região de imagens utilizando teste de hipótese baseado em distâncias estocásticas: aplicação a dados polarimétricos**. Tese (Doutorado em Sensoriamento Remoto) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2013. 73, 74

SKOLNIK, M. **Radar Handbook**. McGraw-Hill Education, 2008. (Electronics electrical engineering). ISBN 9780071485470. Disponível em: <https://books.google.cz/books?id=76uF2Xebm-gC>. 9, 12

SNEATH, P. et al. **Numerical taxonomy. the principles and practice of numerical classification.** [S.l.: s.n.], 1973. 33

SOBRINO, J. A. **Recent advances in quantitative remote sensing**. [S.l.]: Universitat de València, 2002. 24

SONG, H.; YANG, W.; XU, X.; LIAO, M. Unsupervised polsar imagery classification based on jensen-bregman logdet divergence. In: EUROPEAN CONFERENCE ON SYTHETIC APERTURE RADAR, . (Ed.). **Proceedings...** [S.l.: s.n.], 2014. p. 1–4. 4

STRUYF, A. et al. Clustering in an object-oriented environment. **Journal of Statistical Software**, v. 1, n. 4, p. 1–30, 1997. 47

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern recognition & Matlab intro**. [S.l.]: Academic Press, 2010. 42, 173

TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 63, n. 2, p. 411–423, 2001. 36

TORRES, L.; SANT'ANNA, S. J.; FREITAS, C. da C.; FRERY, A. C. Speckle reduction in polarimetric sar imagery with stochastic distances and nonlocal means. **Pattern Recognition**, v. 47, n. 1, p. 141–157, 2014. 2, 13, 23

TSAI, C.-Y.; CHIU, C.-C. A purchase-based market segmentation methodology. **Expert Systems with Applications**, v. 27, n. 2, p. 265–276, 2004. 34

WILEY, C. A. Synthetic aperture radars. **IEEE Transactions on Aerospace and Electronic Systems**, n. 3, p. 440–443, 1985. 11

WOLFF, C. **Side Looking Airborne Radar (SLAR)**. 2018. Disponível em: <http://www.radartutorial.eu/20.airborne/ab06.en.html>. 10

WU, F.; WANG, C.; ZHANG, H.; ZHANG, B. Vessel detection and analysis combining sar images and ais information. In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, . (Ed.). **Proceedings...** [S.l.], 2016. p. 6633–6636. 1

XIANG, D.; TANG, T.; HU, C.; FAN, Q.; SU, Y. Built-up area extraction from polsar imagery with model-based decomposition and polarimetric coherence. **Remote Sensing**, v. 8, n. 8, p. 685, 2016. 136

XU, R.; WUNSCH, D. Survey of clustering algorithms. **IEEE Transactions on Neural Networks**, v. 16, n. 3, p. 645–678, 2005. 34

YANG, W.; LIU, Y.; XIA, G.-S.; XU, X. Statistical mid-level features for building-up area extraction from high-resolution polsar imagery. **Progress In Electromagnetics Research**, v. 132, 2012. 2, 23

YAO, W.; DUMITRU, C. O.; LOFFELD, O.; DATCU, M. Semi-supervised hierarchical clustering for semantic sar image annotation. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 9, n. 5, p. 1993–2008, 2016. 4

YELLOWBRICK. **Clustering visualizers**. 2019. Disponível em: <https://www.scikit-yb.org/en/latest/api/cluster/index.html/>. 36

YING-HUA, W.; CHONG-ZHAO, H. Polsar image segmentation by mean shift clustering in the tensor space. **Acta Automatica Sinica**, v. 36, n. 6, p. 798–806, 2010. 51

ZAÏANE, O. R.; FOSS, A.; LEE, C.-H.; WANG, W. On data clustering analysis: scalability, constraints, and validation. In: PACIFIC-ASIA CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, . (Ed.). **Proceedings...** [S.l.: s.n.], 2002. p. 28–39. 47

ZHAO, Q.; LI, X.; LI, Y. Multilook sar image segmentation with an unknown number of clusters using a gamma mixture model and hierarchical clustering. **Sensors**, v. 17, n. 5, p. 1114, 2017. 4

APPENDIX A - BSC FOR IMAGES FOLLOWING THE GAUSSIAN DISTRIBUTION.

Single-look SAR image from homogeneous region, in-phase or quadrature, can follow the Gaussian distribution. Similarly, multi-look SAR images in amplitude from homogeneous region also can follow the Gaussian distribution. Although optical images are not accurately modeled by Gaussian distributions, this is one of the most applied models when dealing with this kind of images.

Let $x = \{\vec{x_1}, ..., \vec{x_N}\}$ be the dataset of unknown samples to be clustered, where $\vec{x_n}$ is represented as a q-tuple of real numbers, q is the dataset number of bands, and N is the number of samples. If $q > 1$, the dataset follows the multivariate Gaussian distribution denoted as $x \sim \mathcal{N}(\vec{\mu}, \mathbf{\Sigma})$, where $\vec{\mu}$ is the vector of q mean values, and $\mathbf{\Sigma}$ is a covariance matrix. The multivariate Gaussian distribution is represented as:

$$f(x; \vec{\mu}, \mathbf{\Sigma}) = \frac{1}{(\sqrt{2\pi})^q} |\mathbf{\Sigma}|^{-\frac{1}{2}} exp\Big(- \frac{1}{2}(x - \vec{\mu})^T \mathbf{\Sigma}^{-1}(x - \vec{\mu})\Big) \tag{A.1}$$

where$\mu = \frac{1}{N}\sum_{n=1}^{N} \vec{x_n}$, and $\mathbf{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(\vec{x_n} - \vec{\mu})^T(\vec{x_n} - \vec{\mu})$.

The BSC algorithm can be easily adjusted to handle a dataset following the Gaussian distribution. Three adjusts must be done: The first one concerns on the initial parameter determination, where the PDDP should use the regular PCA approach, while the the EM should use the Gaussian Mixture model; the second adjust is about the stochastic distance, which shall be between two Gaussian distribution; and the third, is regarding the entropy calculation.

A.1 Initial parameter determination

A.1.1 PDDP

The PDDP algorithm is based on the PCA routing, which is a method often used to reduce the dimensionality of large data sets. However, in here it is used to get the principal component, which will be later used to divide the dataset into two subsets.

The PCA computes the mean ($\vec{\mu}$) and covariance matrix ($\mathbf{\Sigma}$) of a given dataset, and decomposes the covariance matrix into eigenvectors (U) and eigenvalues (Λ) matrices. The q principal components of the observed dataset x are given by the vectors $\vec{\delta} = U^T(\vec{x_n} - \vec{\mu})$.

The principal direction is given by the eigenvector u_i with the highest eigenvalue λ_i associated to it. The projection of $\vec{x_n}$ on the principal component is given by:

$$\sigma v_n = u_i^T(\vec{x_n} - \vec{\mu}) \tag{A.2}$$

The correspondent samples $\vec{x_n}$ associated to values v_n are splitted into two groups accordingly to the v_n sign. Therefore two groups are created, and their mean and covariance matrices are considered the initial parameter for BSC algorithm.

A.1.2 EM

The EM algorithm is an iterative procedure that uses the maximum a posteriori rule to compute the maximum likelihood of a mixture model distribution, when the data is incomplete, has missing data points, or has unobserved variables. Assuming $x = \{\vec{x_1}, ..., \vec{x_N}\}$ a set of observed samples, the multivariate Gaussian mixture model can be expressed by:

$$f(x; \Psi) = \sum_{k=1}^{K} \pi_k f_k(x; \vec{\mu}, \mathbf{\Sigma}) \tag{A.3}$$

where K is the number of Gaussian within the mixture, $\Psi = \{\pi_1, ..., \pi_K, \vec{\mu_1}, ..., \vec{\mu_2}, \mathbf{\Sigma_1}, ...\mathbf{\Sigma_K}\}$ is the mixture model parameter vector, and π_k is the weighting factor per Gaussian.

Each EM iteration consists on an E-step and an M-step. The E-step uses the current Ψ^t value to compute the log-likelihood of the observed data, from where the π_k membership weights of data sample $\vec{x_n}$ to the distribution f_k can be derived. The log-likelihood of the Gaussian mixture model is expressed in Equation (A.4). In the M-step, the weights π_k, the mean $\vec{\mu_k}$ and covariance matrices $\mathbf{\Sigma_k}$ of the K distribution within the mixture model are updated, as presented in Equation (**??**), (A.6) and (A.7).

$$Q(\Psi, \Psi^t) = \sum_{n=1}^{N} \sum_{k=1}^{K} u_{n,k} \Big[log(\pi_k^t) - \frac{q}{2} log(2\pi) - \frac{1}{2} log(|\mathbf{\Sigma_k^t}|) - \frac{1}{2} (\vec{x_n} - \vec{\mu_k^t})^T (\mathbf{\Sigma_k^t})^{-1} (\vec{x_n} - \vec{\mu_k^t}) \Big] \tag{A.4}$$

where $u_{n,k} = 1$ if the sample n produces a measurement k, and $u_{n,k} = 0$ otherwise.

$$\pi_k^{t+1} = \frac{1}{N}\sum_{n=1}^{N} u_{n,k} \tag{A.5}$$

$$\mu_k^{t+1} = \frac{\sum_{n=1}^{N} u_{n,k}\vec{x_n}}{\sum_{n=1}^{N} u_{n,k}} \tag{A.6}$$

$$\boldsymbol{\Sigma_k^{t+1}} = \frac{\sum_{n=1}^{N} u_{n,k}(\vec{x_n} - \vec{\mu_k})^T(\vec{x_n} - \vec{\mu_k})}{\sum_{n=1}^{N} u_{n,k}} \tag{A.7}$$

The EM algorithm runs until convergence or until the spot criteria is reached. The mean and covariances matrices derived in the M-steps are used as initial parameters for BSC algorithm.

A.2 Stochastic distances

Theodoridis e Koutroumbas (2010) developed the Bhattacharyya and Kullback-Leibler stochastic distance between multivariate Gaussian distributions. These distances are expressed as:

a) Bhattacharyya

$$d_{G_B}(\vec{x_n}, \vec{x_p}) = \frac{1}{8}(\vec{\mu_n}-\vec{\mu_p})^T \left(\frac{(\boldsymbol{\Sigma}_n^{-1} + \boldsymbol{\Sigma}_p^{-1})}{2}\right)^{-1} (\vec{\mu_n}-\vec{\mu_p}) + \frac{1}{2}\log\left(\frac{\frac{(\boldsymbol{\Sigma}_n^{-1}+\boldsymbol{\Sigma}_p^{-1})}{2}}{\sqrt{|\Sigma_n||\Sigma_p|}}\right) \tag{A.8}$$

b) Kullback-Leibler

$$d_{G_{KL}}(\vec{x_n}, \vec{x_p}) = \frac{1}{2}(\vec{\mu_n}-\vec{\mu_p})^T(\boldsymbol{\Sigma}_n^{-1}+\boldsymbol{\Sigma}_p^{-1})(\vec{\mu_n}-\vec{\mu_p}) + \frac{1}{2}Tr(\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\Sigma}_p\boldsymbol{\Sigma}_n\boldsymbol{\Sigma}_p^{-1} - 2I_q) \tag{A.9}$$

A.3 Multivariate Gaussian entropy

In BSC algorithm, the entropy is used to compute the information gain, which defines the clusters to be splitted. The entropy of the multivariate Gaussian is equal to:

$$H = \frac{1}{2}log((2\pi e)^N det(\boldsymbol{\Sigma})) \tag{A.10}$$

A.4 Results

The image used as input for the Gaussian version of BSC is the San Francisco Landsat 7 (Figure A.1). This images is from the same region as presented in Section 5.1.1.2. The image is composed by man made structures, as the Golden Gate Bridge and the Bay Bridge, and natural features such as the Marin Headlands and the complex currents in San Francisco Bay. This image is provided by the EROS Data Center Landsat 7 Data Handling Facility Image Assessment System (IAS) (PRZYBORSKI, 2020).

Figure A.1 - San Francisco Landsat 7.

SOURCE: Przyborski (2020).

As presented in chapter 6, the best classification results were achieved with the BSC using the RPDDP algorithm, and the Bhattacharyya and Hellinger distances. Therefore, in this section only the BSC-PDDP-B (BSC using the regular PDDP and Bhattacharyya) results are presented.

The classification result of the BSC-PDDP-B algorithm for the Landsat 7 San Francisco, and its confusion matrix are shown in Figures A.2 and A.3. As presented in Figure 5.6(b), the San Francisco area has six main classes: Mountain, Water, Vegetation, High Density Urban (HDU), Low Density Urban (LDU), and Developed Urban (DU). Note that the ALOS/PALSAR data, presented in Figure 5.6(a), were recorded in 2007, and the Landsat 7 image is from 2005, therefore some differences in the land cover can be found.

Figure A.2 - BSC-PDDP-B San Francisco Landsat 7 classification result.

SOURCE: Author's production.

The San Francisco six main classes are presented on the Truth Legend of Figure A.2, while the classes generated by BSC-PDDP-B are described in the Outcome Legend. The BSC-PDDP-B generate eight classes: Mountain/Vegetation, LDU/HDU, New 1, New 2, New3, New4, New5, and DU. As presented in A.3, the Mountain and Vegetation were clustered together, the same happened to LDU and HDU. The DU class was fairly well classified, while the Water class, although it present a good result in the confusion matrix, it was divided into four new classes, as shown in Figure A.2. The BSC-PDDP-B classification result overall accuracy is 76.98 %.

Figure A.3 - BSC-PDDP-B San Francisco Landsat 7 confusion matrix.

SOURCE: Author's production.

Figure A.4 presents the BSC-PDDP-B dendrogram, where the clustering bisecting history can be followed. The San Francisco Landsat 7 image was initially divided into cluster ID2, which gathers the urban classes, i. e., LUD, HDU, and DU; and cluster ID3, which groups the Mountain, Water, and Vegetation classes.

In the sequence, the cluster ID3 was divided into cluster ID4, where the Mountain and Vegetation classes are grouped; and into cluster ID5 where the Water classes samples are clustered. Note, that, differently from SAR images which have effects as the foreshortening and layover, optical images alone do not provide ways to differentiate altitude, therefore the mountain vegetation is note so different from the vegetation from low altitudes, for this reason the BSC-PDDP-B algorithm wasn't able to divide these classes. On the other hand, as can be checked in Figure A.1, the water region has different shades, what influences the splitting result.

Figure A.4 - BSC-PDDP-B San Francisco Landsat 7 dendrogram.

SOURCE: Author's production.

APPENDIX B - BSC WEBPAGE AND HOW RUN THE CODE

The BSC algorithm were developed in Matlab, which is a high-performance language for technical computing. This tool were chosen due to it environment, which integrates computation, visualization, and programming. Also, Matlab has a vast collection of computational algorithms ranging from elementary functions like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix eigenvalues, which were exhaustively explored by BSC algorithm.

As discussed in Chapter 6, the BSC algorithms were evaluated with real and simulated PolSAR images. The Matlab Codes and the set of simulated PolSAR image can be found at the address https://sites.google.com/view/bscalgorithm. Figure B.1 presents the initial page of the BSC algorithm webpage. In this page there are two main icons. The first one is a link for the set of 100 simulate PolSAR images. The second icon is the link to download the BSC algorithm Matlab code.

Figure B.1 - BSC algorithm webpage.

SOURCE: Author's production.

The simulated images are available in three version. The first one is a image with three bands s_{hh}, s_{hv}, and s_{vv}; the second version is the image represented by PolSAR covariance matrix, and the third is the PolSAR coherence matrix.

After the code downloading, the user shall edit the main.m file. Figure B.2 shows the code lines where the edits must be done. The user shall inform the initial parameter algorithm, if EM or RPDDP; the stochastic distance name (Bhattacharyya or Kullback-Leibler, for images following the Gaussian distribution, or Bhattacharyya, Hellinger or Kullback-Leibler, for images following the Wishart distribution); and the image type, if optical or if PolSAR. If optical image are select, the BSC will run the Gaussian distribution version, while, if PolSAR images are chose, the BSC will run the Wishart distribution version. The user shall inform the folder were the input images are and the folder where the classified image and the dendrogram will be saved (folderTosave).

Figure B.2 - BSC main.m file configuration.

```
%% 1. Initial configuration==================================================
level_max = 30;
use_size_weigth = TRUE;
enable_plan_H_alpha = FALSE;
initial_parameter_algo = 'EM'; %'RPDDP'; %
image_type =  'Optical'; %'PolSAR'
L = 5; %Number of Looks;
distance_name = 'Bhattacharyya'; % 'Hellinger';
folder = 'C:\Users\ncarvall\Documents\PersonalCodes\Matlab\Doutorado\Imagen
folderTosave = 'C:\Users\ncarvall\Documents\PersonalCodes\Matlab\Doutorado\

%==========================================================================
```

SOURCE: Author's production.

The user shall inform the input image name. The BSC can handle PolSAR images which are in ENVI standard formats. Therefore the images can have extensions .dat or .bin, and must have a header file, which contains information the image size, data type, byte order, interleave type (bil, bsq, or bip). The optical images must have three bands (RGB).In Figure B.3, the red rectangles show where the image names should be informed. After all configuration be done, the algorithm will aromatically

generate the classified image and dendrogram, and save in the indicated folder.

Figure B.3 - BSC image name.

```
%% 2. Load image==========================================================
if (strcmp(image_type, 'PolSAR'))
    im_filename = 'SIRC_Lband_L5.dat'; %'ALOS_SanFrancisco_L5.dat'; %
    hdr_filename = 'SIRC_Lband_L5.hdr'; %'ALOS_SanFrancisco_L5.hdr'; %
    imcov = openPolSARimage(folder, im_filename, hdr_filename);
    [n_row, n_col, n_band] = size(imcov);
    figure;
    plotPolSARimage( imcov );
    if (n_band == 9)
        q = 3;
    elseif(n_band == 18)
        q = 6;
    end
    pdf_name = 'wishart';
elseif (strcmp(image_type, 'Optical'))
    im_filename = 'san_francisco.jpg';
    address = fullfile(folder, im_filename);
    imcov = double(imread(address))/255;
    figure;
    imagesc(imcov);
    axis equal
    [n_row, n_col, n_band] = size(imcov);
    q = 3;
    pdf_name = 'gaussian';
end
%=========================================================================
```

SOURCE: Author's production.

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.