

Applications of Hyper-spectral and Radar Remote Sensing analysis: a case study of forest landscapes in Costa Rica

Mauricio Vega-Araya

Faculty of Forest Sciences and Forest Ecology

Georg-August University of Göttingen

A thesis submitted for the degree of

Doctor of Philosophy

November 2012

D7

1. Gutachter: Prof. Dr. Christoph Kleinn
2. Gutachter: Prof. Dr. Steffen Kuntz
3. Prüfer: Prof Dr. Alexander Knohl

Tag der mündlichen Prüfung: 12.11.2012

Diese Arbeit ist durch die Niedersächsische Staats- und Universitätsbibliothek, SUB-Göttingen, elektronisch veröffentlicht.

(Dedication)

Dedication

To Paula, Camila and Ludmila.

I would not have been able to complete my doctoral program without their love, support and companionship.

— Göttingen Germany, November 2012 —

Acknowledgements

This research would not have been possible without financial support from DAAD and CONARE through the Alecosta (Costa Rica-Alemania) agreement, the National University of Costa Rica and a teaching assistantship and support from the Chair of Forest Inventory and Remote Sensing from the University of Göttingen.

I am also deeply thankful to the Infoterra GmbH team at Astrium Geoinformation Services, who made possible the three months of an internship in Immentaad headquarters.

I would like to thank CICANUN at University of Costa Rica, who let me use their portable spectrometer. In particular, I would like to thank Professor Dr. Javier Bonatti who designed the portable system for his tremendous support and friendship during all the study process.

Another critical part of this research was the access to the *Gmelina arborea* plantations in the south of Costa Rica. Many thanks to all owners who trusted in me and gave me full access to their plantations. Many thanks also to Henning Aberle and Paul Magdon for his unconditional help and support.

I am very grateful to INISEFOR from the National University and the project "Mejoramiento de la capacidad productiva de pequeños y medianos reforestadores de la zona sur de Costa Rica", for the logistical support in the field campaign.

I also acknowledge PRIAS-CENAT and the CARTA project for providing the HyMap imagery and other invaluable information.

I would like to thank my main academic mentor in my doctoral research, Professor Dr. Christoph Kleinn who taught me to appreciate the importance of field work and its link with remote sensing. He propelled me forward with his knowledge, enthusiasm, confidence and teaching abilities. In addition, I would like to thank my committee member, Professor Dr. Stefen Kuntz, for his academic guidance and confidence in helping me complete my dissertation. I would like to thank Dinora Frutos and Alina Kleinn for the help in English corrections.

(Acknowledgements)

Thank you to my parents, Edwin Vega Barrantes and Maria Virginia Araya Gonzalez, as well as my brothers Edwin Eduardo and Andrea Maria. Also my parents in love, Juan Frutos and Ludmila Svatek, all of them have been a solid pillar of emotional support and a wealth of wisdom during my entire adventure in Germany.

I would like to thank past and present INISEFOR and National University staff, especially Lic. Orlando Chinchilla, Lic. Sandra León and Dr. Olman Segura, for their assistance in helping bring my research to fruition. I am greatly indebted to the Junta de Becas from the National University staff, who implemented the Alecosta agreement.

I also thank my field workers in the south of Costa Rica for their friendship, comprehension, availability and help.

— Göttingen Germany, November 2012 —

Abstract

The main objective this research was to assess two types of emerging remote sensing technology, hyper-spectral and *SAR* sensors, for an exploratory data analysis of land covers in the south of Costa Rica. Hyper-spectral data contain information in several narrow spectral bands in the optical domain, which give information on the biochemical and structural properties of vegetation, while the *SAR* data, as an active system, can penetrate the clouds making it a promising tool for ecosystem monitoring. The main hypothesis was that these two datasets would permit greater understanding of the spectral confusion between different land covers. From the hyper-spectral point of view, this knowledge could help to select and derive spectral signatures which serve as training data sets in supervised species classification in the optical domain. In the microwave domain, fusion and derived bands increase the separability and permit greater forest/non-forest classification accuracy in non-flat terrains.

The hyper-spectral information is based on two information sources. The first comes from two scenes of the space-borne Earth Observing-1 mission with the Hyperion sensor and two scenes of the airborne hyper-spectral sensor *HyMap*. The second hyper-spectral data source was acquired from the field-based hyper-spectral clip-prove system. Furthermore, the microwave information corresponds to the TerraSAR-X *HH* and *VV* polarized images.

Working in different land covers including *Gmelina arborea* plantations in the south of Costa Rica, individual Regions of Interest were manually digitized with reference to high spatial resolution aerial photographs datasets. Principal components of hyper-spectral space and airborne data were derived to perform a classification using two different approaches of Hierarchical Cluster Analysis.

Spectra from field-based hyper-spectral clip-prove data was acquired from *Gmelina arborea* leaves in three plantations of 6, 8 and 18 years. Other reference spectra of land covers were also measured.

With seven TerraSAR-X polarized images, principal component analysis as fusion technique and derived bands ratios were generated in order to evaluate the availability of reducing the speckle noise in non flat terrains to classify forest in

the south of Costa Rica.

The highest scene based spectra variability was in the Near Infra-Red portion of the electromagnetic spectrum. Hierarchical Cluster Analysis applied to the hyper-spectral scenes showed that cluster solutions of the PCs spectra from the two sensors present different separability solutions. The clusters solutions were subject to systematic differences; only one scene of EO-1 Hyperion and one of *HyMap* PCs spectra did not present spectral confusion among *Gmelina arborea*, palm oil and the forest. That indicates that the same sensor under different conditions will give different spectra and different cluster results. These results suggested that hyper-spectral imagery need not to be acquired at a very high spatial resolution to provide adequate discrimination of land covers. Furthermore, spectra collection and analysis are needed to acquire time series spectral signatures.

The best Hierarchical Cluster Analysis classification was with the Approximately Unbiased *p-values* which permit the identification of clusters that exist at a pre-defined level of significance.

Canopy phenology, a property related to the different acquisition times and atmospheric conditions, was important in clustering land covers.

Regarding the field based spectra, there was spectral confusion in the majority of 18 years of leaves of *Gmelina arborea* and mangrove. Also, 6 spectra of this age were not clustered at all. There was spectral confusion between the spectra of *Gmelina arborea* leaves of 6, 8 and 18 years. However, the reflectance of field based spectrometers should be interpreted with caution. Sampling is a key factor as well as a challenge in leaf spectral analysis.

Hyper-spectral and Synthetic Aperture Radar data was useful for land cover discrimination, and it did provide an unprecedented potential to classify forest and non-forest in tropical environments and avoid spectral confusion with highly related land covers. However, all the associated variability of acquisition parameters has to be taken into account in order to provide acceptable levels of accuracy.

Zusammenfassung

Die Zielsetzung dieser Forschungsarbeit war die Beurteilung zweier sich noch in der Entwicklung befindenden Technologien der Fernerkundung, hyperspektrale und Synthetic Aperture Radar (SAR) Sensoren, im Rahmen einer untersuchenden Datenanalyse von Landbedeckungen im Süden Costa Ricas. Die hyperspektralen Daten enthalten Informationen in vielen schmalen spektralen Bändern im optischen Bereich, die Aufschluss über die biochemischen und strukturellen Eigenschaften von Vegetation geben können. Dagegen können SAR-Sensoren, als aktive Systeme, Wolken durchdringen und stellen somit ein viel versprechendes Werkzeug für das Monitoring von Ökosystemen dar. Die Haupthypothese war, dass diese beiden Datensätze ein besseres Verständnis der spektralen Vermischung unterschiedlicher Landbedeckungsklassen erlauben. Vom hyperspektralen Standpunkt aus könnte dieses Kenntnis dabei helfen spektrale Signaturen zu selektieren und abzuleiten, welche als Trainingsdaten einer überwachten Artenklassifizierung im optischen Bereich dienen. Im Mikrowellenlängenbereich verbessern Fusion und abgeleitete Bänder die Unterscheidbarkeit und erlauben im nicht flachen Terrain höhere Genauigkeiten bei der Klassifizierung von Wald/Nicht-Wald.

Die hyperspektralen Informationen basieren auf zwei Quellen. Bei der ersten handelt es sich um zwei Szenen der EO-1 Mission des Satelliten Hyperion und je zwei Szenen des luftgetragenen hyperspektralen Sensors HyMap. Der zweite hyperspektrale Datensatz stammt von einem feldbasierenden System. Ferner entsprechen die Mikrowellen-Information TerraSAR-X *HH* und *VV* polarisierten Bildern.

Es wurden unterschiedliche Landbedeckungen im Süden Costa Ricas einschließlich *Gmelina arborea*-Plantagen untersucht, wobei einzelne Klassen manuell mit Hilfe von hochaufgelösten Luftbildern identifiziert und digitalisiert wurden. Hauptkomponenten der hyperspektralen Bilder wurden verwendet, um zwei verschiedene Ansätze der hierarchischen Clusteranalyse durchzuführen. Zudem wurden mit einem Feldspektrometer Spektren von *G. arborea*-Blättern in drei Plantagen im Alter von 6, 8 und 18 Jahren aufgenommen. Auch wurden Spektren anderer Landbedeckungen gemessen.

Aus sieben polarisierten TerrSAR-X Bildern wurde eine Hauptkomponentenanalyse als Fusionstechnik durchgeführt sowie Bandverhältnisse generiert, um die Möglichkeit der Reduzierung des Rauschens im nicht flachen Gelände abzuschätzen und Wald im Süden von Costa Rica zu klassifizieren.

Die höchste bildbasierte spektrale Variabilität gab es im nahen Infrarotbereich des elektromagnetischen Spektrums. Die hierarchische Clusteranalyse, angewendet auf die hyperspektralen Szenen, zeigte, dass Cluster-Lösungen basierend auf den Hauptkomponenten der zwei Sensoren unterschiedliche Trennbarkeiten liefern. Die Cluster-Ergebnisse bilden den Gegenstand systematischer Unterschiede; nur eine Szene von EO-1 Hyperion und eines der Hymap Hauptkomponentenbilder zeigten keine spektrale Vermischung von Gmelina arborea, Ipalmen und Wald. Das weist darauf hin, dass derselbe Sensor unter unterschiedlichen Bedingungen verschiedene Spektren und andere Cluster erzeugt. Diese Ergebnisse deuten an, dass hyperspektrale Bilder nicht mit einer hohen räumlichen Auflösung akquiriert werden müssen, um eine gute Unterscheidung von Landbedeckungen zu erreichen. Weiter werden Aufnahmen und Analysen von spektralen Signaturen im Rahmen von Zeitreihen benötigt.

Die beste Klassifizierung aus den hierarchischen Clusteranalysen ergab sich mit den Approximately Unbiased p-Werten, welche eine Identifikation von Clustern erlauben, die bei einem definierten Signifikanzlevel existieren. Die Kronenphänologie, abhängig von den unterschiedlichen Aufnahmezeitpunkten und atmosphärischen Bedingungen, war wichtig beim Clustern von Landbedeckungen.

Bezüglich der im Feld gemessenen Daten gab es Vermischung von Blattspektren vor allem bei den 18 jährigen Individuen von *G.aborea* und Mangroven. Auch konnten sechs Spektren dieser Altersklasse keinem Cluster zugeordnet werden. Die Reflexionswerte des Feldspektrometers sollten mit Vorsicht interpretiert werden. Sampling ist dabei genauso ein Schlüsselpunkt wie die Herausforderung der Analyse von Blattspektren.

Hyperspektrale und Synthetic Aperture Radar Daten stellten sich als nützlich für die Unterscheidung von Landbedeckungen heraus. Zudem bieten sie ein beispielloses Potential für die Klassifizierung von Wald und Nicht-Wald. Dennoch sollte

(Zusammenfassung)

die Variabilität der Aufnahmeparameter beachtet werden, um ein akzeptables Niveau der Genauigkeit zu liefern.

— Georg-August Universität Göttingen, November, 2012 —

Contents

Dedication	i
Acknowledgements	ii
Abstract	iv
Zusammenfassung	vi
Contents	ix
List of Figures	xii
List of Tables	xv
Acronyms and abbreviations	xviii
1 Introduction	1
1.1 Introduction and Problem Statement	1
1.2 Objectives and Hypotheses	3
1.3 Study Area	5
1.3.1 Geographic Location	5
1.3.2 Climate and Vegetation	5
1.3.3 Principal Land Covers of the Study Site	9
2 Literature Review	12
2.1 Forest Monitoring	12
2.1.1 Forest Measuring	14
2.1.2 Vegetation and Climate	16

2.2	Remote Sensing and Earth Observation Monitoring Systems	17
2.3	Hyper-Spectral Data	20
2.3.1	Atmospheric Corrections	23
2.4	RADAR Data	26
2.4.1	Principles of RADAR Data Processing	29
2.4.1.1	RADAR Relief Displacement, Shadows and Speckle	29
2.4.1.2	Surface Roughness and Electrical Characteristics	29
2.4.2	Texture Quantization and Image Segmentation	31
2.5	Data Reduction and Fusion	32
2.6	Cluster Analysis and Linear Discrimination Analysis	34
3	Methodology	38
3.1	Hyper-Spectral	38
3.1.1	EO-1 Hyperion	40
3.1.2	HyMap	43
3.1.3	Selection of the Regions of Interest	44
3.2	The USB650 Field Spectrometer Measurements	45
3.3	Statistical Analysis of Hyper-spectral Data	50
3.3.1	Principal Component Analysis of Hyper-spectral Data . .	51
3.3.2	Hierarchical Cluster Analysis	52
3.4	TerraSAR-X Image Processing	57
3.4.1	The TerraSAR-X Imagery Description	58
3.4.2	TerraSAR-X Data Pre-processing	59
3.4.2.1	Radiometric Correction	59
3.4.2.2	Speckle Reduction	61
3.4.2.3	Terrain Correction	62
3.4.2.4	Selection of TerraSAR-X Processed Bands	64
3.4.3	Image Classification of TerraSAR-X Processed Bands and the RapidEye imagery description	64
3.4.4	Analysis of the Land Cover Distributions	66
4	Results	68
4.1	Hyper-Spectral Analysis from the Four Scenes	68

4.1.1	The Scene’s Spectra	68
4.1.2	The hy_15 scene	73
4.1.3	The hy_14 scene	77
4.1.4	The eo1_pz scene	80
4.1.5	The eo1_kw scene	83
4.1.6	Specific clusters error evaluation	87
4.2	USB650 Field Spectrometer	89
4.3	TerraSAR-X Processing and Land Cover Analysis	95
5	Discussion	107
5.1	Hyper-Spectral Sensors and General Considerations on the Clus- tering Process	107
5.2	The USB650 Field Spectrometer	111
5.3	TerraSAR-X Processing and classification	114
5.4	Climate, Vegetation and Remote Sensing in Costa Rica	120
6	Conclusions	126
6.1	Hyper-spectral Conclusions	126
6.2	TerraSAR-X Conclusions	129
6.3	Remote Sensing and Climate	130
	Bibliography	132
	Appendices	150
A	Climate and Vegetation in Costa Rica	151
A.1	Normalized Difference Vegetation Index	151
A.2	Precipitation	152
A.3	Temperature	153
A.4	Potential Evapo-transpiration	154
A.5	Reconnaissance Drought Index	154
B	Slopes of the Study Area	156
C	List of the Clusters and Edges	157

C.1	Clusters and Edge Number of the Dendrogram of the <i>hy-15</i> PCs Spectral Signatures.	157
C.2	Clusters and Edge Number of the Dendrogram of the <i>hy-14</i> PCs Spectral Signatures.	158
C.3	Clusters and Edge Number of the dendrogram of the <i>eo1-pz</i> PCs Spectral Signatures.	160
C.4	Clusters and Edge Number of the Dendrogram of the <i>eo1-kw</i> PCs Spectral Signatures.	161
C.5	Clusters and Edge Number of the Dendrogram of the USB650 Spectral Signatures.	162
D	Conditional Density Plots for All Bands and the Different Land Covers	164
E	Comparison of the Covers by TerraSAR-X Processed Bands	170
F	Comparison of the Covers by TerraSAR-X Processed Bands	171

List of Figures

1.1	Study site in south of Costa Rica.	6
1.2	Mean values of precipitation, potential evapo-transpiration and Normalized Difference Vegetation Index for the study site.	7
1.3	Original RapidEye true colour composite showing the cloud problems in the study area and the systematic sampling points.	8
2.1	Spectral curves from Lawn Grass and Juniper Bush spectra and the relative spectral response for Landsat ETM+ and MODIS sensors.	23
2.2	Principles of typical RADAR system components	28
2.3	Steps of clustering process	36

List of Figures (List of Figures)

3.1	Overview of the hyper-spectral scenes in the south of Costa Rica	39
3.2	Segment of colour composite of the EO1 Hyperion images	41
3.3	Segment of colour composite of the of the HyMap images	44
3.4	Crown sampling scheme in a tree.	47
3.5	<i>USB650</i> Field Spectrometer.	48
3.6	A schematic diagram of the general approach to perform the hierarchical cluster analysis and the similarity discrimination.	51
3.7	TerraSAR-X work-flow in south of Costa Rica.	57
3.8	TerraSAR-X polarization modes in south of Costa Rica.	60
4.1	Spectral Signatures in the <i>hy_15</i> scene.	69
4.2	Spectral signatures in the <i>hy_14</i> scene.	70
4.3	Spectral signatures in the <i>eo1_pz</i> scene.	71
4.4	Spectral Signatures in the <i>eo1_kw</i> scene.	71
4.5	Silhouette graphs for the <i>PCs</i> spectral signatures of the scenes analysed.	72
4.6	Average silhouette width for different number of clusters solutions	72
4.7	Hierarchical cluster analysis and silhouette graph of 83 <i>PCs</i> spectra of <i>hy_15</i> scene.	74
4.8	Approximately unbiased (AU) <i>p-value</i> of 83 <i>PCs</i> spectral signatures of <i>hy_15</i> scene.	76
4.9	Hierarchical cluster analysis and silhouette graph of 95 <i>PCs</i> spectral signatures of <i>hy_14</i> image.	79
4.10	Approximately Unbiased (AU) <i>p-value</i> of 95 <i>PCs</i> spectral signatures of <i>hy_14</i> image.	80
4.11	Hierarchical cluster analysis and silhouette graph of 73 <i>PC</i> spectral signatures of <i>eo1_pz</i> scene.	82
4.12	Approximately unbiased (AU) <i>p-value</i> of 73 <i>PC</i> spectral signatures of <i>eo1_pz</i> scene.	83
4.13	Hierarchical cluster analysis and silhouette graph of 79 <i>PCs</i> spectral signatures of <i>eo1_kw</i> scene.	85
4.14	Approximately unbiased (AU) <i>p-value</i> of 79 <i>PC</i> spectral signatures of the <i>eo1_kw</i> scene.	86

List of Figures (List of Figures)

4.15	Approximately unbiased (<i>AU</i>) <i>p-value</i> and standard error for the four scenes	88
4.16	Mean \pm standard deviation of reflectance for leaves of three different <i>edges</i> of <i>G.arborea</i> plantations in the south of Costa Rica. . .	90
4.17	Silhouettes of a clustering with $k=14$ of 122 spectral signatures measured from the <i>USB650</i> field spectrometer.	91
4.18	Dendrogram with $k=14$ cut height of 122 spectral signatures measured from the <i>USB650</i> field spectrometer.	92
4.19	Approximately unbiased (<i>AU</i>) <i>p-value</i> of 122 <i>PCs</i> spectral signatures of <i>USB650</i> field spectrometer.	93
4.20	Approximately unbiased (<i>AU</i>) <i>p-value</i> and the standard error for the <i>USB650</i> spectral signatures	94
4.21	Segment of TerraSAR-X processed <i>HH</i> intensity image in the south of Costa Rica	96
4.22	Linear density graphs for the <i>ROIs</i> of forest in undulating and rolling slopes for the 9 bands	98
4.23	Example of TerraSAR-X Red (band B8), Green (band B5) and Blue (band B1) composite in a segment of the study area in the south of Costa Rica.	99
4.24	Conditional density plot for all bands for the land cover <i>forest</i> in flat terrain (<i>raffia</i>).	100
4.25	Example of TerraSAR-X penetration in soils in a segment of the study area in the south of Costa Rica.	102
4.26	Unsupervised Classification of the TerraSAR-X imagery in the study area.	106
B.1	Slopes of the TerraSAR-X study area.	156
D.1	Conditional density plot all bands for the land cover <i>forest</i> in gently undulating slopes.	164
D.2	Conditional density plot for all bands the land cover <i>forest</i> in rolling slopes.	165
D.3	Conditional density plot for all bands for the land cover <i>herbland</i> in flat slopes.	165

D.4	Conditional density plot for all bands for the land cover <i>infrastructure</i>	166
D.5	Conditional density plot for all bands for the land cover <i>mangrove</i>	166
D.6	Conditional density plot for all bands for the land cover <i>palm oil</i> (old plantation).	167
D.7	Conditional density plot for all bands for the land cover <i>palm oil</i> (young plantation).	167
D.8	Conditional density plot for all bands for the land cover <i>pasture</i>	168
D.9	Conditional density plot for all bands for the land cover <i>bare soil</i>	168
D.10	Conditional density plot for all bands for the land cover <i>water</i>	169
F.1	Conditional density plot for the Band 1 and all land covers.	171
F.2	Conditional density plot for the Band 2 and all land covers.	172
F.3	Conditional density plot for the Band 3 and all land covers.	172
F.4	Conditional density plot for the Band 4 and all land covers.	173
F.5	Conditional density plot for the Band 5 and all land covers.	173
F.6	Conditional density plot for the Band 6 and all land covers.	174
F.7	Conditional density plot for the Band 7 and all land covers.	174
F.8	Conditional density plot for the Band 8 and all land covers.	175
F.9	Conditional density plot for the Band 9 and all land covers.	175

List of Tables

1.1	Slopes of the study area.	8
2.1	Characteristics of some selected <i>SAR</i> sensors	28
3.1	Codification and number of selected regions of interest per hyper- spectral scenes.	40

List of Tables (List of Tables)

3.2	Codification used and acquisition dates for to the four hyper-spectral scenes	41
3.3	Codification of selected covers for the <i>USB650</i> hyper-spectral field spectrometer.	49
3.4	Characteristics of Optical Sensors	50
3.5	Interpretation key for the Silhouettes values.	55
3.6	Interpretation key for the average silhouette width.	56
3.7	Characteristics of TerraSAR-X mission products	59
3.8	Acquisition parameters for <i>Stripmap</i> TerraSAR-X imagery.	60
3.9	Multilook parameters of the TerraSAR-X imagery.	62
3.10	Selected regions of interest for the density distribution analysis of the TerraSAR-X process.	66
4.1	Hierarchical clustering parameters for the four remote base sensor spectral signatures	73
4.2	Statistics for the approximately unbiased <i>p-values</i> and the <i>SE</i> for the <i>PCs</i> of the <i>hy_15</i> scene data	75
4.3	Statistics for the Approximately Unbiased <i>p-values</i> and the <i>SE</i> for the <i>PCs</i> of the <i>hy_14</i> scene data	81
4.4	Statistics for the Approximately unbiased <i>p-values</i> and the <i>SE</i> for the <i>PCs</i> of the <i>eo1_pz</i> scene data	84
4.5	Statistics for the approximately unbiased <i>p-values</i> and the <i>SE</i> for the <i>PCs</i> of the <i>eo1_kw</i> scene data	87
4.6	Approximately unbiased <i>p-values</i> for the two highest standard error per scene.	88
4.7	Statistics for the approximately unbiased <i>p-values</i> and the <i>SE</i> for the <i>USB650</i> field spectrometer data	94
4.8	Approximately unbiased <i>p-values</i> for the three <i>edges</i> with the highest standard error in the <i>USB650</i> field spectrometer data	94
4.9	TerraSAR-X selected bands.	97
4.10	Summary area and percentage per class of the unsupervised classification of the TerraSAR-X imagery in the study area	105

List of Tables (List of Tables)

4.11 Error matrix of forest/non-forest classes for the classification of TerraSAR-X imagery in south of Costa Rica.	105
E.1 Comparison of the covers by TerraSAR-X processed bands.	170

Acronyms and abbreviations

α_λ	Absorbance
λ	wavelength
ρ_λ	Reflectance
τ_λ	Transmittance
ALOS-2	Advanced Land Observing Satellite-2
ASAR	The SAR on board on ENVISAT
ATREM	ATmospheric REMoval program
AU	Approximately Unbiased Test
AVIRIS	Advanced Very High Resolution Radiometer
BIL	Band-Interleaved-by-Line
BIP	Band Interleaved by Pixel
CARTA Mission	Costa Rica Airborne Research and Technology Application
CBM	Convention on Biological Diversity
CICANUM	Research Centre on Nuclear and Molecular Sciences, University of Costa Rica

(Acronyms and abbreviations)

CosmoSkyMed	Constellation of Small Satellites for Mediterranean Basin Observation
DBH	Diameter at Breast Height
DEM	Digital Elevation Model
DEM	Digital Elevation Model
DLR	German Aerospace Centre
DN	Digital Number
EEC	Enhanced Ellipsoid Corrected
EnMap	Environmental Mapping and Analysis Program
ENVISAT	ENVironmental SATellite
EO-1 Hyperion	High Resolution Hyper-spectral push-broom imager on board of Earth Observing-1 mission
ERS	Earth Remote Sensing Satellite
ESA	European Space Agency
FAO	Food and Agriculture Organization of the United Nations
FCPF	Forest Carbon Partnership Facility
FLAASH	Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes
FRA	Global Forest Resources Assessment of FAO
FWHM	Full Width at Half Maximum
G.arborea	<i>Gmelina arborea</i>

(Acronyms and abbreviations)

GCP	Ground Control Point
GEC	Geocoded Ellipsoid Corrected
GeoTIFF	Geographic Referenced Tagged Image File Format
GIS	Geographic Information Systems
GIS	Geographical Information Systems
HCA	Hierarchical Cluster Analysis
HH	Horizontal-Horizontal polarization
HyMap	Airborne Hyperspectral Sensor
I.D.I.O.T	InSAR Deformation Inspection and Observation Tool
IEMR	Incident Electromagnetic Radiation
IFOV	Instantaneous Field of View
Infoterra GmbH	A subsidiary of EADS Astrium which is responsible for the commercial marketing of the TerraSAR-X products, nowadays Astrium GEO-Information Services
Landsat 7 ETM+	Landsat 7 Enhanced Thematic Mapper Plus
LDA	Linear Discrimination Analysis
LULC	Land Use and Land Cover
MGD	Multilook Ground Range Detected
MODIS	MODerate resolution Imaging Spectroradiometer
MRV	Measurement, Reporting and Verification
NASA	National Aeronautics and Space Administration

(Acronyms and abbreviations)

NDVI	Normalized Difference Vegetation Index
NEST	Next ESA Toolbox
NFIs	National Forest Inventories
NIR	Near-Infra-Red
NPP	Net Primary Production
PAZ	The first Spanish RADAR observation satellite
PCs	Principal Components
PET	Potential Evapo-transpiration
PRCR	Costa Rican Cadastre and Register Regularization Program
PRE	Precipitation
QGIS	Quantum GIS is a cross-platform free and open source desktop geographic information systems
R	R is a language and environment for statistical computing and graphics
RADARSAT-2	Canada's next-generation commercial radar satellite
RapidEye	Space Borne Satellite Image Products for Geo-Information Solutions
RDI	Reconnaissance Drought Index
REDD+	Reduction of Emissions from Deforestation and Forest Degradation
ROIs	Regions of Interest

(Acronyms and abbreviations)

SAR	Synthetic Aperture Radar
SE	Standard Error
Sentinel-1	ESA SAR C-band satellites
Sentinel-2	ESA satellites with high-resolution optical capabilities
SPOT	Satellites Pour l'Observation de la Terre or Earth-observing Satellites
SRTM	Shuttle Radar Topography Mission
SSC	Single Look Slant Range Complex
SWIR	Short-Wave Infra-Red
TanDEM-X	TerraSAR add-on for Digital Elevation Measurements
TerraSAR-X	Operational Advanced SAR-satellite system for scientific and commercial applications
UCR	University of Costa Rica
UNA	National University of Costa Rica
USB650	USB650 RedTide Fibre Optic Spectrometer from Ocean Optics
UTM	Universal Transversal de Mercator
VEGETATION	SPOT VEGETATION ten daily synthesis archive
VHR	Very High Resolution
VIS	Visible region of the spectrum
VI _s	Vegetation Indices

(Acronyms and abbreviations)

VNIR	Visible and Near-Infra-Red
VV	Vertical-Vertical polarization
WGS84	World Geodetic System

Chapter 1

Introduction

1.1 Introduction and Problem Statement

The global climate change is a fact with an international impact that has been discussed extensively within the framework of international agreements such as the Kyoto Protocol, the Convention on Biological Diversity (CBM) and the United Nations Convention on Climate Change. The need for scientific information is very time sensitive due to the impacts these changes have on the different ecosystems including human environments.

The ability to predict climate change and climate-change related impacts critically depends on understanding the role vegetation plays in modifying changes in the concentration of atmospheric CO₂ and the global energy balance [[Jones and Vaughan, 2010](#)].

The CO₂ concentration levels are a global issue, and it does not matter where they are produced because they affect the entire planet. From this perspective the international community is responsible for maintaining the equilibrium of CO₂ concentrations. Currently, only a multinational monitoring system can provide the scientific information needed on global climate change, carbon sequestration, forest growth and biodiversity [[Lorentz et al., 2005](#)].

The Global Forest Resources Assessment (FRA) of the Food and Agriculture Or-

ganization of the United Nations (FAO), is the most important initiative that provides quantitative and qualitative global forest information. This information combines data provided by several countries, however, the *FAO* uses supplementary remote sensing data due to the inconsistencies, availability and quality of the data provided [FAO, 2001]. Among these efforts, there are two main aspects that are worth mentioning, first, the needed continuously gather forest related information, because climate is not static, hence the idea of monitoring. The second aspect is the need standardize definitions in order to make the reports comparable in time and consistent between the different countries.

The basic question of, "what constitutes a forest?" has been discussed by the international community, however, no consensus has been achieved. No solution has been reached for a harmonized definition of forest, not even a definition which could be implemented for the needs of remote sensing. Similarly, it would be important to have a definition that differentiates forest and other land cover types on various scales. For example, it would be important to make a distinction at the regional level (a region within a country), country level and even at the global level.

Currently, only *FAO* uses a global standard definition of forest [Fagan and DeFries, 2009]. In Appendix 2 of the terms and definitions for the national reporting tables for *FRA* 2005 can be read: "Land spanning more than 0.5 hectares with trees higher than 5 metres and a canopy cover of more than 10 percent, or trees able to reach these thresholds *in situ*. It does not include land that is predominantly under agricultural or urban land use" [FAO, 2006]. The lack of standard, acceptable and appropriate definition for forest cover makes it difficult to create forest maps with standard remote sensing methods. However, and based on the case-study of Costa Rica Kleinn et al. [2002] "do not believe that the problem can be resolved in a homogenization-of-definitions manner on a global level [Kleinn et al., 2002]".

The use of remote sensing to provide complementary forest measurements is needed because there are gaps in data from less accessible regions of some countries or entire countries, especially countries with several economy or social problems.

In forest mapping and monitoring, remote sensing has been proven to be a well-established method [Fagan and DeFries, 2009; Kuntz, 2010], however, there are several challenges. These challenges can be categorized into technical and logistic issues. One technical challenge could be because the selection of the imagery is not guided by scientific principles. For example, limitations in the budget in a specific project or the availability of a set of cloud free imagery in a specific region.

Under a global perspective, Fagan and DeFries [2009] pointed out that it is difficult to distinguish between different types of forest (i.e. between primary and secondary forest or tree plantations) with actual remote sensing imagery. It is also challenging to detect forest degradation in which a forest is partially cleared by human activity [Fagan and DeFries, 2009]. However, at country level, new Very High Resolution (VHR) imagery can supply detailed information which help to distinguish between forests. For example, the Pléiades platform¹, with 50-cm resolution and up to 100 km x 100 km in strip mapping mode, can archive the information needed to map and monitor the vegetation and distinguish between the different types.

Despite the limitations of the technique, principally at global scale, it is recognized that remote sensing can provide the spatial information needed to manage ecosystems in a time and cost effective manner.

1.2 Objectives and Hypotheses

This study was conducted against the background of requiring Forest Area Maps and Remote Sensing as a tool to measure, report and verify the related information in the context of international processes. The two main objectives were to find the potential of hyper-spectral data to differentiate within and between species and other land cover groups. The second aim was to evaluate the use of TerraSAR-X data for the purpose of identifying forest classes in an adaptation of

¹<http://www.astrium-geo.com/en/52-pleiades-very-high-resolution-satellite-imagery>. Access 20.09.2012

the (*FAO*) forest definition of classes forest/non-forest.

Hyper-spectral specific objectives

1. To evaluate the potential of two different data sensors (one space borne and the other air-borne) to discriminate between and identify different land cover groups in the southern region of Costa Rica.
2. To ascertain the difference in reflectance in the visible and near-infrared (NIR) radiation of different *Gmelina arborea* plantations in the southern region of Costa Rica, using a field-base spectrometer.

Synthetic Aperture Radar specific objectives

1. By fusion of two polarization modes of Single Look Slant Range Complex TerraSAR-X imagery, to evaluate the possibility of reducing the speckle noise in non flat terrains to classify forest in the south of Costa Rica.

Additionally, in this study, the inter-annual climate variability is assessed to analyse the influence and the relationship of the climate on the vegetation based on the Normalized Difference Vegetation Index *NDVI* in sub-climatic regions in Costa Rica.

The given goals can be translated into the following research questions or hypotheses:

1. Are there spectral confusions in the classification of HyMap and EO-1 Hyperion sensors for some of the most important land covers in the south of Costa Rica.
2. Is it possible to develop suitable methods to increase the potential of TerraSAR-X in target recognition for monitoring land covers as an option for continuous monitoring in humid tropical areas with high cloud coverage.

By understanding the spectral patterns and detecting the spectral confusion from the hyper-spectral and microwave domain in different land covers in the south of Costa Rica, better methods for classifying forest and non-forest are studied

1.3 Study Area

1.3.1 Geographic Location

Costa Rica is located between Nicaragua and Panama on the Central American isthmus with an area of only 51.100 km². The country has a diverse topography (with elevations from 0 to 3820 m.a.s.l), climatic conditions (e.g. precipitation from 1400 to 8000 mm/a, averaged 3297 mm/a), vegetation forms (from semi-deciduous dry forests to tropical moist forests and mountain cloud forests), and land use systems (intensive monoculture agriculture like sugar cane and banana, and highly diverse agro-forestry systems). Costa Rica's land area represents just about 0.03% of all land on Earth but this area houses around 5-6 % of all documented species in the world. There are two mountain systems that run from the north western region to the south eastern one, dividing the country in three general areas: the so called Pacific influence, the Caribbean influence, and the Central Valley. The mountain system and the impact of the oceans define the climatic regions of Costa Rica.

1.3.2 Climate and Vegetation

The southern region site, located in Central America, Costa Rica (Figure 1.1 a and b), consists of three *cantones* (or counties) called Osa, Corredores y Golfito. The lowlands are dominated by big farms with various forms of agricultural use e.g. rice, melon and palm oil. There are some forest plantations of teak and *Gmelina arborea* (G.arborea) in the lowlands.

In the regions of average slopes, small farmers produce a diversity of products such as plantain, yucca and recently palm oil. On the steep slopes there are some remaining patches of primary and secondary forests. On the coast there are large areas covered by mangroves.

The study site (which is in the South Pacific Climatic Region) has monthly precipitation that varies between 69 mm in February and 597 mm in October, the



Figure 1.1: Study site in south of Costa Rica.

mean value is 300 mm. There is a decrease in rainfall in the months of January, February, March and April. Also in those months, there is a rise in the values of potential evapo-transpiration the maxima being in April with 115 mm and the minima in December with 95 mm, the mean value of potential evapo-transpiration is 106 mm (Figure 1.2 after Vega-Araya et al. [2012]). The minimum value of *NDVI* is 0.6 in October and the maximum is 0.71 in January. Despite this decrease in October, the *NDVI* in this zone is more or less stable with a standard deviation of 0.033 units of *NDVI* [Vega-Araya et al., 2012].

Figure 1.2 is a summary of the results computed for the south pacific climatic region following the methodology presented in Appendix A. The variables presented in this graph are: *PRE* is precipitation, *PET* is potential evapo-transpiration and *NDVI* is the normalized difference vegetation index.

Table 1.1 list the types of slopes the study area for the TerraSAR-X analysis. 48.6 % of the slopes are classified as moderate steep and steep. Most of the plantations in this region are flat, especially the palm oil plantations, however, recently with the rise of international prices of palm oil, some small farmers have planted the palm in undulating slopes. Another practice is to drain the flat lowlands which traditionally are swampy areas in order to establish more palm oil plantations.

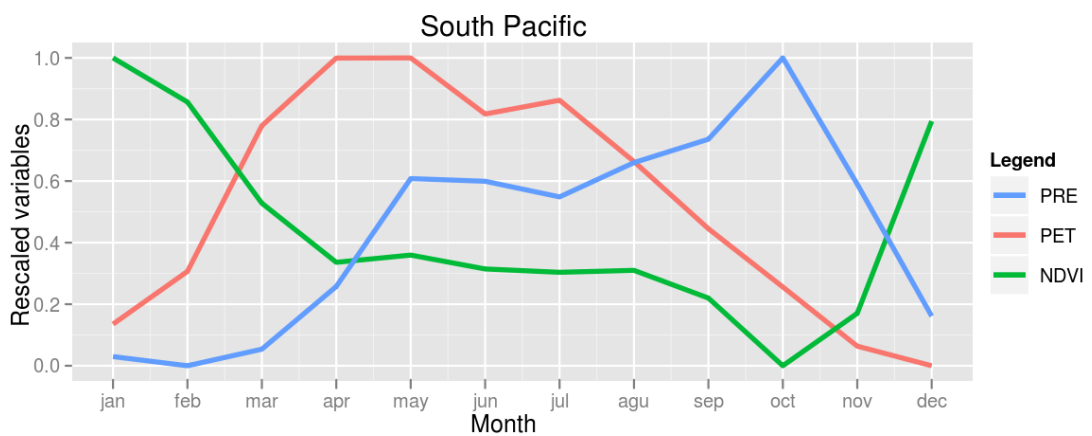


Figure 1.2: Mean values of precipitation (PRE), potential evapo-transpiration (PET) and Normalized Difference Vegetation Index (NDVI) for the study site [Vega-Araya et al., 2012].

Figure 1.3 is a mosaic of four images in a true colour composite of high spatial (5m) resolution RapidEye optical sensor. Most of the raw images used in this study were taken between 3 and 4 pm (UTC). During this hours there is an inherent presence of clouds in Costa Rica as is the case for tropical areas in general, even in the months of February and March, where there is a decrease in the rate of the precipitations.

Table 1.1: Slopes of the study area.

Slope category	Range(%)	Area (ha)	Area (%)
Flat	0-3.5	14237.15	27.2
Gently undulating	3.5-7	4050.77	7.7
Undulating	7-12.3	4484.34	8.6
Rolling	12.3-17.6	3835.01	7.3
Moderate steep	17.6 - 36.4	11599.36	22.1
Steep	36.4 - 120	13901.13	26.5
Highly steep	> 120	193.46	0.4
no data	–	116.40	0.2
Total	–	52417.60	100

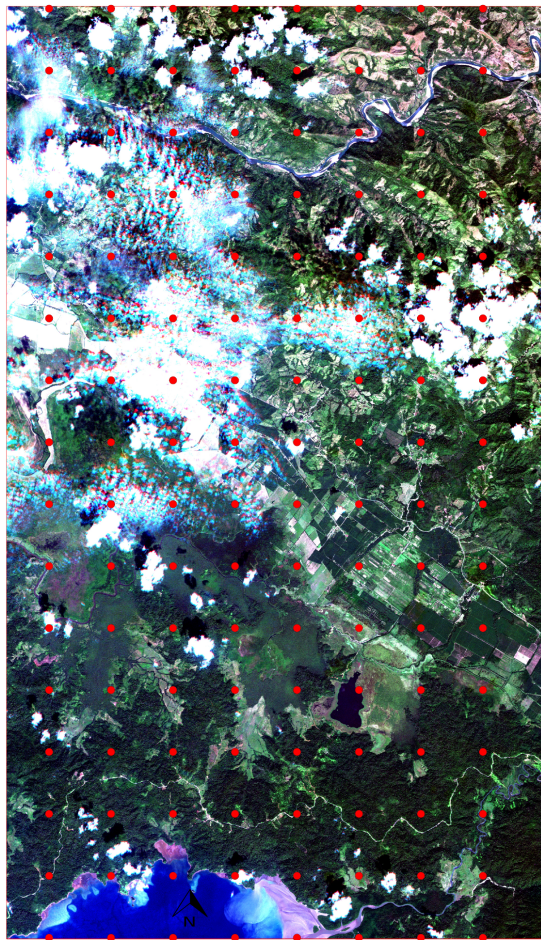


Figure 1.3: Original RapidEye true colour composite showing the cloud problems in the study area and the systematic sampling points.

1.3.3 Principal Land Covers of the Study Site

In the study area there's a total of 21 natural ecosystems. From those, 5 are different types of dense tropical forest, 7 are different types of dense tropical shrublands, 8 types of dense tropical herblands, and one water body [Kappelle et al., 2002]. However, some of the differences between the ecosystems, are due to the altitudinal distribution.

This section describes the main land cover groups which are useful to select the different Regions of Interest (*ROIs*). As will be explained in Chapter 3, different *ROIs* were selected for the hyper-spectral scenes analysis and for the TerraSAR-X image processing.

The description below of a three level grouping system could be useful for a posterior classification scheme in the south of Costa Rica as well as for a remote sensing based approach for the generation of standardized land cover maps at national level.

Level I (Vegetated areas)

Forest (Level II)

Description: **Tree cover (1):** are all wooded pixels with different tree canopy cover where the canopy is composed of single stemmed and woody plants. This class also includes riverine vegetation with sparse grass cover, mangrove trees, dense tropical evergreen lowland dominated by *raffia* (*Raphia taedigera*) or by *hog plum* (*Symphonia globulifera*) [Kappelle et al., 2002]. This study takes a special consideration of the ecosystem of *dense tropical broad-leave evergreen lowland swamp forest dominated by raffia* [Kappelle et al., 2002], because of the similarity with palm plantations. This ecosystem is located in lowlands and is frequently flooded.

Non Forest (Level II)

Description: **Shrubs (2):**. Sparse woodland or scattered trees giving a canopy cover but not in form of a tree cover. This class also includes a varying density of small shrubs and bushes. This cover is normally about 2m in height *in situ*. There are small areas with dense tropical shrubland dominated by *raffia* (*Raphia taedigera*), however, in the description of this ecosystem [Kappelle et al. \[2002\]](#) indicates that it is sometimes a woody herbland.

Description: **Herblands and natural non wooded lands (3):** in this class the herblands are grouped to include the shrubby tropical evergreen herbland dominated by graminoids and the woody tropical evergreen herbland dominated by graminoids [[Kappelle et al., 2002](#)]. In this class there are both natural herblands and pastures. Normally in gently undulating and undulating terrains it is common to find pastures in an association with a palm called royal palm (*Attalea butyracea*).

Description: **Cultivated managed areas (4):**. This class includes agricultural areas currently with crops. It also includes palm oil, rice, plantain and banana plantations, among others. Most of the plantations in the study area are in flat terrains.

There is a company called *Palmatica* which is the largest owner of palm oil plantations in the south of Costa Rica. This company buys the fruit of the palm from the small owners through contracts. Thus, there is an established market that makes this activity profitable.

In the case of *G.arborea*, the area planted has been decreasing, principally because of the increase in the international prices of palm oil. Currently most of the plantations are cut, or the owners have cut the largest trees and let the rest in a semi non-managed system. This is possible because *G.arborea* can re-sprout from the stumps.

Level I (No Vegetation)

Water bodies (Level II)

Description: **Water bodies (5)**, both natural and artificial are grouped in this category. In the study area the *Laguna de Sierpe* represents the largest body of freshwater in the study area, the remaining freshwater bodies are rivers. In Figure 4.21 (d) it is represented by a small black spot in the lower right part of the segment. The principal salt water body in the study area is the *Golfo Dulce* which flows into the Pacific Ocean. The rivers carry high sediment loads (see Figs.1.3 and 4.23).

Description: **Bare land (6)** is non-vegetated areas such as bare rocks or areas with very little vegetation cover where soil exposure is clearly apparent. This class also includes landslides. There are some natural small bare lands on the banks of the river *Grande de Térraba*, however, most of the bare lands are terrains which are being used for agricultural purposes.

Description: **Settlements (7)** are areas where there is permanent concentration of people with man-made structures. Except for the town of *Palmar*, there is not a big concentration of infrastructure in the study area. The remaining settled land scape consists of man-made constructions that are scattered or are at the sides of the main streets.

Level I (No Data)

No data (Level II)

Description: This class includes the **shadow areas (8)** and the **foreshortened areas (9)**. The latter is subdivided in foreslope and backslope. In the optical domain this class includes the clouds and the shadows of the clouds.

Chapter 2

Literature Review

2.1 Forest Monitoring

Remote sensing, together with other tools such as geographic information systems have been increasing the potential of ecological research and applications that are explicitly spatial and temporal [Cohen and Goward, 2004]. The models and algorithms developed by the scientific community have to be applicable at the global scale but also take into account the regional scale [Tansey et al., 2004]. Other authors such as Asner et al. [2000], believe that one of the principal objectives of remote sensing is carbon accounting in the world's forests via biomass monitoring.

It is generally accepted that forests provide a series of ecosystem services such as climate moderation, protection of streams and soils, oxygen generation, supply of wood and non-timber products, biodiversity beauty and recreation [Haines-Young and Potschin, 2007]. These ecosystems play a large role in the world's carbon budget and its dynamics [Clark et al., 2001].

The Food and Agriculture Organization of the United Nations (FAO) began reporting information about global forests in 1948 [FAO, 1948], and has subsequently published reports up to date. Recently, the organization published Forest

Resource Assessment (FRA) reports for the years 1980, 1990, 2000, and 2005¹. However, the inaccuracy of the *FAO* reports has been criticized [Waggoner, 2009]. Errors in estimating forest area are attributed to changes in the definitions of forests, revisions of estimates based on conflicting data, unreliable national inventory estimates, and data aggregation errors [Grainger, 2008; Houghton, 2005]. Historical *FAO* estimates of forest area are demonstrably assumption and do not exist for some countries [Grainger, 2008; Houghton, 2005]. As Waggoner [2009] points out, it is unclear from the *FAO* statistics whether the global forest area is declining or growing.

Any comparison of the current status and trends of forest resources among countries presumes that the nations apply definitions and concepts that produce comparable estimates. Presently however, cross-country estimates are not comparable [Vidal et al., 2008].

It is acknowledged that climate change will have a severe global impact, and this is one of the main contemporary discussion themes in the world. As one way of addressing this challenge, the Reduction of Emissions from Deforestation and Forest Degradation (*REDD+*) takes an important role on issues like conservation and the enhancement of carbon stocks in tropical forests. *REDD+* also includes sustainable forest management as a general goal. An important component of the *REDD+* mechanism is the Measurement, Reporting and Verification (MRV) standard which allows a transparent and comprehensive monitoring of the entire process. Monitoring encompasses collecting and assessing information over time which, for *REDD+*, refers to monitoring at a national level. Information (data) and services which can provide the data and infrastructure are also defined along the nationally available capacities. In addition, history-based reference emission levels must be determined in order to compare them with the emission reduction effects of the newly implemented measures.

¹available in <http://www.fao.org/forestry/fra/2620/en/>

2.1.1 Forest Measuring

The need for relevant information about forests and their function at both national and international levels is increasing [Vidal et al., 2008].

Forests play an ecological role in addition to the traditional industrial productivity role. The scope of National Forest Inventories (NFIs) has broadened and new variables for assessment have been introduced to address both national needs and the need for common reporting at the international level [Vidal et al., 2008].

However, published forest cover data often differ considerably [Kleinn et al., 2002]. Class confusions are the most frequent problem in remote sensing based landscape classifications because in tropical areas green vegetation has a very complex spectrum of different plant compositions and arrangements.

REDD+ is a financial initiative. Its main purpose is to reduce emissions of greenhouse gases from deforestation and forest degradation. In *REDD+*, one of the challenges is how to differentiate deforestation and degradation because this distinction is deeply intertwined with questions of how to measure either category. However, both concepts are connected in the political-economical and scientific dimension. In addition, in *REDD+* mechanisms, all actors need a base to negotiate or to implement one specific initiative. Normally, this base is a forest map.

Only remote sensing can provide the information needed to produce a forest map in regular time and large scale dimensions at a reasonable cost. Remote sensing plays an important role in cases in which it is not suitable to carry out terrestrial inventories such as for remote and hardly accessible locations or due to security reasons.

Remote platforms provide the only means of viewing large portions of the Earth surface at regular intervals and the selective absorption and reflectance of light by plants allows optical sensors to gather large amounts of information [Ollinger, 2011] as well as to provide essential input for climate and ecosystem models [Ustin et al., 2004].

For passive sensors, the amount of light reflected from a surface is determined by

[[Jackson and Huete, 1991](#)]: the amount and composition of solar irradiance that strikes the surface and the reflectance properties of the surface. Here we face two different issues: one is that the composition of solar irradiance varies with the atmospheric conditions the second is related to the intrinsic characteristics of the target itself.

From a global perspective, [Fagan and DeFries \[2009\]](#) pointed out, that it is difficult to distinguish between different types of forest (i.e primary, secondary forests or tree plantations) with actual remote sensing imagery. It is also challenging to detect forest degradation in which a forest is partially cleared by human activity [[Fagan and DeFries, 2009](#)].

Hyper-spectral imagery is a promising technology to solve some difficulties in discrimination of green vegetation. With this technology, the image is sub-divided in several narrow bands across the spectrum. Several recent studies have used hyper-spectral remote sensing [[Asner et al., 2008](#); [He et al., 2011](#); [Ustin et al., 2002](#); [Ustin and Santos, 2010](#)] to assess the spatial distribution of plant species.

Passive sensors such as the multi-spectral sensor have largely succeeded at classifying whole pixels, however, deep analysis related to its constituent substances is limited by a relatively low number of spectral measurements [[Keshava and Mustard, 2002](#)].

Hyper-spectral scanners have the ability, on the other hand, to collect data in over 200 very narrow continuous bands [[Jones and Vaughan, 2010](#)]. Therefore, this imaging spectrometers can discern the reflectance spectrum of each pixel. These spectra may then be compared with spectra measured from field samples or spectra from laboratory measurements collected in spectral libraries.

In the last few years in the optical domain, the typical patterns of the radiation in the electromagnetic spectrum are well described [[Jacquemoud et al., 2009](#); [Jones and Vaughan, 2010](#); [Ollinger, 2011](#)]. These patterns can be divided and summarized into three main spectral regions; the primary region emerges from the reflectance spectrum of a leaf. The first (400-700 ηm) pattern is predominantly due to light absorption by pigments (especially chlorophyll), the second (700-1200 ηm) pattern is influenced by light scattering at air-cell- all interfaces and

characterized by high reflectance and transmittance, and the third (1200-2500 nm) pattern is governed by water absorption.

Nevertheless, and in spite of the fact that the healthy leaves have a similar characteristic spectral signature [Daughtry and Walthall, 1998], the variability that exists between and within leaves, plants [Price, 1994] and ecosystems presents great challenges. The success of hyper-spectral approaches therefore depends on the ability to interpret reflectance data [Ollinger, 2011].

Although remote sensing has been proven to be a well-established method to map forest [Fagan and DeFries, 2009; Kuntz, 2010], there are several challenges that are still there. These challenges, can be summarized and categorized into technical, logistic and political issues.

In terms of logistic issues, the selection of satellite imagery is not always guided by scientific principles. Many times the selection of the imagery simply depends on availability and accessibility (i.e. the cloud coverage or the budget of a specific project).

From the technical point of view, there are problems associated with the spectral heterogeneity in transition zones where areas exhibit properties of two or more land cover types.

The political aspect is related to the definition of the classes, for example, the basic question of "what is a forest?" has been discussed by the international community. There is no harmonized definition of forest, not even one which could be implemented for the purpose of remote sensing. A similar challenge to the issue of defining a forest is posed by the task of differentiating forests and other land cover classes into various scales. For example, on a local level (a region within a country), a country level or even a global scale perspective.

2.1.2 Vegetation and Climate

For estimating national level forest cover only remote sensing can provide a view back in time (given that imagery is available). Even though accuracy and preci-

sion are always and issue, it also is the only data source that allows wall-to-wall mapping at an acceptable cost. Currently, most of the national forest maps have been produced with optical passive sensors. Though seemingly straightforward, remote sensing based forest mapping is a complex undertaking and involves analyses and interpretation. Vegetation Indices (*VI*s) as indicators of vegetation characteristics are widely used in this context [Glenn et al., 2008]. However, the *VI*s are calculated by different band transformations from optical sensors [Fuchs et al., 2009]. The calculation of *VI*s also depends on many factors that are not specific to the vegetation characteristics, e.g. solar zenith angle, view angle, soil background and climatic conditions such as atmosphere and moisture [Jackson and Huete, 1991].

*VI*s are widely used in the state of plant growth and plant phenology studies and are recognized, at a large regional scale, as an important indicator of ecological conditions [Weishou et al., 2011]. Other authors have established correlations between climate variables and *VI*s. For example, Shisanya et al. [2011] studied the rainfall variability and its impact on Normalized Difference Vegetation Index (NDVI) in Kenya, Ji and Peters [2003] evaluated the vegetation response (through NDVI) to drought in the Northern Great Plains and Zaitchik et al. [2007] studied the inter-annual climatic variability to characterize the climatic sensitivities of vegetation in the Middle East.

Given that climatic variables have a significant impact on the determination of *VI*s, it becomes clear that *VI*s for one and the same area vary over time. Analysis of time-series of *VI*s over different seasons will therefore be useful.

2.2 Remote Sensing and Earth Observation Monitoring Systems

Satellite remote sensing data provide an invaluable continuous, temporal and spatial information, which helps us understand the processes, dynamics and disturbances in the biosphere, and the impacts of environmental changes on terrestrial ecosystems [Zhao and Running, 2008].

Earth Observation (EO) is a technique that uses sensor data collection techniques, which implies that the information about the objects is collected from a certain distance. Remote sensing collects the data using sensors from different platforms, for example, space borne or from the air using an aircraft's airborne systems.

Up to present time, forest assessment has been implemented worldwide using remote sensing at different intensities and at different scales, ranging from the stand level to a national, multinational and global scale. These different views are changing some paradigms in terms of the way that we see forest systems [Lorentz et al., 2005].

Today there are 994 (at 12.31.2011) satellites operating, however, just 13% are for Earth science, Earth observation or meteorology [Union of Concerned Scientist (UCS), 2012]. Many of the systems have a military use.

With the launch of the first Landsat¹ satellite in 1972, the era of civilian satellite image on repetitive acquisition began.

Remote sensing information that is acquired in conjunction with ground measurements at the project level, can help in the development of regional and national estimates of the impacts of carbon sequestration initiatives [Sanchez-Azofeifa et al., 2009].

In remote sensing, it is possible to obtain relevant information about objects by the characterization of the radian flux, when it interacts with the surface. Carefully monitoring the exact nature of this incoming radiant flux in selective wavelengths is a critical step in remote sensing [Jensen, 2007].

Earth observation satellites provide data by using different portions of the electromagnetic spectrum at different spatial, temporal, radiometric and spectral resolutions [Pohl and van Genderen, 1998]:

The spectral resolution refers to the ability of the sensor to distinguish between two closely thematic objects in an image or in the electromagnetic spectrum [Jones and Vaughan, 2010]. In other words, it specifies the number of spectral bands in a determined wavelength range in which the sensor can collect reflected

¹<http://geo.arc.nasa.gov/sge/landsat/tofc.html> Access 01.02.2010

radiance. The narrowest spectral feature that can be resolved by the spectrometer is defined by the full width at half maximum (FWHM) of an instrument's response to a signal. Another important concept is the spectral sampling interval which is the interval between data points in wavelength units.

A temporal resolution is simply how often the sensor observes the same place [Lillesand et al., 2008]. For example, meteorological satellites have a very high temporal resolution in terms of minutes, such as the second generation of Meteosat which has temporal repeat cycles of 15 minutes and twelve spectral bands. On the other hand there are systems that have a very large repeat cycle, for example, the EO-1 Hyperion with 200 days [Beck, 2003].

The radiometric resolution refers to the number of possible brightness values in each data band, thus, after the data quantization process, the number of bits into which the recorded energy is defined [Jensen, 2007; Jones and Vaughan, 2010].

Spatial resolution is a measure of the smallest angular or linear separation between two objects that can be resolved by the sensor [Jensen, 2007; Jones and Vaughan, 2010]. In other words, it is the detail of discernible objects in an image which is represented by each pixel or grid cell in the image that is the ability to separate closely spaced objects on an image. Spatial resolution of passive sensors depends on their Instantaneous Field of View (IFOV). The *IFOV* defines the nominal spatial resolution, which is in turn defined as a dimension in meters where the diameter of the circle in the field (D) is a function of the *IFOV* (β) times the altitude (H) of the sensor above ground level [Jensen, 2007].

$$D = \beta \times H \tag{2.1}$$

The final system of nominal spatial resolution is composed of a matrix of picture elements, or pixels, of 30×30 meters in the case of Landsat TM+, for example.

2.3 Hyper-Spectral Data

Hyper-spectral scanners have the ability to collect data in over 200 very narrow continuous bands [Goetz et al., 1985; Jones and Vaughan, 2010], a process also called imaging spectroscopy [Smith, 2002]. For this reason, imaging spectrometers can discern the reflectance spectrum of each pixel. These spectra may then be compared with spectra measured from field samples or spectra from laboratory measurements collected in spectral libraries. For example, Roberts et al. [1999] utilize high-resolution *AVIRIS* to augment a regionally specific spectral library for the Santa Monica Mountains for use in posterior Multiple End-member Spectral Mixture Analysis [Roberts et al., 1999].

In typical multi-spectral sensor samples the given wavelength window is similar to the hyper-spectral one in terms of the range covered (400 to 2500 ηm), but it uses several broad band-passes, thus leaving large gaps between the bands. In contrast, hyper-spectral sensors sample the spectral window with very narrow band-passes (see Fig 2.1).

Table 3.4 shows the differences between hyper-spectral imagery and multi-spectral imagery, such as that of Landsat Thematic Mapper (TM) and three hyper-spectral sensors. For example, the high spectral resolution of individual channels in MODIS imagery is less than 10 ηm wide over a spectrum.

Spectral anomalies can be detected due to the narrow widths of the channels. If these were not given, as for example in multi-spectral imagery, the differences could be masked by broader bands.

The specific problems associated with hyper-spectral remote sensing spectral image analyses arise from any combination of the following [Mernyi, 1998] cited by Mernyi [1999]:

- The spectral patterns are high dimensional (more than 50 and less than a hundred bands);
- The number of data points (image pixels) can be as large as several million;
- The pixels are mixed: Several different materials contribute to the spectral

signature detected from each pixel;

- Given the richness of data, the goal is to separate many cover classes;
- Different surface materials may be distinguished by very subtle differences in their spectral patterns;
- Very little training data may be available for some classes; and classes may be represented very unevenly.

To compare different spectra it is necessary to convert the data to reflectance. It does not matter whether the spectra come from airborne, sensor or field data. The most critical step in most imaging spectrometer data analysis strategies is to convert the data to reflectance so that individual spectra can be compared directly with laboratory or field data for identification. This is one of the main results of the atmospheric correction procedure (see Section 2.3.1). Passive remote sensing is based on interaction between light and material. Each vegetation (leaves or group of leaves) has its own spectral response, this is the basis for the working principle of imaging spectrometers. Therefore, if remotely sensed data have high spectral resolution with a continuous wavelength range, the materials can be distinguished by using their spectral features. In addition, the narrower the band width, the higher the discrimination capacity of the sensor [Jensen, 2007; Sanchez-Azofeifa et al., 2009].

Remote sensing at the proximal visible and near-infra-red (VNIR, 400-2500 ηm) has been used to estimate the variability in leaf properties. For example, at the species level Castro-Esau et al. [2006] showed the sample variability of some meso-American trees. Studying these variabilities, the authors analysed implications for potential tree crown classifications. Dennison and Roberts [2003] performed a time series with hyper-spectral analysis to determine effects of vegetation penology in southern California Chaparral and the implications in posterior classifications. Clark et al. [2005] investigated the utility of high spectral and spatial resolution imagery for the automated species-level classification of individual tree crowns in Costa Rican tropical rain forest.

The combination of narrow band absorption features and reflectance provides

a unique avenue for understanding changes in the biophysical and biochemical characteristics of ecosystems [Asner, 1998]. For example, Schmidt and Skidmore [2001] analysed the spectral reflectance from eight species of range-land grasses in Kenya using a laboratory-based spectrometer. They found statistically significant differences in the spectral reflectance between species. A result which is encouraging for future work on identifying, classifying, mapping and monitoring rang-eland ecosystems from hyper-spectral imagery.

In spite of these efforts, the spectral properties of vegetation and soils must still be better understood, not just to identify plant species, but also to estimate plant productivity from remotely sensed data [Daughtry and Walthall, 1998]. However, other considerations such as scale and temporal dimensions are necessary. The ability to go from a leaf to a canopy-level increases the challenges [Ollinger, 2011].

From the field-work perspective, recent improvements in portable *VNIR* spectro radiometers have made this technology faster, lighter, and easier to use which is a requisite and a necessity for tree-level measurement [Fagan and DeFries, 2009]. In addition, the availability of space-borne hyper-spectral platforms makes the collection and analysis of data possible at low-cost, which may increase the number of examples for each species in spectral libraries.

Figure 2.1 demonstrates the potential and the differences of hyper-spectral data to distinguish between vegetation covers. The y axis is reflectance, which is defined as the ratio of reflected radiation to the incoming radiation. However, when the radiation reflected from a surface is measured, this measure is the spectral radiance, which is the radiant flux density emanating from the surface in $W m^2 sr^{-1} nm^{-1}$ [Jones and Vaughan, 2010].

The division of the optical domain from 400 to 2500 ηm are typically classified into three parts [Elach and van Zyl, 2006; Jacquemoud and Ustin, 2003]: "the visible region (400-700 ηm) characterized by a strong absorption of light by chlorophyll in a green leaf. In the near-infra-red region (700 to 1200 ηm), the absorption is limited to dry leaf matter where there is multiple scattering within the leaf and thus a strong reflectance due to the refractive index discontinuity between air and the leaf cell structure. Finally the middle-infra-red region (1200-2500

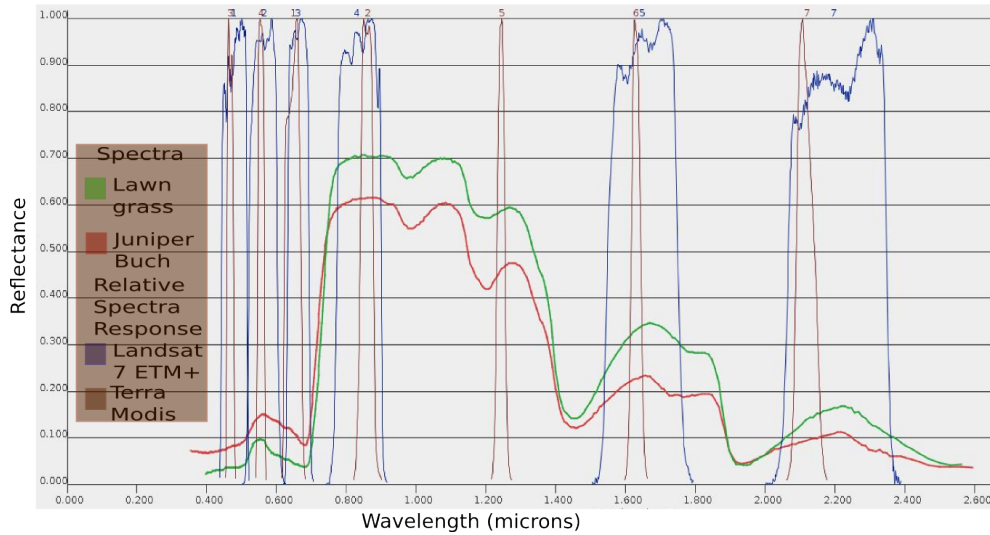


Figure 2.1: Spectral curves from Lawn Grass and Juniper Bush spectra and the relative spectral response for Landsat ETM+ and MODIS sensors. [U.S Geological Survey, 2010]

ηm) which is also a zone of strong absorption, primarily by water in fresh leaves and secondarily, by dry matter (dry carbon compounds like cellulose and lignin, nitrogen, sugars, and other plant compounds) when the leaf wilts and dries. All of these observations are a prerequisite to extracting biophysical information” [Elach and van Zyl, 2006; Jacquemoud and Ustin, 2003].

The interaction of the radiation with plant leaves and canopies and its respective magnitude of spectral reflectance (ρ_λ), spectral absorbance (α_λ), and spectral transmittance (τ_λ), depends not only on the wavelength but also on the range of the structural and chemical composition, the leaf age, the leaf thickness, the leaf structure, and the water content, among others [Jones and Vaughan, 2010].

2.3.1 Atmospheric Corrections

The Digital Number DN as a form to recorded the radiation emanating from a surface, can be converted to a spectral radiance at the sensor ($Wm^{-2}sr^{-1}$) as a previous step to obtain the reflectance at the surface. However, one needs to remove the effects of atmospheric absorptions and scattering of the electromag-

netic radiation travelling between the sun, the surface and the sensor [Jones and Vaughan, 2010].

When the electromagnetic radiation interacts with the Earth's atmosphere, the degree of interaction depends on the wavelength of the radiation and the local characteristics of the atmosphere (gases and aerosol composition) [Song et al., 2001; Tso and Mather, 2009]. For example, the Earth's atmosphere absorption is relatively small at frequencies lower than 10 GHz (3 cm wavelength), frequencies where clouds are also transparent. However, when the frequencies rise, the absorption increases because of the presence of water vapour and oxygen [Elach and van Zyl, 2006].

The source of incident electromagnetic radiation (IEMR) for passively acquiring remote sensors is the sun. The visible (VIS) (400-700 ηm) and near infra-red (NIR) (700-1300 ηm) regions contain the maximum amount of solar energy, which falls off in energy into the *SWIR* short-wave infra-red (SWIR) (1300-2500 ηm). Electromagnetic radiation must travel through the atmosphere of Earth before interacting with surface materials [Jensen, 2007].

The earth's atmosphere contains CO_2 , O_2 , O_3 , H_2O , CH_4 , CO , NH_4 , N_2O , and other nitrogen gasses, which interact with approximately 50% of the *IEMR* spectrum over the region of 300-2800 ηm [Gao et al., 1993]. Gases and particles in the atmosphere absorb and scatter significant amounts of the *IEMR* over this region. The absorption features of these gases are typically narrow. Particular regions of the electromagnetic spectrum, such as 900 ηm , 1100 ηm , 1400 ηm and 1900 ηm wavelengths, are almost opaque to sensors. Regions that do not contain significant absorption features due to atmospheric gasses are called transmission windows. It is these transmission widows that are exploited by remote sensing sensors [Jensen, 2007].

The carbon dioxide (CO_2), oxygen (O_2), methane (CH_4), carbon oxide (CO) and ammonium ion (NH_4^+) gases remain in relatively constant concentrations through the atmosphere, whereas ozone (O_3) predominantly occurs in the stratosphere. The concentration of water (H_2O) can vary with altitude whereas the other gases can differ with the environment's setting [VerMOTE et al., 2006].

Remotely sensed spectral imagery of the Earth's surface can be used to the fullest advantage only when the influence of the atmosphere has been removed and the data are reduced to units of reflectance [Adler-Golden et al., 1999].

Atmospheric correction (or compensation) of spectral imagery refers to the retrieval of surface reflectance spectra from measured radiances [Matthew et al., 2000] and it is a prerequisite to enhance and to exploit hyper-spectral imagery data analysis approaches [Borengasser et al., 2008]. The need for atmospheric correction depends on the objectives of the analysis. For instance, when multi-sensor or multi-date images are being classified, or where the aim of pattern recognition is to identify land cover change over time [Tso and Mather, 2009].

A number of studies have been carried out on the inter-comparison of the techniques and performance of some of the various atmospheric correction schemes [Goetz et al., 2003; Kruse, 2004].

Some cases in which the advantages of radiometric and atmospheric correction in remote sensing become apparent are: (1) when multi-temporal data applications such as in land use/cover change detection are desirable; (2) across scene (across path) comparison of spectral information of land cover types, for example, a scene of HyMap can be several km long; (3) multi-sensor data applications such as multiple image mosaic to spatially produce a large image, multi-sensor data fusion; (4) quantitative analysis by combining field survey data with spectral data for applications such as biomass estimation; and (5) band ratio operations such as vegetation indexes (after Lu et al. [2002]).

The magnitude of the electromagnetic energy in the visible and near-infra-red region of the spectrum that is detected by a sensor above the atmosphere is dependent on the magnitude of incoming solar energy (irradiance), which is attenuated by the process of atmospheric absorption and by the reflectance characteristics of the ground surface. For this reason, energy received by the sensor is a function of incident energy (path radiance), and atmospheric absorption [Tso and Mather, 2009].

There are basically three types of atmospheric correction approaches. The first is a scene-based empirical approach, the second is the radiative transfer modelling

approach; and the third are the hybrid approaches [Gao et al., 2006]. One example of an radiative transfer modelling approach is the Fast Line-of-sight Atmospheric Analysis of Spectral Hyper-cubes (FLAASH) which is an atmospheric correction of hyper-spectral and multi-spectral imagery. *FLAASH* handles data from a variety of hyper-spectral and multi-spectral sensors, supports off-nadir as well as nadir viewing, and incorporates algorithms for water vapour and aerosol retrieval and adjacency effect correction. [Adler-Golden et al., 1999; Matthew et al., 2000].

2.4 RADAR Data

The principle of **RA**dio **D**etecting **A**nd **R**anging (RADAR) is that it does not need electromagnetic radiation from the sun. Therefore it is called active remote sensing, which is based on the transmission of long wavelength microwaves through the atmosphere followed by recording the amount of energy backscatter from the surface [Jensen, 2007]. In Figure 2.2 the antenna sends a beam through the transmitter, this energy interacts with the objects and is then recorded in a digital recorder.

The *RADAR* systems are capable of crossing most atmospheric gases and also the clouds, but this ability depend on the wavelength (λ) of the system. The bands are associated with some letters (see Tab. 2.1). The general idea is that the larger the wavelength, the higher the penetrability of the beam.

However, there are other important characteristics of *RADAR* which complement the understanding of *RADAR* functionality, these parameters are [Jensen, 2007; Richards, 2009]:

- Azimuth direction, refers to the in-flight direction along the orbit or flight path. Since *RADAR* is a side looking system, the objects near the line of the trajectory of the antenna are close to the antenna, that is why it is called near-range. On the other hand, the objects situated far from the antenna are called the far-range. The azimuth direction is directly dependent on the size of the antenna. A significance improvement of *RADAR* systems is archived by synthesising a very long antenna electronically by combining

many pulses along-track.

- Range direction, is the looking direction of the *RADAR* perpendicular to the azimuth, thus the direction of the beam that is at a right angle to the direction of the sensor's trajectory. The range direction is dependent on the incidence angle and the band width.
- Incidence angle, refers to the angle between the incident radar beam and the direction perpendicular to the ground surface. The interaction between microwaves and the surface depends on the incident angle of the radar pulse on the surface.
- Polarization, when the electric pulse oscillates along the horizontal direction, the wave is H polarized. If the electric field vector oscillates perpendicular to the horizontal direction, the wave is V polarized.

As the *RADAR* systems send their own pulses of energy (active sensors), they are capable of night-time operation. The current *RADAR* systems are capable of synthesizing a much larger antenna, using the forward linear motion of the space platform, and are for that reason called Synthetic Aperture Radars (SAR) [Richards, 2009].

Some of the existing *SAR* sensors are COSMO-SkyMed and TerraSAR-X in X-band and RADARSAT-1/2, ERS-2 in the C-band [ITT Visual Information Solutions, Nd; Union of Concerned Scientist (UCS), 2012]. ERS-2 and RADARSAT-1 provide data based on a single polarization (VV for ERS-2, and HH for RADARSAT-1). With the sensor ASAR it is possible to acquire images in dual-polarization mode (HH/HV, HH/VV, VV/VH). The radar incidence angle is 23° for ERS-2, between 20° and 50° for RADARSAT-1, and ranges from 14° to 45° for ASAR. The nominal swath width for ASAR and ERS-2 is greater or equal to 100 km, with a spatial resolution of, at best, 25m (12.5m pixel size). For RADARSAT-1, the nominal swath width is greater than or equal to 50 km, with a spatial resolution of, at best, 10 m (6.25 m pixel size).

Today it is possible to order *SAR* data with multiple polarization configurations, incidence angle, radar wavelength (X-, C- and L-bands), and at high spatial reso-

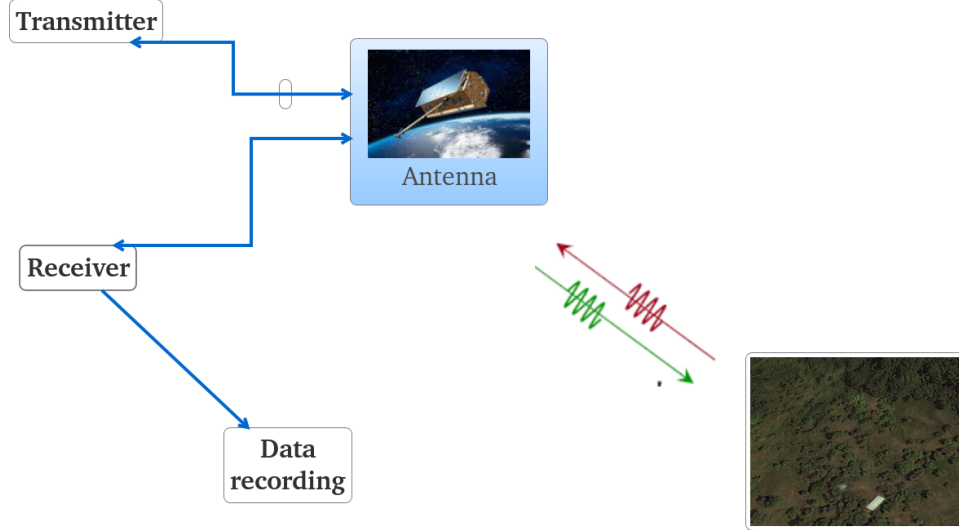


Figure 2.2: Principles of typical RADAR system components (adapted after Jensen [2007]).

lution (metric for TerraSAR-X and COSMO-SkyMed, ≈ 3 m for RADARSAT-2, and 6m for PALSAR) [Liao et al., 2009]. The increasing number of *SAR* satellites and the high revisit interval of new sensors (TerraSAR-X 11 days and COSMO-SkyMed every 15 days) makes it possible to follow the vegetation changes with high temporal frequency [Union of Concerned Scientist (UCS), 2012].

Table 2.1: Characteristics of some selected *SAR* sensors

Band	Wavelength(cm)	Platforms
P	30-100	AIRSAR
L	15-30	ALOS PALSAR, JERS-1 SAR
S	7.5-15	ALMAZ-1
C	3.8-7.5	RADARSAT-1-2, ENVISAT, ERS-1-2
X	2.5-3.8	TerraSAR-X, COSMO-SkyMed
Ku	1.7-2.5	Military domain
K	1.1-1.7	Military domain
Ka	0.8-1.1	Military domain

2.4.1 Principles of RADAR Data Processing

2.4.1.1 RADAR Relief Displacement, Shadows and Speckle

For a focus on forests in slopes (undulating and rolling terrains), there remain several difficulties despite the cloud penetrability of the radar images which are summarized here, based on [Jensen \[2007\]](#); [Kuntz \[2010\]](#); [Richards \[2009\]](#):

- Contrary to the passive sensors, where the light may be scattered into the shadow, there is no information within the shadow areas of the radar images that contain shadows, due to the side-looking effect of the *SAR* antenna. A possible solution for this can be to use ascending and descending orbits.
- The shape of objects depends on *SAR* imaging geometry and also on the imaging mode. That is the reason the same object may appear differently (e.g. with different shadows) depending on the acquisition parameters.
- The grainy salt-and-pepper *speckle* phenomenon causes random constructive and destructive interference. Due to this, some features appear as bright or as dark areas. The speckle can be reduced by a process called *multiple looks*.
- All the mountains that have a slope towards the looking direction will appear compressed or *foreshortened* relative to the opposite side of the same mountain (*back-slope*). In general, the foreshortening depends on the object height and the incidence angle. This effect will cause that higher objects to be more foreshortened and the intensities recorded by the sensor to be brighter at the top of the mountain and darker in the back-slope. Also, the smaller the incidence angle is, the greater the foreshortening.

2.4.1.2 Surface Roughness and Electrical Characteristics

Vegetation canopy forms a complex, heterogeneous volume which consists of different sizes, shapes and orientation components [[Jones and Vaughan, 2010](#)]. The size of vegetation or of a group of objects is related to the roughness characteris-

tics and to the surface texture characteristics. [Jensen \[2007\]](#), makes a comparison of the terminology used in the interpretation of aerial photography. For example, the division into rough, intermediate and smooth (fine) surface texture. The author then extended the analogy to include the concepts of micro-scale, meso-scale and macro-scale. This categorization takes the height of the objects into account and is dependent on the wavelength of the system. For example, at a wavelength of 3 cm the micro-scale level is defined by a surface with a height less than 0.17 cm (smooth) for the intermediate level the range is between 0.17 cm and 0.96 cm. Finally, the rough level is attributed to all objects higher than 0.96 cm. The meso-scale and macro-scale function for large areas, for example a patch of forest (meso-scale) or entire landscape that takes into account the topography (macro-scale).

A smooth surface acts like a specular reflection, where the incident radar pulse is reflected from the antenna in the same angle as the incidence. Thus, very little energy is scattered back to the *RADAR* sensor. The opposite case is a rough surface which reflects the incident radar pulse in all directions. Here part of the energy is recorded by the sensor.

As different objects at different scales have different roughness characteristics, the properties have been studied in a diverse manner. For example, [Gade and Stelzer \[2000\]](#) studied the radar backscattering from wet sediments, and found that the backscattering depends on the wet sediments' surface roughness properties. However, surface roughness is another major limiting factor in soil moisture retrieval for which simple correction procedures are extremely difficult to develop [[Barrett et al., 2009](#)].

[Park et al. \[2010\]](#), reviewed and developed effective methods of extracting geophysical information of tidal wetlands in Suncheon Bay, South Korea. However, they pointed out that relating radar derived roughness parameters to fundamental environmental processes in tidal flats has to be further investigated through time series of polarimetric *SAR* data sets and detailed in-site measurements.

Different types of terrain conduct the electricity better than others. This characteristic is the complex dielectric constant [[Jensen, 2007](#)]. The dielectric constant

of the surface materials depends strongly on the moisture content of the landscape components [McDonald et al., 2002]. It also depends on the size, shape, distribution of the individual object's constituents (i.e. the leaves, reproductive organs, branches, stems and boles) which are primary determinants of the radar signature [Kasischke and Christensen-Jr, 1990].

To link variation in canopy conductance to soil moisture in the rooting zone of a forest dominated by *Pinus taeda*, Oren et al. [1998] measured several variables such as through-fall, volumetric soil moisture, transpiration from trees, and vapor pressure deficit. They found that canopy leaf conductance, calculated from these variables, was strongly related to volumetric soil moisture in the upper soil layer.

2.4.2 Texture Quantization and Image Segmentation

In many machine visions and image processing algorithms, simplifying assumptions are made about the uniformity of intensities in local image regions. However, images of real objects often do not exhibit regions of uniform intensities. For example, the image of a wooden surface is not uniform but contains variations of intensities which form certain repeated patterns called visual texture. The patterns can be the result of physical surface properties such as roughness or oriented strands which often have a tactile quality, or they could be the result of reflectance differences such as the colour of a surface [Tuceryan and Jain, 1988]. Texture could be defined as a structure composed of a large number of more or less ordered similar elements or patterns without one of these drawing special attention [van Gool et al., 1985].

A texture measure attempts to characterize the texture within an image [Wiebe, 1998] and describes an innate property of objects which contains important information about the structural arrangement of surfaces [Castro-Esau et al., 2006; Tso and Mather, 2009].

In remote sensing the images are not composed of regions of uniform intensities which exhibit regular patterns called visual texture [Tuceryan and Jain, 1988]. In an image classification approach different textured regions are labelled with their

proper category label. A further goal of texture segmentation is to separate the regions in the image which have different textures and identify the boundaries between them [Tuceryan and Jain, 1988].

2.5 Data Reduction and Fusion

Many authors believe that hyper-spectral images are highly redundant because of band correlations [Mernyi, 1999]. Therefore data reduction techniques are necessary. For example, multispectral remote sensing data are usually highly correlated [Song et al., 2001], thus, Principal Component Analysis (PCA) is widely used. For example, [Theseira et al., 2003] found that 96.9 % of the variance of the imagery information, using the Along Track Scanning Radiometer 2 (ATSR-2) imagery in central Finland, is explained by the first principal component. But, in hyper-spectral remote sensing, *PCA* is used not only because of the redundancy of some bands, but also because of the dimensionality of the data which makes the process a challenge [Robila, 2005].

One important aspect of spectral imaging is that it generates a large amount of data which may be difficult to handle, be viewed and interpreted [Scholten et al., 2005]. The size of a spectral image cube can easily reach several tens of megabytes and post-processing methods are needed in order to fully exploit the information contained in the image cube. The main goal of hyper-spectral data processing is to distinguish and identify materials or land covers, to detect signs of degradation, to enhance the visibility of faint or obscured features from their spectral signatures and to spatially group pixels with similar characteristics [Klein et al., 2008].

As the hyper-spectral data are obtained in hundreds of bands for many applications, it may be inefficient and undesirable to utilize the data from all the bands (i.e., features) due to the increased computational time in data processing [Cheng et al., 2006].

To exploit all the potential of the different multi-sensor, multi-temporal, multi-resolution and multi-frequency image data, the fusion of digital image data has

become a valuable tool in remote sensing image evaluation [Pohl and van Genderen, 1998]. If a monitoring system is established, the fusion techniques are also useful in the case of a sensor's breakdowns. For instance, the Landsat program has been used widely since 1972 [Union of Concerned Scientist (UCS), 2012] with a specific and expected lifetime which was longer than originally planned. However, the Landsat 7 ETM+, the Scan Line Corrector (SLC) in the instrument, failed before its expected life time. To fix this error in the meantime, it is possible to use of multi-temporal image combinations where the gaps of one images are automatically filled by the next image.

Principal Components Analysis (PCA) is a statistical technique that has applications in fields such as face recognition and data reduction as well as image compression, pattern recognition and image analysis. It is a common technique for finding patterns in data of high dimension and to reduce the data dimensionality [Izenman, 2008; Smith, 2002]. For example, Pu and Gong [2004] used *PCA* as a feature extraction methods to map forest crown closure and leaf area index (*LAI*) with EO-1 Hyperion data. The use of *PCA* as data reduction is more widely used with different sensor types. For instance, Khan-Ullah and Jian [2007] used *PCA* to shrink the dimension of the feature vector in a classification performance.

PCA explores the variance-covariance or correlation structure of a sample set in vector form and it serves data (or dimensionality) reduction and interpretation [Johnson and Wichern, 2007].

The orthogonal axes in *PCA* analysis, are determined by performing an eigen-decomposition of the sample covariance matrix of the data [Liang, 2004]:

$$\Gamma = \frac{1}{N} \sum_{i=1}^N (R_i - \mu_R) (R_i - \mu_R)^T \quad (2.2)$$

where μ_R is the mean vector of the pixel reflectance R_i . The covariance matrix can be decomposed into eigenvector matrix U and eigenvalues $\Gamma = U \Sigma U^T$, where Σ is the diagonal matrix of eigenvalues.

The reduction is done when the first few largest eigenvalues and the corresponding eigenvectors are selected and retained for a successful unmixing in the lower dimension [Liang, 2004].

Spectral component substitution techniques have been developed principally to fuse multi-spectral and panchromatic images where the panchromatic band usually has a much higher spatial resolution, as well as for *RADAR* and optical imagery [Musa and Hussin, 2000; Sadidya et al., 2004]. They are based on replacing a spectral component of the low-resolution multi-spectral image by the radio-metrically adjusted panchromatic image. The most frequently used spectral component substitution techniques include hue-intensity-saturation (HIS) methods, *PCA* methods, and regression methods [Liang, 2004].

2.6 Cluster Analysis and Linear Discrimination Analysis

Cluster analysis is the process of classifying objects or homogeneous natural groups [Hoffbeck and Landgrebe, 1995] into subsets that have meaning in the context of a particular problem [Jain and Dubes, 1988]. It is an unsupervised learning method that categorizes objects into groups, where no assumptions are made concerning the number of groups or the group structure, distribution of patterns and interesting correlations in large data sets [Becker and Plumley, 1996; Halkidi et al., 2001; Kaufman and Rousseeuw, 2005]. For the objects, there are no specifications by which group they will be clustered. In other words the clusters are not known in advance [Maindonald, 2004].

This is in fact a contrast to the supervised classification (such as Linear Discriminant Analysis, Decision Tree, Spectral Information Divergence, Naive Bayes, Maximum Likelihood, Parallelepiped and some others) where the user defines the training classes prior to clustering objects and the computer identifies the best classification rules.

There are basically two categories of clustering, hierarchical clustering and par-

tition clustering [Commaniciu and Meer, 2002; Johnson and Wichern, 2007]. In the first approach, the user categorizes the objects into a hierarchy similar to a tree-like diagram called dendrogram [Halkidi et al., 2001]. Partition clustering in contrast (such as k means clustering and expectation-maximization algorithm), will construct non-overlapping groups [Johnson and Wichern, 2007].

In the hierarchical clustering the principal steps are [Chen et al., 2002; Halkidi et al., 2001]:

- Preprocessing of data may be necessary prior to their utilization in clustering.
- To create a hierarchical structure that reflects the order in which groups are merged or divided, a clustering algorithm is used which consist of two steps. The first is the proximity measure (distance) that quantifies how similar two data points are, for example. The second is the clustering criterion that is a cost function, and is used to find the pair of sets $\{S_i, S_j\}$ from the list that is the *cheapest* to merge. Once merged, S_i and S_j are removed from the list of sets and replaced with $S_i \dots S_j$ in an agglomerative method, which builds the hierarchy by merging the objects which initially belonged to a list of singleton sets S_1, \dots, S_n .
- This process iterates until all objects are in a single group. Complete linkage, average linkage, and single linkage methods use maximum, average, and minimum distances between the members of two clusters, respectively.
- Validation of the results has to be verified using appropriate criteria and techniques.
- Finally, the interpretation of the results means that the results of the clustering have to be analysed in order to draw the right conclusions.

Different variants of agglomerative hierarchical clustering algorithms may use different cost functions [Chen et al., 2002]. In summary the user specifies the rule of clustering in terms of how the distances between clusters are computed.

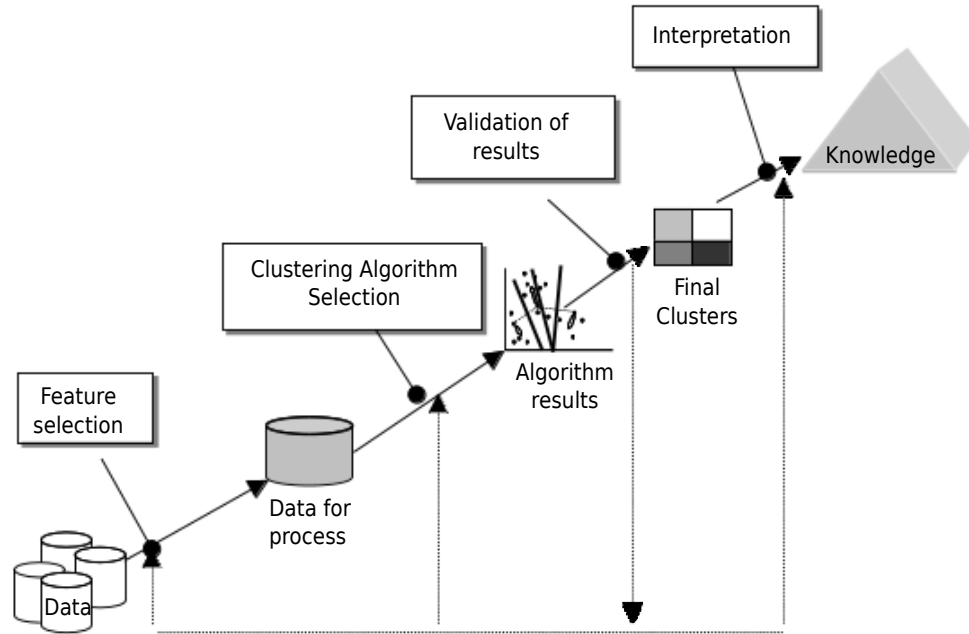


Figure 2.3: Steps of clustering process after [Halkidi et al. \[2001\]](#).

Referring to the validation phase, however, there are some natural questions that are difficult to answer [[Rezaee et al., 1998](#); [Rousseeuw, 1987](#)]:

- How good are the partitions?
- Are the dissimilarities within clusters small when compared to the dissimilarities between clusters?
- Which objects appear to be well-classified, which ones are misclassified?
- What is the overall structure of the data like?
- Can we estimate the number of clusters that are really present?

The validation implies the application of statistical methods and testing hypotheses [[Abonyi and Feil, 2007](#)]. For example, [Rousseeuw \[1987\]](#) proposed the silhouette approach, which tries to answer these questions. [Shimodaira \[2002\]](#) developed a method for assessing the uncertainty in hierarchical cluster analysis which is called an approximately unbiased (AU) test that uses a newly devised multiscale bootstrap technique.

In addition, if the number of classes within the data is not known a priori, a validation index may help to find out the optimal number of classes [Rezaee et al., 1998].

The Linear Discrimination Analysis (LDA) [Fisher, 1936] uses a priori knowledge about the agreement of different groups in the data, separating the groups using linear functions which are built by estimating the central location and the multivariate spread covariance for each group. This method can be used as an estimator. Once the functions are estimated, new observations can be assigned to a group [Reimann et al., 2008].

The group covariances are estimated using the covariance matrix of the data for each group. For this reason data outliers plays an important role in the estimations of the group centres' covariances [Reimann et al., 2008].

Chapter 3

Methodology

Chapter 3 involves the processing of the hyper-spectral data as well as the TerraSAR-X imagery. Two work-flow diagrams are presented in Figures 3.6 and 3.7, for the hyper-spectral and *SAR* process respectively. In addition, in Appendix A the long term time series from the Costa Rican dataset and its relation to the climatic data information in Costa Rica is presented. Specifically, in this appendix, the methodological steps to extract the 10-year long time series in Normalized Difference Vegetation Index (NDVI) and more than 30 years of climatic data is presented. The main idea is to identify trends and changes in vegetation over a decade to understand the dynamics of the local and Costa Rican terrestrial ecosystems and its links with the imagery used in this study.

3.1 Hyper-Spectral

The relationship between land use and land cover (LULC) is often used as a fixed notion and since remote sensing data usually allows only the assessment of the Earth's cover [Breidenbach et al., 2009], in this study the term land cover is henceforth used to avoid ambiguity.

In this chapter, three hyper-spectrometers with different resolution are tested in order to explore the spatial-scale dependency of spectral reflectance in the remote

identification of forest classification.

A codification for hyper-spectral scene was used (see Tab. 3.2). These are the prefixes for HyMap (*hy*) and for Earth Observing-1 (*eo1*).

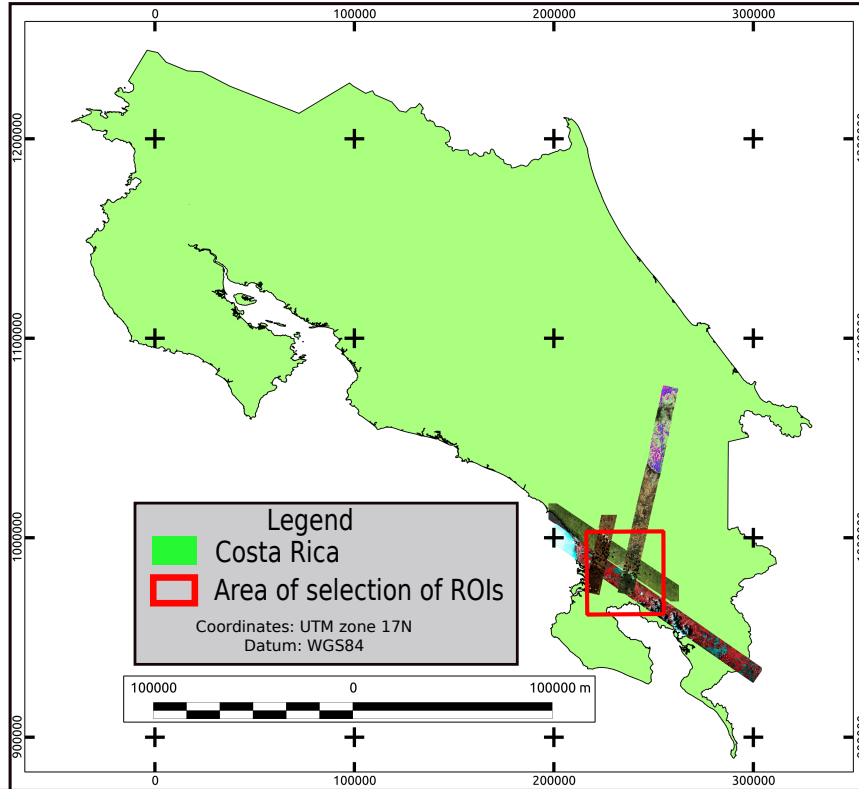


Figure 3.1: Overview of the hyper-spectral scenes in the south of Costa Rica

In this study three types of hyper-spectral data were used. First, two scenes of the EO-1 Hyperion sensor, second, the HyMap (also two scenes). Third, the *USB650* field spectrometer (Ocean Optics) which was used to measure several sample of leaves of *G.arborea* trees and also another samples of palm oil, teak, rise and plantain.

Table 3.1 list the codes used for the selections of the *ROIs*. Not all the scenes have the same quantity of this *ROIs*. The *hy-15* has three covers that are not included. These covers are: clouds, shadows and mangrove. In the *hy-14* scene there is not grass. For the *eo1* scenes it was possible to archive all the covers

presented in the Table 3.1.

Table 3.1: Codification and number of selected regions of interest per hyper-spectral scenes.

Region	Codes	ROIs per scene			
		hy_15	hy_14	eo1_pz	eo1_kw
G.arborea	X1000	10	4	8	3
palm	X22	6	9	10	4
forest	X33	14	6	7	8
soil	X44	7	9	8	5
grass	X55	12	NA	7	5
mangrove	X66	4	5	5	1
clouds	X77	NA	5	8	2
shadows	X88	NA	5	6	2
water	X99	3	5	6	4
Total		56	48	65	34

3.1.1 EO-1 Hyperion

The swath width of the Hyperion is 7.5 km where the characteristic scene size is 7.5 x 100 km, a narrow strip [Beck, 2003]. The Level 1 radiometric product has a total of 242 bands but only 198 bands are calibrated, there are only 196 unique channels. Calibrated channels are 8-57 for the *VNIR*, and 77-224 for the *SWIR* and the reason for not calibrating all 242 channels is mainly due to the detectors' low responsiveness. The bands that are not calibrated are set to zero in those channels [Beck, 2003]. Level 1Gst (L1Gst) is radio-metrically corrected and re-sampled for geometric correction and registration to a geographic map projection with a swath width of 7.7 km (see Tab. 3.4).

In Table 3.4, other useful characteristics of the hyper-spectral data are presented.

Table 3.2: Codification used and acquisition dates for to the four hyper-spectral scenes

Scene name	Code used	Acquisition date
hy2_20050315_s01	hy_15	2005/03/15
hy2_20050314_s01	hy_14	2005/03/14
EO1H0140532005074110PZ	eo1_pz	2005/03/15
EO1H0140532006054110KW	eo1_kw	2006/02/23

The final *L1R* data provides a total of 198 bands representing 427-2395 ηm of continuous spectra with 10 ηm spectral resolution. The data files are organized in Band-Interleaved-by-Line (BIL) format.

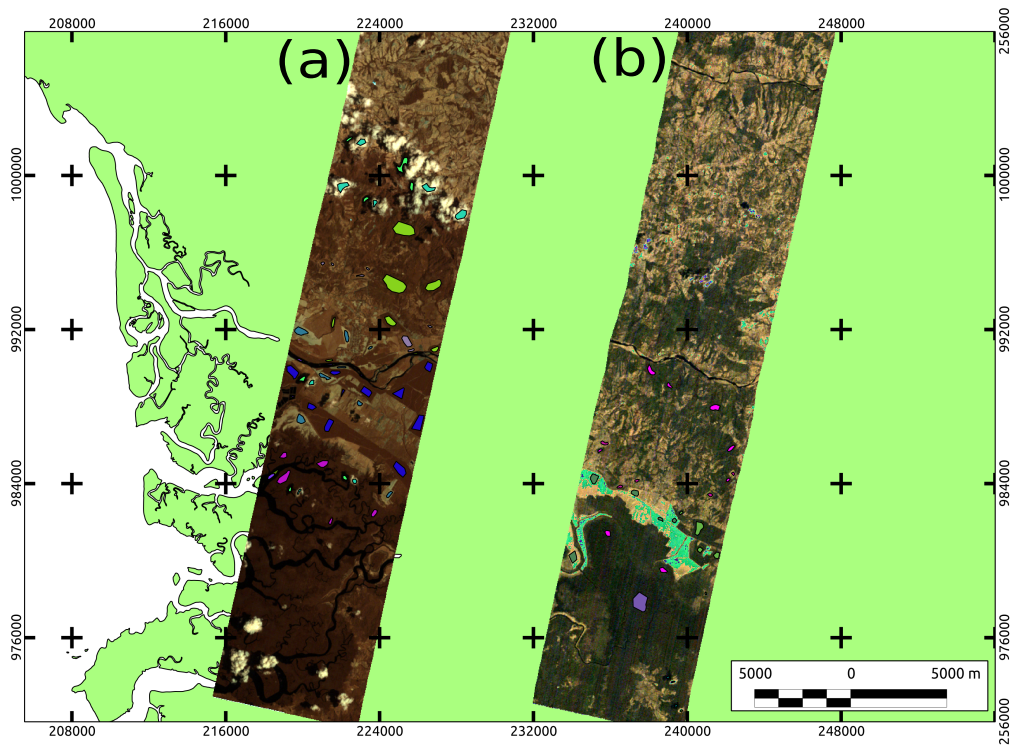


Figure 3.2: Segment of colour composite of the EO-1 Hyperion images and its *ROIs*(a) *eo1_pz*, (b) *eo1_kw*

Two scenes from EO-1 Hyperion images were chosen. One (*eo1_pz*) was acquired on March 15, 2005. The second one (*eo1_kw*) was acquired on February 23, 2006.

Data from the Earth Science Data Interface (ESDI) at the Global Land Cover Facility was downloaded. The projection of the EO-1 Hyperion images was *UTM*, reference ellipsoid was *WGS84*, and zone number was 17N. The data format was in Geo Tagged-Image File Format (GeoTIFF) radio-metrically corrected (Level 1Gst) data, written as *BIL* files stored in 16-bit signed integer radiance values [U.S. Geological Survey, 2006].

The un-calibrated image bands from the original 242 channels were removed. For example, the bands 1-7 (356-417 ηm), bands 58-70 (936-1058 ηm), bands 71-76 (852-902 ηm) and bands 225-242 (2406-2578 ηm) are not calibrated and set to zero in the *L1R* product. Thus, the final *L1R* data provides a total of 198 bands representing 427-2395 ηm continuous spectra with 10 ηm spectral resolution.

The original EO-1 Hyperion scenes to level 1 radiance were processed using *ENVI* [ITT Visual Information Solutions, 2012] and Equations 3.1 and 3.2. The gain, offset, and mean band centre and mean wavelength information were provided by the U.S. Geological Survey [2006] in order to edit the band headers. The units of the radiance scenes are in $Wm^{-2}sr^{-1}\eta m^{-1}$ (Watts per stereo-radians per square meter)

$$VNIR_{(L)} = DN/40 \quad (3.1)$$

$$SWIR_{(L)} = DN/80 \quad (3.2)$$

Where L are the radiance values, and DN are the digital numbers.

An atmospheric correction of the EO-1 Hyperion radiance images was performed using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (*FLAASH*) module in *ENVI*. The headers of the scenes were edited again to obtain the correct wavelength, *FWHM*, and scale factors in accordance with the parameters of the U.S. Geological Survey [2006]. *FLAASH* uses the MODerate resolution atmospheric TRANsmission (*MODTRAN*) radiative transfer model to archive the at-surface reflectance.

After atmospheric correction with *FLAASH*, the final products have units of at-

surface reflectance, which is a ratio between the reflected spectral radiant flux density at the surface and the direct beam spectral solar irradiance at the surface [Jones and Vaughan, 2010].

3.1.2 HyMap

HyMap (Hyper-spectral Mapper), is series of airborne hyper-spectral sensors that deliver up to 128 bands covering the 440 to 2500 η m spectral region and two thermal bands. The design is fundamentally modular which allows the configuration of the spectral and spatial characteristics to suit the customer's specific requirements [Cocks et al., 1998].

The HyMap imagery was acquired between March and April of 2005 in a mission called "Costa Rica Airborne Research Technology and Application" (CARTA-2005). The specific acquisition dates are in Table 3.2. The images at surface reflectance were provided by the National Aeronautics and Space Administration (NASA). The HyMap images were converted to apparent surface reflectance using the *HyCorr* atmospheric correction software (Hyvista Corp., Sydney, Australia). This is a modified version of the Atmospheric Removal (ATREM) algorithm [Gao et al., 1993]. The HyMap data was provided in 126 bands because two bands (band 1 and 32 of the VIS module) were deleted during the pre-processing steps.

Both images were geo-corrected using the geo-correction data files provided by the *CARTA* Mission to the *UTM* zone 17 north and ellipsoid WGS84 as the final projection.

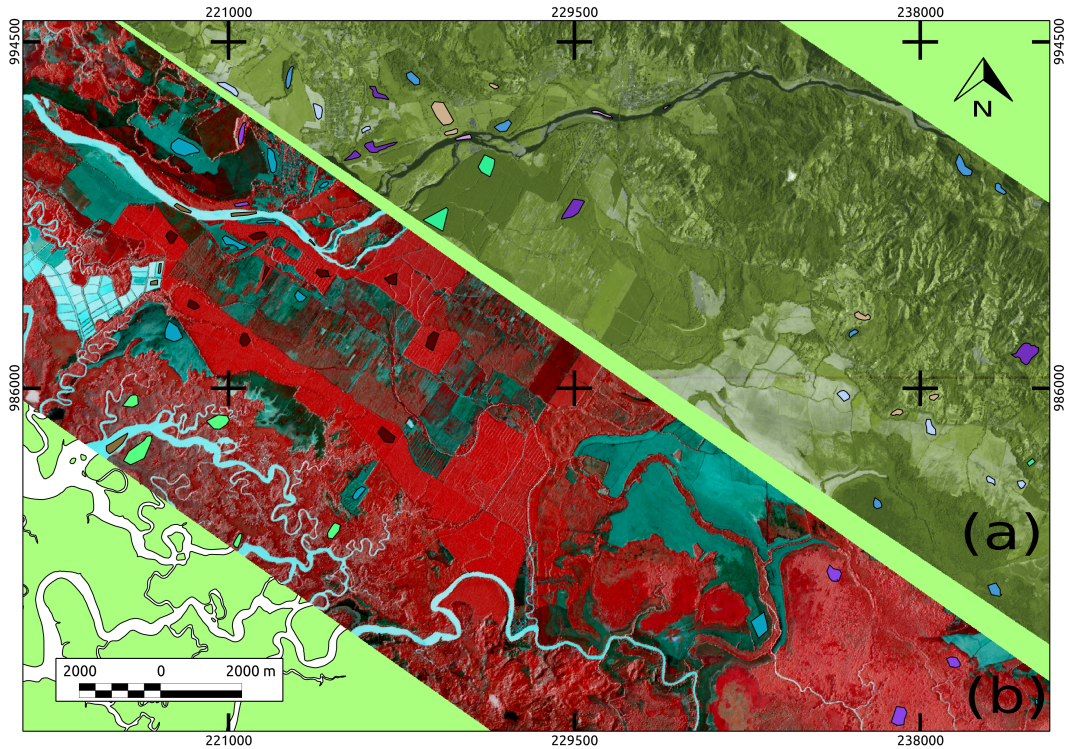


Figure 3.3: Segment of colour composite of the HyMap images (a) hy_15, (b) hy_14

3.1.3 Selection of the Regions of Interest

In the Quantum GIS geographical information system [Quantum GIS Development Team, 2012] the *ROIs* were selected based on the very high spatial (50 cm proximately) aerial ortho-rectified-photographies, provided by the Cadastre and Register Regularization Program (PRCR) in Costa Rica. A revised field-based and old *G.arborea* data base register of plantations was used to identify pure *G.arborea* stands over the study area.

The *ROIs* selection is a combination of pure spectra, similar to the concept of *Endmembers*, which can be measured in the laboratory as well as be extracted from imagery [Dennison and Roberts, 2003]. For all the *ROIs*, a buffer of 60 m was carefully drawn from the edge of each selected cover to the polygon centre

of mass in order to avoid the edge mixture pixels. A special effort was made to accurately locate these *ROIs* in order to be sure of the representativeness of each pixel and avoid lateral scattering.

The VIPER Tools to extract the spectral signatures via *ROIs* were used. This application is made available by the Department of Geography at University of California Santa Barbara¹. For all the spectra sets a Frost filter was applied [Frost et al., 1982]. The Frost filter is an exponentially damped circularly symmetric filter that uses local statistics [Frost et al., 1982; Shi and Fung, 1994]. Also, other adaptive filters in *ENVI* were tested but the Frost filter emerged as the best option to reduce the noise.

3.2 The USB650 Field Spectrometer Measurements

Section 3.2 involves the measuring of the leaves with the *USB650* field spectrometer for later comparisons of *G.arborea* plantation spectra. Here the description of the measurements of the leaf radiative properties of *G.arborea* leaves is presented. This approach is known as directional-hemispherical values for transmittance (τ_λ) or reflectance (ρ_λ) [Jones and Vaughan, 2010].

The Red Tide *USB650* fibre optic spectrometer from Ocean Optics [Ocean Optics, 2006] was used. The *USB650* was interfaced with the PASCO Xplorer GLX data-logger [PASCO, Nd]. A contact probe designed by CICANUM, from the University of Costa Rica was used. In order to take the samples from fresh leaves it uses a fibre optic *led* illumination source and a leaf clip assembly with a black panel face. The responsiveness of the *USB650* is from 350 to 1000 ηm (VNIR portion of the spectra) and it utilizes a Sony ILX511 linear silicon CCD array detector. The sensor has a sensitivity of up to 75 photons/count at 400 ηm of wavelength and an optical resolution of 2.0 of Full Width at Half Maximum (FWHM) of an instrument's response to a signal [Ocean Optics, 2006].

¹<http://www.vipertools.org>

The light enters the *USB650* and passes through the different components of the spectrometer, like SMA Connector, Slit, Filter, Collimating Mirror, Grating, Focusing Mirror. Then the detector *VNIR* collects the light received from the Focusing Mirror or L4 detector Collection Lens and converts the optical signal into a digital signal. Each pixel on the detector responds to the wavelength of light that strikes it, creating a digital response. The spectrometer then transmits the digital signal to the *Pasco* data-logger [[Ocean Optics, 2006](#)] (see Fig. 3.5). A reference and dark measurements to correct for instrument response variables were stored.

All *G.arborea* plantations in the south of Costa Rica were searched in order to measure the spectra. In the preparation survey, all information about area, location, accessibility, permissions, age and other conditions was extracted. The area was acquired from RapidEye images from 2009-2010, but secondary sources such as the owners were also consulted.

A basic inventory with a systematic sampling design for these plantations was performed. On circular plots of 12.64 m radius, each tree bigger than 5 cm diameter at breast height (DBH) (1.30 m); and the height (total and commercial) of three closest trees to the centre of the plot were measured. The main reason of using systematic sampling over other sampling was its simplicity, and the assurance that the entire *G.arborea* plantations were evenly sampled.

In order to take the destructive samples of leaves in trees of *G.arborea*, the permission for three different *G.arborea* plantations from edges of 6, 8 and 18 years was obtained. This was done in order to select three trees for each group of ages.

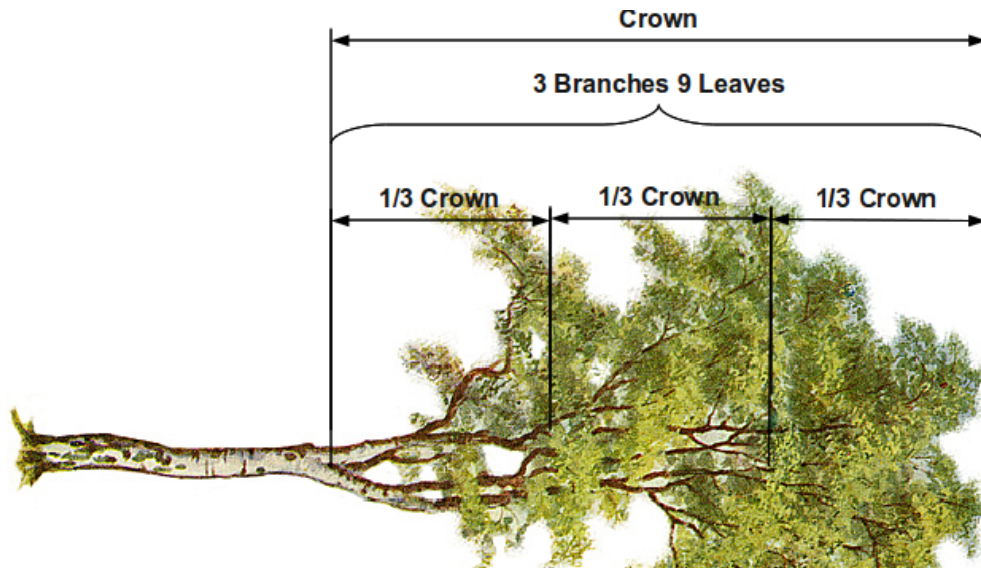


Figure 3.4: Crown sampling scheme in a tree.

With the basic inventory data, an analysis of diameter frequency distribution was made. Then all trees were divided in tree diameter classes. A sample of 9 cut trees was selected. For each tree the crown was divided in three sections (see Figure 3.4). In each crown section one branch was selected randomly and in each branch three leaves were also randomly selected. *In situ*, at each selected leaf four spectral signatures were taken with the fibre optic field spectrometer *USB650* [Ocean Optics, 2006]. This handle system was adapted by Research Centre on Nuclear and Molecular Sciences (CICANUM) from the University of Costa Rica. In total 81 leaves and 281 spectra were measured. A brief description of the system is presented below.



Figure 3.5: *USB650* Field Spectrometer.

A codification for the sampled leaves was defined. For example, the code *X103311055* was divided into five parts (see Tab. 3.3). The first one was assigned to the cover, in this case it was X10 for *G.arborea*. The second level was for the plantation age which has three levels 6, 8 and 18 years with their respective 11, 22, and 33 codes. In our example the code was a 33 which means that this sample has an age of 18 years. The third level was for the tree identification within the plantation. In this case it was number 1, meaning that this sample corresponded to the tree number one of the particular plantation of *G.arborea*. The fourth level was for the crown strata which have three crown levels (see Fig. 3.4). Again number 1 appeared, meaning that this leaf was taken from the low crown stratum in this tree. Finally, the fifth level was assigned to the leaf within the samples of the specific plantation. This three digit number is just a counting number which facilitated the work in the data frame format.

As shown in Table 3.3, all the covers except the *G.arborea* have only the first and the fifth level. That is because a sample design for taking the spectra of these covers was not performed. For these covers the important number is the first one because it permits to detect whether there are some spectral confusions with the *G.arborea*.

In the same field survey, samples of several leaves of teak, mangrove, rice, palm oil, and plantain were taken (see Tab. 3.3). To do so a block or a plantation of these land covers was selected. Then azimuth and a distance (within the block or in the plantation) from the access point were randomly selected. Once in this point, the samples were chosen. There were special difficulties with the palm oil, mangrove and teak trees, because of their height. Thus, only the lower part of these trees was sampled.

Table 3.3: Codification of selected covers for the *USB650* hyper-spectral field spectrometer.

Cover	Cover code	Plantation age (years)	Age code	Tree/code	Crown strata	Crown level code	Leaf code
G.arborea	X10	6	11	1	low	1	001
		8	22	2	medium	2	
		18	33	3	upper	3	
Teak	X20						001
Mangrove	X30						002
Grass	X40						003
Rice	X50						004
Palm	X60						005
Platain	X70						006

With the interpretation key presented in Table 3.3 it is possible to read the results presented in the Appendix C.5 as well as in Figure 4.19.

These spectral signatures were averaged by leaf and filtered with the "A Robust Adaptive On-line Filter for Multivariate Time Series" (*madore* filter) in the R *robfilter* package [Fried et al., 2011].

The final data frame used is composed of 100 spectra of *G.arborea* and 22 spectra of other covers. To obtain this reduced number of spectra for the *G.arborea*, the mean spectra per leaf was calculated. For the remaining 22 spectra the mean value was also calculated but this value is for the site where the sample was taken. For example, with teak (code X20), there were five spectra, this means that several samples from five different teak plantations were taken and the average for each

place was calculated. For plantain (code X70) there is just one mean spectra value, because only one plantain plantation in the study site was found in the field campaign.

Table 3.4 shows some characteristics of the hyper-spectral sensors used. In addition, the RapidEye characteristics are also presented.

Table 3.4: Characteristics of Optical Sensors (adapted from Beck [2003]; Cocks et al. [1998]; Jung-Rothenhusler et al. [2007]).

Parameters	Platform			
	RapidEye	EO-1 Hyperion	Hymap	USB 650
Spectral range (ηm)	440-850	400-2500	450-2500	350-1000
Spatial resolution (m)	5	30	2-10	-
Swath width (km)	77	7.7	6-9	-
Spectral resolution (ηm)	± 74	10	17	?
Temporal resolution (days)	1	200	independent	independent
Spectral coverage	discrete	continuous	continuous	continuous
Number of bands	5	242	100-200	NA
System	space borne	space borne	air borne	handled

3.3 Statistical Analysis of Hyper-spectral Data

This section involves the processing of the spectra of two sets of hyper-spectral datasets and the later analyses by Hierarchical Cluster Analysis (HCA). Using the same distance algorithm and linkage method, two different approaches of *HCA* were applied to the dataset to identify the spectral confusions in the spectra.

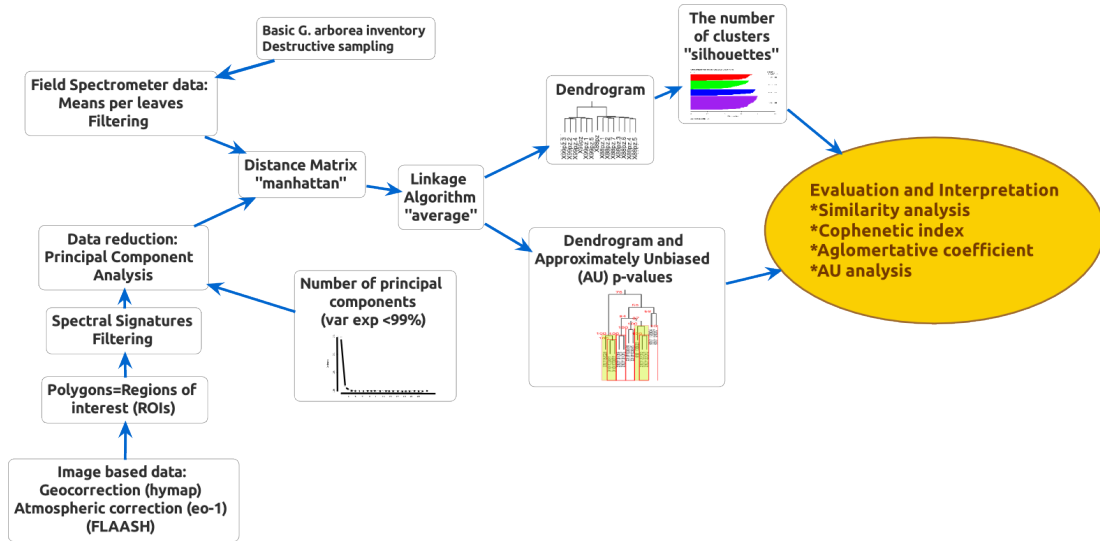


Figure 3.6: A schematic diagram of the general approach to perform the hierarchical cluster analysis and the similarity discrimination.

3.3.1 Principal Component Analysis of Hyper-spectral Data

Principal Component Analysis (PCA) was performed only for spectra of the four scenes (*hy_15*, *hy_14*, *eo1_pz* and *eo1_kw*) as the data reduction technique in order to select the highest variability of the spectra of the different *ROIs*.

All the filtered spectra were imported as data frames into R [R Development Core Team, 2011], and then into the *hyperSpec* package [Beleites, 2009]. After running the *PCA* algorithm for every group of *ROIs*, the scree-plots were plotted and the variability information of each result was printed. Following the suggestion of Izenman [2008] only the first principal components containing 99 percent of the variability were chose. The scree-plots graphed the component number against the explained variance of each group of spectra [Cattell, 1966]. A test of data normality was not applied because it was not essential [Reimann et al., 2008].

After running the *PCA* algorithm for every group of *ROIs*, the scree-plot was plotted and the variability information of each result was printed.

In addition to the scree plots, the *prcomp* function from the *stats* package was applied [Becker et al., 1988] as a method to analyse information contained in each *PCs*. The result is a list containing the coefficients of each component (sometimes referred to as loadings), the principal component scores, etc. The cut point was chosen when more than 98 % of the variability was explained by the first components.

3.3.2 Hierarchical Cluster Analysis

Hierarchical Cluster Analysis *HCA* is an unsupervised classification algorithm which generates a cluster of trees to partition a data set into subsets or clusters, where each cluster is based on a first, pairwise distance between all objects, and on a linkage or similarity algorithm [Duin et al., 1997; Torrecilla et al., res].

In the R statistical software [R Development Core Team, 2011], *HCA* was performed to all derived spectra. That includes the spectra derived from *PCs* in the case of the hyper-spectral scenes and the averaged spectra derived from the *USB650* field spectrometer.

The spectra data were modelled as a matrix. The matrix is of size $n \times m$, where $n =$ is the wavelengths at different intervals and $m =$ is the number of pixels that give a spectral signature. The rows, $i = 1, \dots, n$, are the spectral signatures which, for example, come from each original pixel on the hyper-spectral cube. In R statistical software [R Development Core Team, 2011], this refers to the element in row i and column j of matrix X .

As the first clustering approach, agglomerative *nesting* clustering with *agnes* function included in the *cluster* package [Maechler, 2012] of the R statistical environment was performed [R Development Core Team, 2011]. The *average* was used as clustering method. After several tests with different distance methods, the *manhattan* distance, which is the sum of absolute differences, was used [Izenman,

2008]:

$$d(x_i, x_j) = \sum_{k=1}^i |x_{ik} - x_{jk}| \quad (3.3)$$

The *average* method was used as a clustering algorithm. It linked the distance between two clusters, finding the average dissimilarity of each item in the first cluster to each item in the second cluster by computing [Izenman, 2008]:

$$d_{IJ,K} = \sum_{i \in IJ} \sum_{k \in K} \frac{d_{ik}}{(N_{IJ} N_K)} \quad (3.4)$$

where N_{IJ} and N_K are the numbers of items in clusters IJ and K , respectively.

Because of the availability of different clustering algorithms, the majority of used clustering methods was tested. Specifically the *ward*, *single*, *complete*, *average*, *mcquitty* and *manhattan* methods were computed. As the selection of the distance is the determining factor in the all clustering process, several distance methods specifically the *euclidean*, *maximum*, *manhattan* and *minkowski* were tested. The package *clue* was used [Hornik, 2012] in R to compute the dissimilarity between (ensembles) of hierarchies produced for every method.

To determinate (as an indicator) the number of cluster of each image spectra in R [R Development Core Team, 2011], the silhouette graphs were calculated. A silhouette graph is a graphical display technique where each cluster is represented based on the comparison of its tightness and separation [Rousseeuw, 1987].

To perform a second validation of the intrinsic groups, the Cophenetic Index [Rohlf, 1982; Sokal and Rohlf, 1962] and the Agglomerative Coefficient were used [Kaufman and Rousseeuw, 2005].

The Cophenetic Index (r_C) is a measure of how precisely two dendrograms preserve the pairwise distances between data objects [Rohlf, 1982; Sokal and Rohlf, 1962]. This index is computed from the cophenetic matrix (C) associated with each dendrogram. The elements of a cophenetic matrix ($c_{i,j}$) encode the distance between two objects (i, j), representing in the dendrogram the height of the link

at which those two objects are first joined. This height is the distance between the two clusters that are merged by this link. The Cophenetic Index r_C represents the correlation between two cophenetic matrices (C_1 and C_2) [Torrecilla et al., 2012].

$$r_C = \frac{\sum_i \sum_j (C_{1_{ij}} - \bar{C}_1)(C_{2_{ij}} - \bar{C}_2)}{\sqrt{\left(\sum_i \sum_j (C_{1_{ij}} - \bar{C}_1)^2\right) \left(\sum_i \sum_j (C_{2_{ij}} - \bar{C}_2)^2\right)}} \quad (3.5)$$

The Agglomerative Coefficient measures the clustering structure of the dataset [Kaufman and Rousseeuw, 2005]. For each observation i , $m(i)$ denotes its dissimilarity to the first cluster it is merged with, and is divided by the dissimilarity of the merger in the final step of the algorithm. The agglomerative coefficient is the average of all $1-m(i)$. It can also be seen as the average width (or the percentage filled) of the banner plot [Maechler, 2012].

These two agreement indices take values between 0 and 1. They indicate how closely two partitions of the same data set match each other. The closer to 1 the value, the more similarity between the two partitions, where 1 means a perfect match.

The standard R [R Development Core Team, 2011] functions to estimate the Cophenetic index, and the package *cluster* [Maechler, 2012] to access the Agglomerative Coefficient were used.

Based on the analysis presented in Reynolds et al. [2006], it is the *right* solution when the average silhouette width is the highest and when this value is closest to the real number of groups that are obtained for each scene.

The Silhouettes are constructed by [Rousseeuw, 1987]:

Let A be a specific cluster which contains i objects, then calculate The silhouettes

are defined as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3.6)$$

where:

$a(i)$ = average dissimilarity of i to all other objects of cluster A

$b(i)$ = minimum $d(i, C)$ where there are more than two clusters, and $d(i, C)$ is the average dissimilarity of i to all objects of a specific cluster.

The interpretation key in Table 3.5 was used as an interpretation key due to the silhouettes values [UNESCO, 2008]:

Table 3.5: Interpretation key for the silhouettes values (after [UNESCO, 2008]).

Values of $s(i)$	Interpretation
1	within dissimilarity $a(i)$ is much smaller than the smallest between dissimilarity. In other words, object i has been assigned to an appropriate cluster. The second best cluster B is not nearly as close as the actual cluster A .
0	$a(i)$ and $b(i)$ are approximately equal. Hence, it is not clear whether i should be assigned to A or B . It can be considered as an intermediate case.
-1	Object i is badly classified. When s is close to negative one, the object is poorly classified. Its dissimilarity with other objects in its cluster is much greater than its dissimilarity with objects in the nearest cluster. Why it is not in the neighboring cluster?

Pam [Maechler, 2012] was run several times with different values of k and then the optimal number of clusters based on highest average silhouette width SC was decided. The interpretation key as follows [UNESCO, 2008]:

$$SC = \max \bar{s}(k) \quad (3.7)$$

The maximum SC is a dimensionless measure which is taken over all k and this value summarizes how appropriate each object's cluster is. The silhouette of a cluster is a plot of the $s(i)$ ranked in decreasing order of all the objects i . The plot is a horizontal line whose length is proportional to $s(i)$. The silhouette shows which objects lie well within the cluster and which ones are merely somewhere in between clusters. A wide silhouette indicates large $s(i)$ values and hence a pronounced cluster [UNESCO, 2008].

Kaufman and Rousseeuw [2005] suggest that SC can be used for the selection of the best value of k , by choosing that k for which SC is maximal.

Table 3.6: Interpretation key for the average silhouette width.

Range of $s(i)$.	Interpretation
0.71-1.0	A strong structure has been found
0.51-0.70	A reasonable structure has been found
0.26-0.50	The structure is weak and could be artificial
< 0.25	No substantial structure has been found

As a second HCA and also as an evaluation technique R was used [R Development Core Team, 2011]. The package *pvclust* [Suzuki and Shimodaira, 2011], with the *manhattan* (see Eq.3.3) distance as the rule of clustering and an *average* (see Eq.3.4) linkage as an agglomerative method was used with 10000 bootstrap replications. With this method the Approximately Unbiased (AU) p -values was calculated. The probability p of 0.95 if the one cluster does not exist [Shimodaira, 2002] was taken. The null hypothesis was that the samples of the different spectra arrangements are merely due by chance, which means that the cluster does not exist at a defined significance level. In other words, the groups that are statistically significant cannot be reasonably divided into smaller clusters.

This method evaluates the accuracy of the clusters, where the accuracy means the certainty of the existence of a cluster. For example, if its p -value is less than α , the cluster is rejected at the α level of significance [Suzuki and Shimodaira, 2004].

The hypothesis H_i (the cluster does not exist) was rejected when $P_i < \alpha$ for a

level of significance of $0 < \alpha < 1$, of 0.05.

The *p-value* calculated by multi-scale bootstrap re-sampling is an approximation, however, it is less biased than bootstrap probability, which is also an approximation of *p-value* calculated by bootstrap re-sampling [Suzuki and Shimodaira, 2004].

3.4 TerraSAR-X Image Processing

Section 3.4 involves the processing of the TerraSAR-X imagery from the pre-processing to the accuracy assessment of the unsupervised classification. Figure 3.7, presents a schematic view of the TerraSAR-X process.

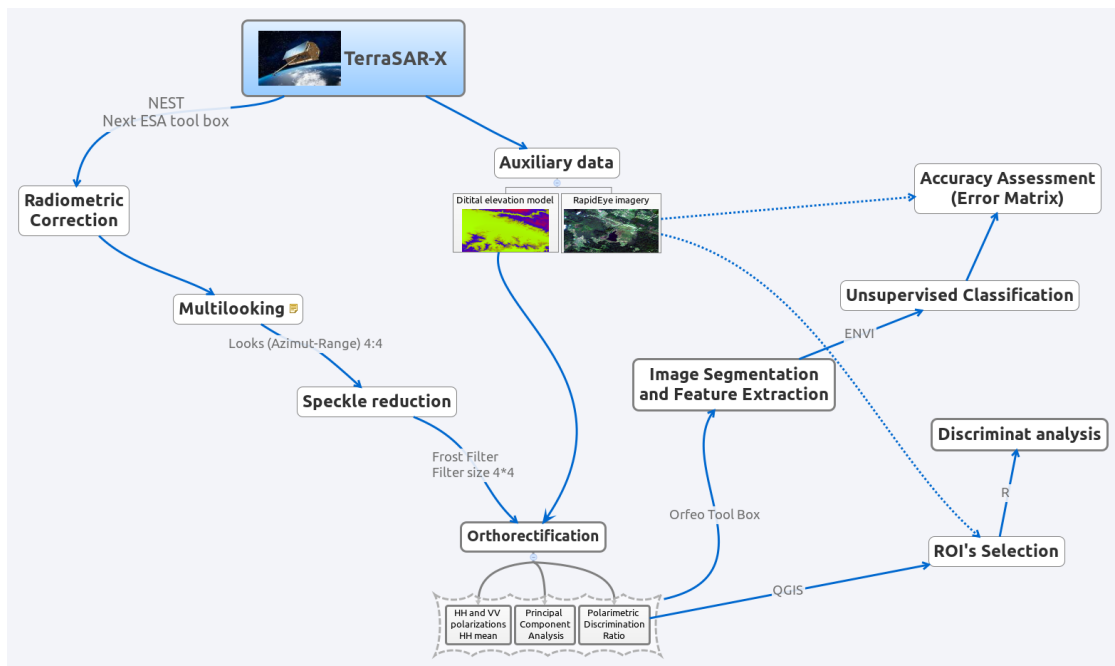


Figure 3.7: TerraSAR-X work-flow in south of Costa Rica.

3.4.1 The TerraSAR-X Imagery Description

All images acquired are in Single Look Slant Range Complex (SSC) mode. See Table 3.8. The main focus of *SSC* is the radar signal and the pixels are spaced equidistantly in azimuth and in slant range. The product contains all intensity information and phase information so that any other product can be generated at a later time [Roth et al., 2004].

In Strip Map Mode, the ground swath is illuminated with a continuous sequence of pulses while the antenna beam is fixed in elevation and azimuth. This results in an image strip with continuous image quality in azimuth [Roth et al., 2007].

Horizontal-Horizontal (*HH*), and Vertical-Vertical (*VV*) sets of polarization images were used as it is shown in Figure 3.8 and in Table 3.8 in both ascending and descending look directions.

Besides the polarization the differences are in incidence angles between two of the sets of *VV* images. These differences have an impact on the backscatter values.

As it is shown in Table 2.1, the TerraSAR-X is operated at X-band at 9.65 GHz or 3.1 cm with 15 revolutions around the world per day, 11 days revisit and an altitude at the Equator of 514.8 km. The TerraSAR-X mission was launched in 2007 [Scheuchl et al., 2009]. The X-Band mainly describes the surface of the objects because the radar beam is backscattered very close to the surface instead of penetrating the vegetation as the L-Band beams do. As Jensen [2007] pointed out, the general rule of thumb is that the microwave energy that could penetrate into a dry surface is equal to the wavelength of the *RADAR* system.

There are three different modes and polarizations (single and dual) where TerraSAR-X operates, which are: the ScanSAR, Stripmap and Spotlight, each of varying geometric resolution from 1 to 18 m. Quad polarization and along-track interferometry are possible on an experimental basis [DLR, 2007; Kuntz, 2010; Scheuchl et al., 2009].

The basic image products for the TerraSAR-X mission are [Faller and Weber, 2007]:

Table 3.7: Characteristics of TerraSAR-X mission products

Basic Image	Scene Size (km)	Polarization	Resolution	Price ^a
High Resolution SpotLight (HS)	10 x 5	Single VV or HH	up to 1 m	6750
SpotLight (SL)	10 x 10	Dual HH and VV	up to 2 m	6750
StriMap (SM)	30 x 50	All ^b	up to 3 m	3750
ScanSAR (SC)	100 x 150	Single VV or HH	up to 18 m	2750

^aStandard Scene in €

^bSingle VV or HH; Dual HH, VV and HV or VH

- Single Look Slant Range Complex (SSC) with amplitude and phase information in slant range geometry
- Multilook Ground Range Detected (MGD) corrected to WGS84 with an average terrain height for slant range to ground range projection
- Geocoded Ellipsoid Corrected (GEC) corrected to WGS84 with an average terrain height
- Enhanced Ellipsoid Corrected (EEC) corrected to WGS84 with a digital elevation model (DEM) .

3.4.2 TerraSAR-X Data Pre-processing

3.4.2.1 Radiometric Correction

In order to make the different measurement comparable, the images were calibrated. This process also facilitates the co-registration process and *PCA* fusion analysis. In *NEST* [ARRAY-ESA, 2011], for TerraSAR-X imagery, the co-registration operator performs absolute radiometric calibration by applying the simplified approach where the Noise Equivalent Beta Naught is neglected. Only calibration constant correction and an incidence angle correction are applied. For

Table 3.8: Acquisition parameters for *Stripmap* TerraSAR-X imagery.

Product ^a	Incidence angle ^b	Polarization	Pass direction ^c	Date	GRR ^d	AR ^e	Track/Orbit
SSC	40.99	HH	asc	2010/05/03	2.689	3.30	112/15977
SSC	40.89	HH	asc	2010/05/13	2.694	3.30	105/16137
SSC	39.20	HH	des	2010/04/10	2.791	3.30	112/15643
SSC	23.70	VV	des	2010/05/13	2.912	3.30	112/16144
SSC	40.98	VV	asc	2010/05/23	2.689	3.30	105/16304
SSC	23.70	VV	des	2010/05/24	2.912	3.30	112/16311
SSC	40.98	VV	asc	2010/06/03	2.689	3.30	105/16471

^a(SSC) Single Look Slant Range Complex.

^bAt scene centre coordinate.

^c(asc) ascending (des) descending

^d(GRR) Ground Range Resolution

^e(AR)Azimuth Resolution

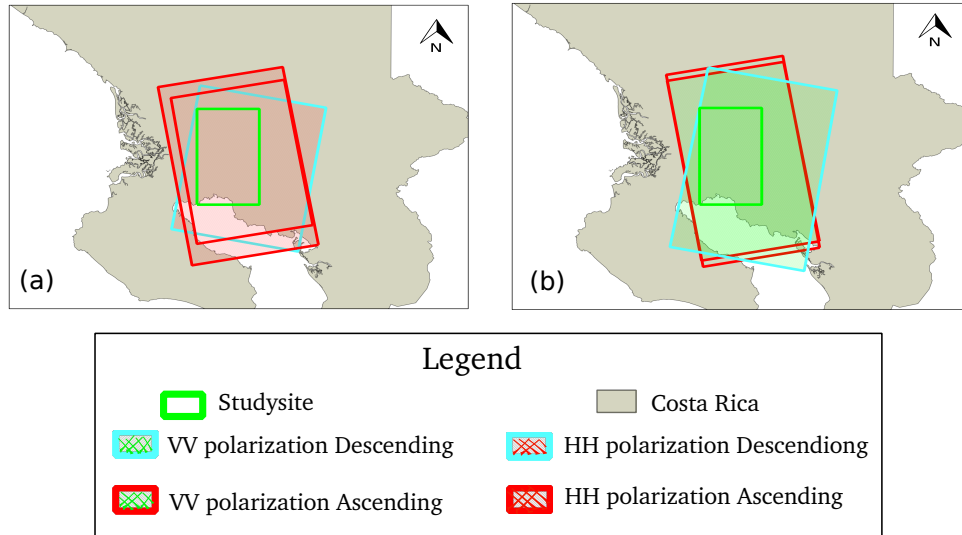


Figure 3.8: TerraSAR-X polarization modes in south of Costa Rica. (a) *VV* polarization and (b) *HH* polarization

the TerraSAR-X products calibration algorithm, the *NEST* algorithm is based in [Infoterra \[2008\]](#).

After the calibration process a set of images is generated in Sigma Naught (σ^0 , radar scattering coefficient in units of m^2m^{-2}), Beta (β^0) and Gamma (γ^0) values as shown below. The values of radar brightness β^0 are computed by:

$$\beta_{dB}^0 = 10 * \log_{10}(k_s * |DN|^2) \quad (3.8)$$

Where DN is the digital number in the imagery, k_s is the calibration factor provided by TerraSAR-X products.

The σ^0 values (*radar reflectivity*) per unit area in the ground range are derived from β^0 , but this time the incidence angle plays an important role in the equation, as follows:

$$\sigma_{dB}^0 = \beta_{dB}^0 + 10 * \log_{10}(\sin\theta_{loc}) \quad (3.9)$$

Where the θ_{loc} is the local incidence angle between the radar beam and the normal to the illuminated area, and dB are decibels.

3.4.2.2 Speckle Reduction

The speckle is a grainy, salt-and pepper pattern in radar images [[Jensen, 2007](#)]. Speckles are caused by random constructive and destructive interference of the de-phased but coherent return waves scattered by the elementary scatters within each resolution cell. Speckle noise reduction can be applied either by spatial filtering or multi-look processing. The reduction of this inherent speckled appearance is generally known as multi-look processing [[ARRAY-ESA, 2011](#)]. In the terminology of *RADAR* image processing these simultaneous (or sub-) images are called *looks* [[Richards, 2009](#)]. The number of looks is a function of the pixel spacing in the azimuth direction, pixel spacing in the slant range, and the incidence angle.

In order to enhance radio-metrically the set of images the number of looks was increased. However, this process reduced the geometric resolution, compared with the spatially enhanced product, for example. For that reason, the speckle of the radio-metrically enhanced image is reduced, resulting in a better radiometric resolution.

In *NEST* [ARRAY-ESA, 2011] the multi-look images were produced by a frequency-domain method using the sub-spectral band width. This means that the multi-look process is produced by space-domain averaging of a single look image, either with or without specific 2D kernels by convolution [ARRAY-ESA, 2011]. The number of range looks and the number of azimuth looks for every image that was used is shown in Table 3.9. This option was computed based on the ground range spacing and the azimuth spacing. The window size is then determined by the number of range looks and the number of azimuth looks. As a result, an image with approximately square pixel spacing on the ground is produced.

After the multi-look process, in *NEST* [ARRAY-ESA, 2011], a *Frost* filter was applied [Frost et al., 1982] in order to further reduce the speckle. The filter size in X and Y was 4×4 , with a Damping Factor of 2.

Table 3.9: Multilook parameters of the TerraSAR-X imagery.

Product.	Looks-[Azimuth-Range]	Ground Range Resolution (meters)
HH-2010/04/10	4:4	8.76
HH-2010/05/01	4:4	7.94
HH-2010/05/12	4:4	8.17
VV-2010/05/03	4:4	7.94
VV-2010/05/13	4:6	9.31
VV-2010/05/23	4:6	9.31
VV-2010/06/03	4:4	7.94

3.4.2.3 Terrain Correction

An interpolated Digital Elevation Model (*DEM*) at 5m spatial resolution from the digitalized iso-lines available in the topography maps of Changuena [National

Geographic Institute of Costa Rica, 1980] and Rincón was used [National Geographic Institute of Costa Rica, 1979].

The main steps in the terrain correction algorithm, after ARRAY-ESA [Nd] are:

- *SAR* simulation: Generate simulated *SAR* image using *DEM*, the geocoding and orbit state vectors from the original *SAR* image, and mathematical modelling of *SAR* imaging geometry. The simulated *SAR* image will have the same dimension and resolution as the original image.
- Co-registration: The simulated *SAR* image (master) and the original *SAR* image (slave) are co-registered and a *WARP*¹ function is produced. The *WARP* function maps each pixel in the simulated *SAR* image to its corresponding position in the original *SAR* image.
- Terrain correction: Traverse *DEM* grid that covers the imaging area. For each cell in the *DEM* grid, compute its corresponding pixel position in the simulated *SAR* image is computing using *SAR* model. Then its corresponding pixel position in the original *SAR* image can be found with the help of the *WARP* function. Finally, the pixel value for the ortho-rectified image can be obtained from the original *SAR* image using interpolation.

A 1000 Ground Control Point (*GCP*) was used in the final co-registration. The interpolation methods used was the cubic convolution.

The final product is a projected and re-sampled Universal Transversal Mercator *UTM*, zone 17 north using the *WGS84* reference ellipsoid. The final pixel spacing is equidistant in northing and easting. Terrain induced distortions are corrected and the pixel location of these products is very accurate.

The *DEM* from the Shuttle Radar Topography Mission (*SRTM*) was also tested visually [USGS- U.S. Geological Survey, 2006] with 30 m spatial resolution supplemented with elevation data of similar quality from other sources, but the best results were from the interpolated topo-maps *DEM*.

¹The *WARP* function is a type of geometric transformation, that is concerned with the registering of an image to a common coordinate system or one data model.

3.4.2.4 Selection of TerraSAR-X Processed Bands

To perform the *PCA* to the TerraSAR-X imagery as a fusion technique, first a layer stacking for each polarization mode was made. This new multi-band file came from the exported and processed TerraSAR-X images from previous steps.

For discriminating the soil and vegetation properties of the scattering surfaces, the polarimetric discrimination ratio (*PDR*) was used [Singh, 2006]:

$$PDR = \frac{\sigma_{VV} - \sigma_{HH}}{\sigma_{VV} + \sigma_{HH}} \quad (3.10)$$

A mean band of *HH* processed bands was calculated but the average of the *VV* images was not calculated because it was considered that in this case the *VV* images did not increase the information for the performed analysis. However, the *VV* images were used to calculate the *PCs* and to calculate one average of this information (see Tab. 4.9).

3.4.3 Image Classification of TerraSAR-X Processed Bands and the RapidEye imagery description

RapidEye is commercial constellation of 5 sensors that deliver 5 spectral bands (blue, green, red, red-edge, NIR) covering the range of 440 to 850 nm of the spectrum. The design system allows regular monitoring of large areas because of the 77 km swath width. There is daily access to any location on the globe and processing of the image data into information products is done within 24hrs [Jung-Rothenhusler et al., 2007]. Other characteristics are presented in Table 3.4.

The RapidEye imagery was acquired between December 2009 and January of 2010 in GeoTIFF product format in *UTM* zone 17N and WGS84 ellipsoid. A mosaic of the RapidEye imagery is presented in Figure 1.3.

The exploratory image classification is divided into two parts. The first consists of the TerraSAR-X bands used (see Tab. 4.9) plus the *PDR* index to perform image segmentation for which the ortho-rectified derived bands were exported to 8-bit

format, following the suggestion of Clausi [2002], who showed that increasing grey levels do not necessarily increase the separability, but also serve as data reduction to reduce the computational effort. Then, the images were exported into *GeoTIFF* format using the Orfeo Toolbox (*OTB*) [OTB Development Team, 2012].

The purpose of the *SAR* image segmentation is to partition an image into regions of different characteristics, in other words, the purpose is to assign a specific class label for each pixel [Zhang et al., 2008].

Descriptive features based on the Gray Level Co-occurrence Matrix (*GLCM*) [Haralick et al., 1973] defined in the software were extracted. The filters computed include; mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation. The *GLCM* method is a way of extracting second order statistical texture features [Albregtsen, 2008]. It is beyond the scope of this thesis to explain each parameter.

After the image segmentation, a subset of input variables was selected by eliminating features with little or no additional information. The idea is to improve the comprehensibility of the resulting classifier models [Kim et al., 2002].

An *isodata* unsupervised classification was used as an exploratory analysis of the results from the previous steps in order to perform a forest/non-forest classification, but also to detect the distribution of the different classes in the whole study area. The *isodata* is an *ENVI* [ITT Visual Information Solutions, 2012] module that calculates the class means evenly distributed in the data space, then iteratively clusters the remaining pixels using minimum distance techniques. All the bands used in the segmentation phase are described in Table 4.9. The segmented image was classified in eight classes, as described in section 1.3.3.

After the *isodata* process, 20 classes were again classified into 9 classes which were the basis for the forest/non-forest classification. The reference data (truth) was acquired using actual RapidEye imagery as well as aerial photographs and the ecosystem descriptions after Kappelle et al. [2002]. For the construction of the error matrix for the classes forest/non-forest a total of 143 sample points was used. The systematic sample design is shown in Figure 1.3. The systematic sampling

design was chosen because it ensures good distribution of samples across the study area landscape and because no other pattern correlated with the landscape was found.

The error matrix is a square array of numbers set out in rows and columns that expresses the number of sample units assigned to a particular category (i.e forest) in one classification relative to the number of sample units assigned to a particular category in another classification [Congalton and Green, 2009].

3.4.4 Analysis of the Land Cover Distributions

Table 3.10 list the codes and a short description of the land covers as *ROIs* that were selected in a subsection of the TerraSAR-X study area. Then a total of 11 *ROIs* was allocated in order to describe the distribution of the pixels contained in these *ROIs* in the different selected bands.

A more detailed description of these land covers is available in section 1.3.3.

Table 3.10: Selected regions of interest for the density distribution analysis of the TerraSAR-X process.

Code	Description
soil	Bare soils
palm_young	Palm oil (young plantations)
forest_gently_undulating	Forest in gently undulating and undulating slopes
pastures	Pastures principally for cattle
palm_old	Palm oil (old plantations, more than 7 years)
infrastructure	Infrastructure (man-made constructions)
herbland	Herblands (natural graminoids)
mangrove	Mangroves
water	Natural and artificial water bodies
forest_flat_raffia	Forest in flat terrains dominated by raffia
forest_rolling	Forest in rolling slopes

In order to avoid mixed pixels only the central part of the image object was mapped by extracting the *ROIs*.

RapidEye actual imagery was used as additional source of actual information, this was however done only when one class could not be assigned from the TerraSAR-X-1 processed bands alone.

A linear discriminant analysis (LDA) was done on the data of the TerraSAR-X processed images (see Tab. 4.9) to evaluate the dissimilarities of selected *ROIs* (see Tab. 3.10) of the study area, and also to evaluate how within each group covariances of the *ROIs* are homogeneous. With the evaluation of the estimation of the group covariances matrix, the possible problems of the processed *TerraSAR-X* bands can be explained.

The histograms were plotted since they are a useful tool for displaying distributions as well as indicating the shape of the underlying density function. The density function was calculated as smooth curves in order to discuss and describe the distributions of the selected covers (*ROIs*). The R [R Development Core Team, 2011] statistical software and the package *ggplot2* was used [Wickham, 2009]. The plots are presented in Appendix D and are divided into two main groups. First, all band per land cover, and then, each land cover per band.

Usually in remote sensing, when a sample is taken from an image, even for exploratory analysis or for an accuracy assessment, the polygons or *ROIs* have different sizes. In this case, the comparison of the distribution with the density kernel function instead of the histograms is recommended [Wickham, 2009].

Chapter 4

Results

This Chapter the results in of the hyper-spectral data as well as the TerraSAR-X imagery are presented. The hyper-spectral part contains the scene and the USB650 spectra.

4.1 Hyper-Spectral Analysis from the Four Scenes

In Section 4.1 the results for the four scenes, two from the space-borne EO-1 Hyperion and two for the airborne HyMap data, are described. These results are divided in three sections. First, the general remarks of the four scenes spectra, the exploratory review of the number of clusters and the evaluation parameters of each scene clustering solution. The second part is the description of the clustering process per scene which includes the two *HCA* approaches. The third section is a specific cluster error evaluation.

4.1.1 The Scene's Spectra

In this section only the standard deviation of the *PCs* of *G.arborea* in Figures 4.1 to 4.4 is presented because the graphical representation becomes illegible if all the standard deviations are shown in one graph.

For the vegetative spectra in the Figures 4.1 to 4.4, the typical patterns are shown. In the *NIR* the pigments do not absorb the electromagnetic radiation strongly. In this region, the *SD* rises for all the scene spectra, and there is strong water influence as chemical contributor to the reflectance patterns.

In these figures the water absorption bands are in the range of ≈ 1300 to 1500 nm and 1800 to 2100 nm . In this range I eliminated all the reflectance values for all the scenes, for that reason there is a straight line between these spectra ranges.

In Figures 4.1, 4.2 and 4.4, it is observed that *G.arborea* and palm oil PCs spectra from the *hy_15*, *hy_14* and *eo1_kw* are similar in shape and magnitude. This prospect of fair discrimination is particular clear in the *NIR* region. For the *eo1_pz* spectra the shapes and magnitudes are similar for *G.arborea*, palm oil and forest.

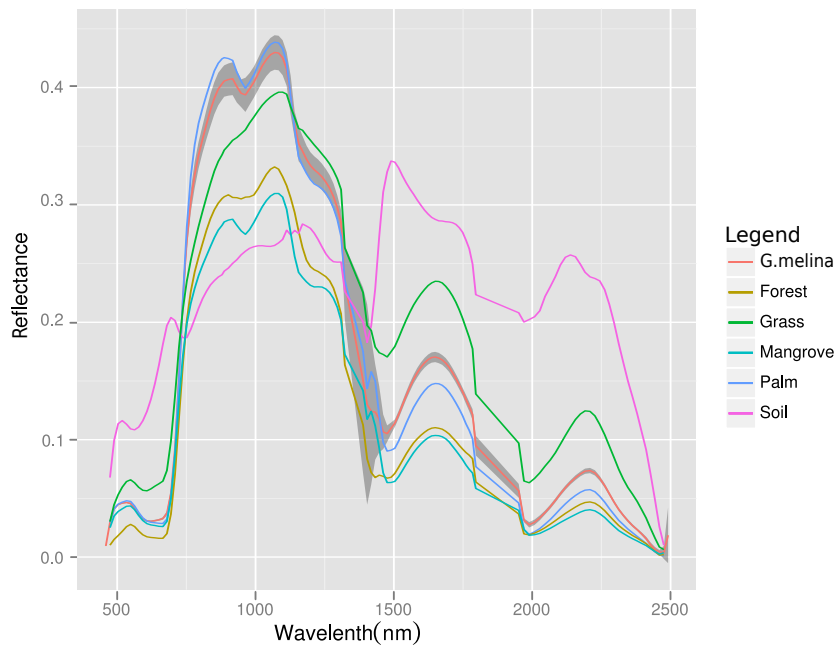


Figure 4.1: Selected spectral signatures in the *hy_15* scene.^a

^aThe shaded area is the mean \pm standard deviation

As shown in Figure 4.1, in the *PCs* spectra for the *hy_15* scene, the shaded area is overlapping with the line of the palm oil in the region of 950 and 1050 nm .

Also in this region the thickness of the line is the greatest.

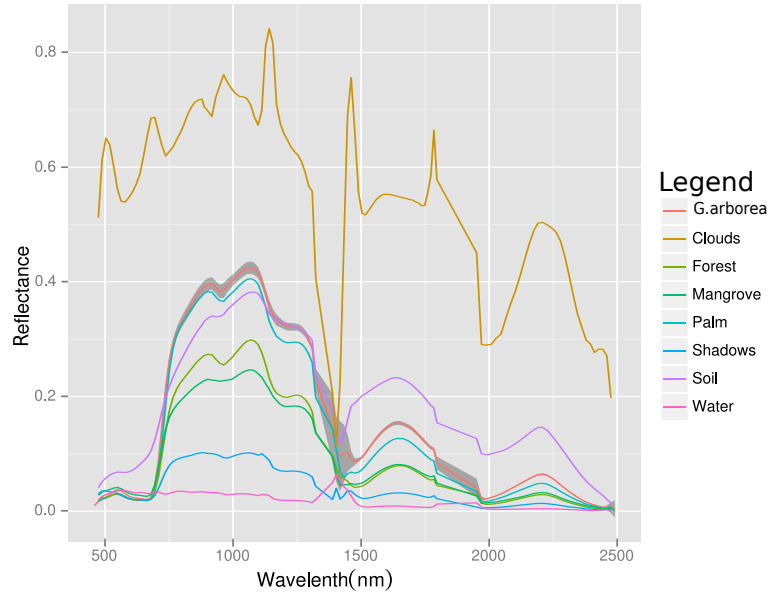


Figure 4.2: Selected spectral signatures in the *hy-14* scene.^a

^aThe shaded area is the mean \pm standard deviation

Table 4.1 lists the cophenetic index as well as the agglomerative coefficient. Both numbers can be used as some kind of measure of how well a particular dendrogram fits.

In Figure 4.6 the average silhouette width by number of clusters is shown. This relationship shows the usefulness of the graphic silhouette. For example, scenes *hy-14* and *eo1_pz* show high values of the average silhouette width, however, this is not the best solution in terms of number of clusters. In general, all the scenes tend to have low values as the number of clusters increases.

The highest average value recorded image is *eo1_kw* with 0.86 for the solution of 5 clusters, however, the number of clusters in the scene *eo1_pz* has one of the lowest values with 0.59.

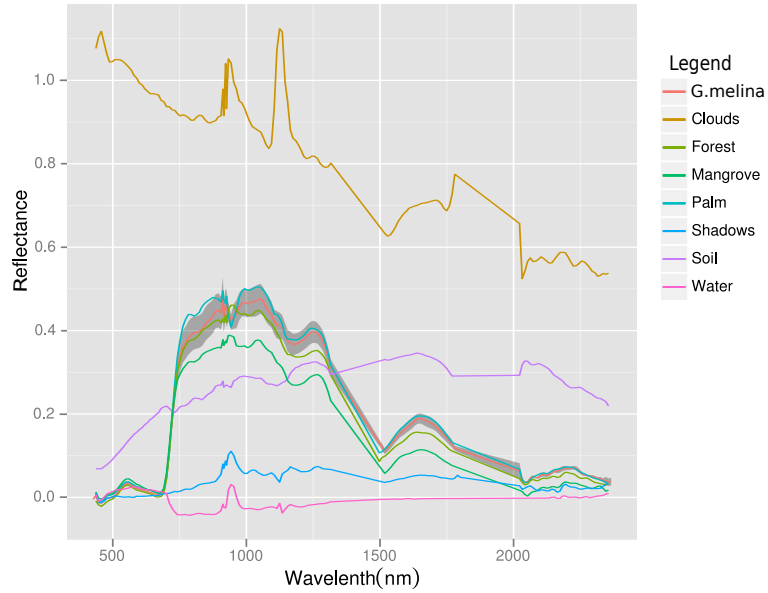


Figure 4.3: Selected spectral signatures in the *eo1_pz* scene.^a

^aThe shaded area is the mean \pm standard deviation

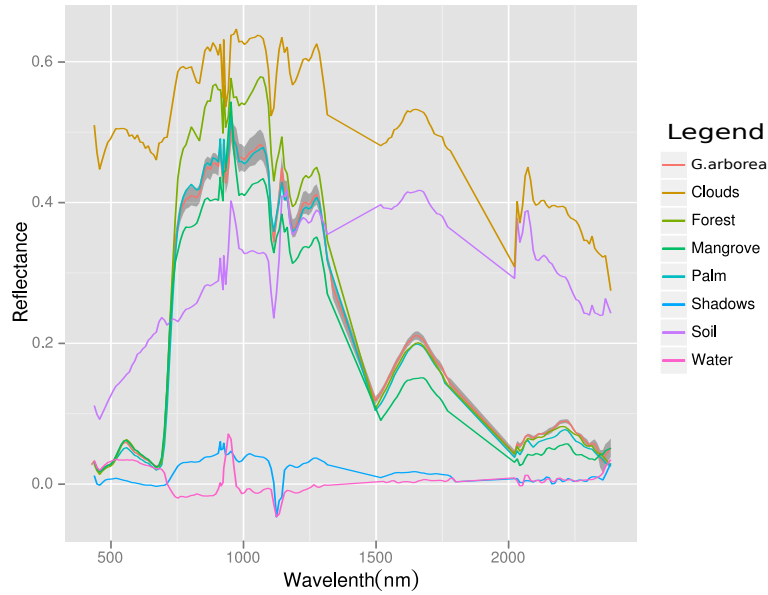


Figure 4.4: Selected spectral signatures in the *eo1_kw* scene.^a

^aThe shaded area is the mean \pm standard deviation

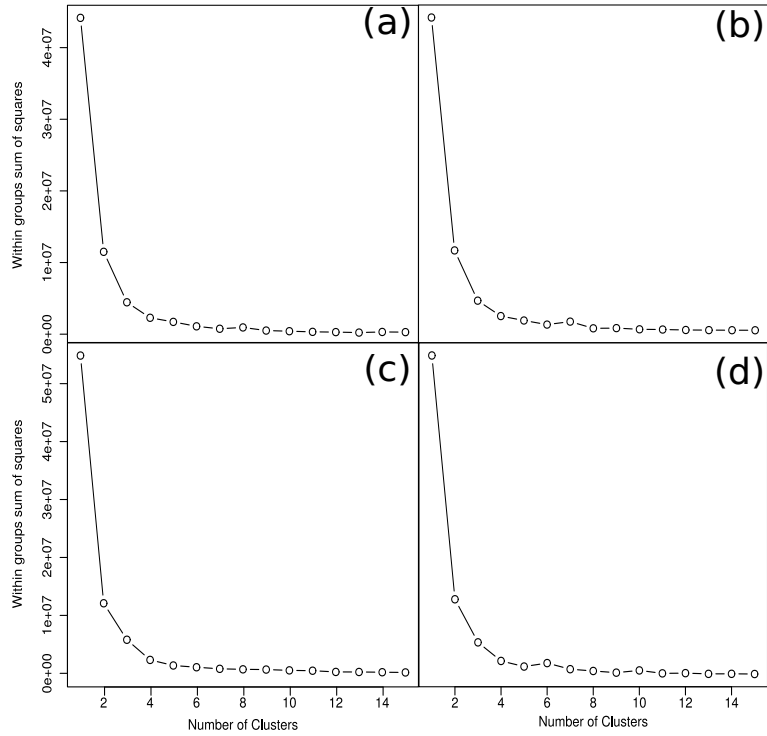


Figure 4.5: Silhouette graphs for the PCs spectral signatures of the scenes analysed.

(a) hy_{15} , (b) hy_{14} , (c) $eo1_{pz}$ and (d) $eo1_{kw}$

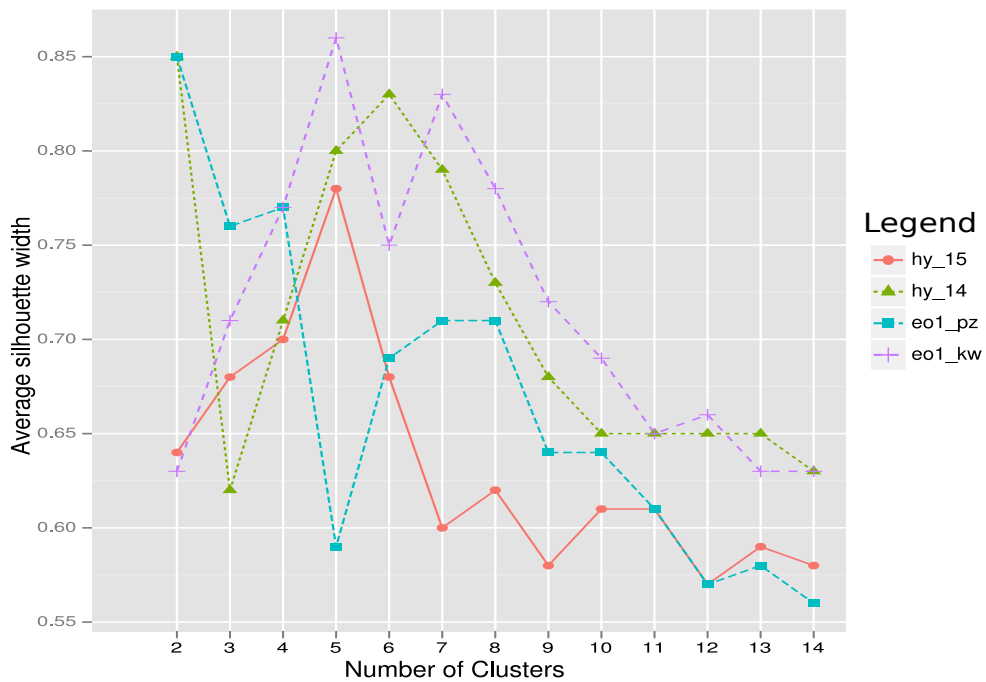


Figure 4.6: Average silhouette width for different number of clusters solutions for the space-borne and air-borne scenes

Table 4.1: Hierarchical clustering parameters for the four remote base sensor spectral signatures

Scene	Cophenetic index	Agglomerative coefficient	Average silhouette
<i>hy_15</i>	0.89	0.95	0.78
<i>hy_14</i>	0.87	0.98	0.83
<i>eo1_pz</i>	0.93	0.99	0.77
<i>eo1_kw</i>	0.92	0.99	0.83

Figures 4.7,4.9,4.11 and 4.13 show the first approach of the hierarchical cluster analysis performed for the input data consisting of the *PCs* of 4 hyper-spectral scenes. The optimal number of clusters (k) is presented in the silhouette graphs. In all the dendrogram graphs, the horizontal line represents the height where the solution k gives the best average silhouette width value.

The average of the silhouette is the *SC* of the Equation 3.7.

The second part of the clustering process (see Fig.3.6) for the *PCs* spectral signature scenes is presented in Figures 4.8,4.10,4.12 and 4.14. The values at the branches are *p-values* in percentage and the coloured boxes are some selected clusters that have a *p-value* higher than the α level of significance or in other words k clusters with $AU \geq 95\%$.

4.1.2 The *hy_15* scene

In this section the tow *HCA* approaches are presented for the *hy_15* scene.

The best average silhouette width for the *PCs* spectra in the *hy_15* scene is 0.78, this value is possible with $k=5$. This differs according to the number of *ROIs* that was originally selected. The agglomerative coefficient and cophenetic index are 0.95 and 0.89 respectively (see Table 4.1). Higher values indicate that a strong structure has been found between the different land covers analyzed for this scene.

Figure 4.5 (a) presents another way to see and to detect how many clusters are present in the *hy_15 PCs* spectra. As shown in this figure, after approximately 5 clusters, the within groups sum of squares tend to be constant.

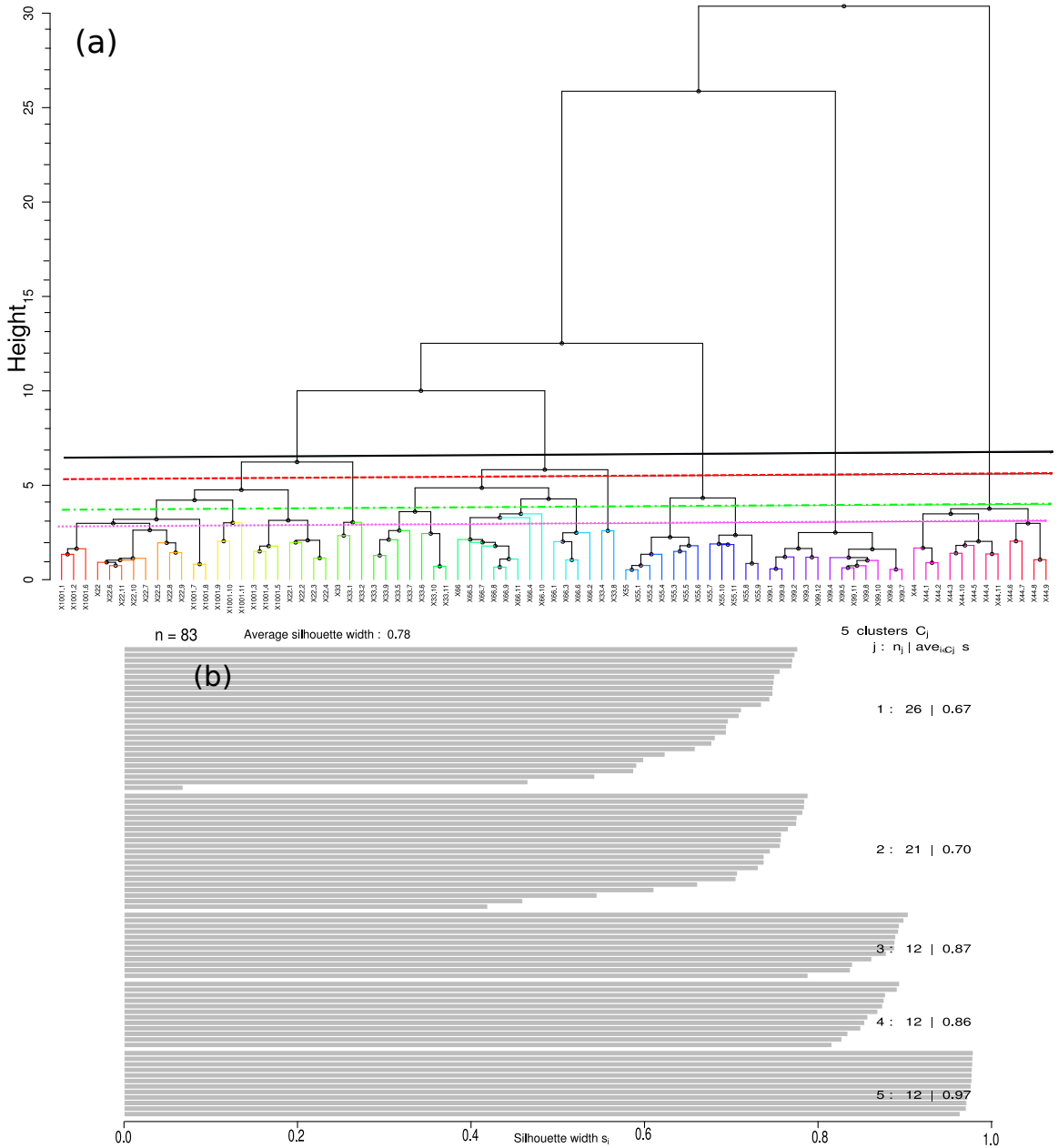


Figure 4.7: Hierarchical cluster analysis and silhouette graph of 83 PCs spectra of hy_{15} scene.

(a) Dendrogram obtained from 83 PCs spectra in the hy_{15} scene. The black horizontal line indicates the number of clusters with the SC maximum. The dashed red line indicates the original number of land covers (7). (b) Silhouette graph for the $k=5$ with the best average silhouette width for the hy_{15} scene.

Figure 4.7 illustrates the results from the cluster analysis applied to the *PCs* spectral signatures of the scene *hy_15* over the entire spectral range from 474 ηm to 2476 ηm . The horizontal line indicates the cut height where the five best k clusters could be cut. If this cut height is taken as the best number of k clusters (see Fig. 4.7 (b)), then there are spectral confusions or it is not possible to separate the *G.arborea* from the palm and from the three spectra of forest. Also the mangroves and the forest present spectral confusions at this reference height. However, there is good separability from all land covers if the cut height is taken as the magenta dotted line is taken as reference. Also, in this figure we can see where the problems are. For example, the first cluster has 26 spectra with an average silhouette width of 0.67. If we follow the decision key presented in Table 3.6, only one reasonable structure has been found for this cluster arrangement.

Table 4.2: Statistics for the approximately unbiased *p-values* and the *SE* for the *PCs* of the *hy_15* scene data

Statistic	AU	SE.AU
Min.	0.5468	0.000000
1st Qu.	0.7503	0.001108
Median	0.9251	0.004328
Mean	0.8777	0.004705
3rd Qu.	0.9951	0.007868
Max.	1.0000	0.031868

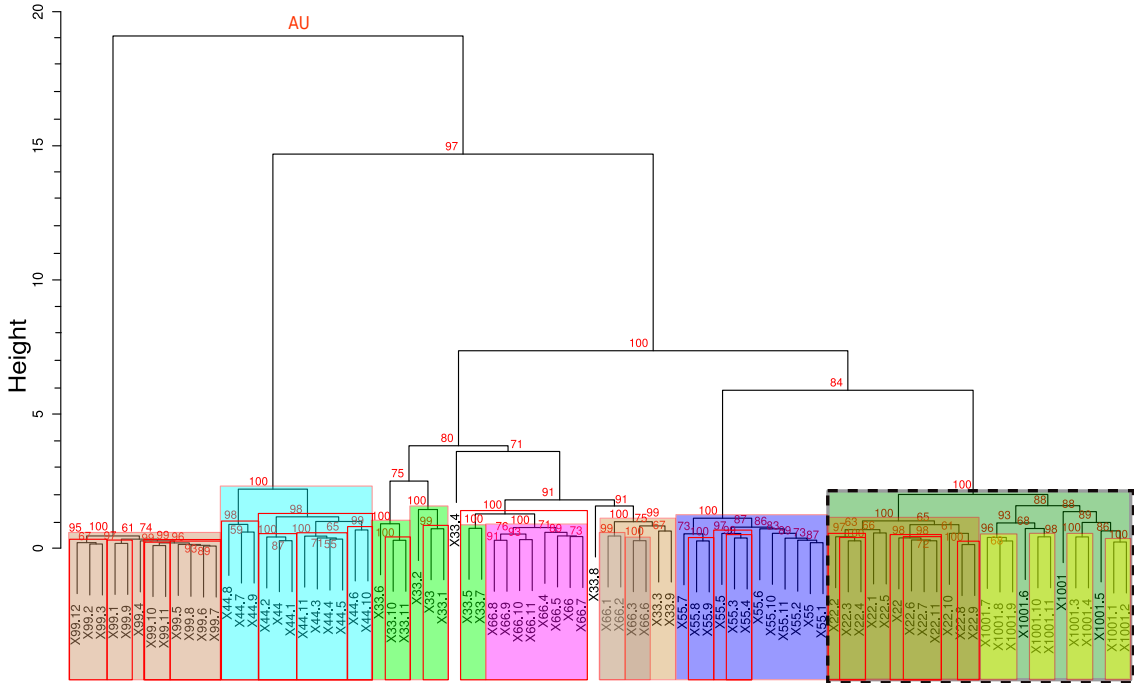


Figure 4.8: Approximately unbiased (AU) p -value with the *manhattan* distance and *average* as the cluster method for 83 PCs spectral signatures of *hy_15* scene.

The dendrogram in Figure 4.8 gives the cluster arrangement with the approximately unbiased p -values. Some of the most important clusters were squared. For example, the dashed line squares a cluster which mixes palm oil and *G.arborea*. However, within this cluster, there are other clusters. One of these separates the palm oil completely.

The median of 0.9251 (see Tab. 4.2) indicates that half of the AU values are below this value. All the clusters with a value less or equal to 0.95 are rejected. For example, and coming back to the cluster within the dashed line, only three spectra of *G.arborea* do not form a cluster with other *G.arborea* despite being close to another *G.arborea* spectra.

4.1.3 The hy_14 scene

The first clustering process for the *PCs* spectral signatures of the *hy_14* scene are presented in Figure 4.9. Also, the clustering behaviour is different with respect to the number of types of *ROIs* that were selected for this image (a total of 8). The agglomerative coefficient and cophenetic index are 0.98 and 0.87 respectively (see Table 4.1). This degree of similarity is very similar to the one obtained for the *hy_15*.

Figure 4.8 shows the results of the uncertainty in the hierarchical cluster analysis applied to the *PCs* spectral signatures of the scene *hy_15*. With this approach, there are spectral confusion with *G.arborea* when the cluster with the dashed line in this figure is selected. The cluster within the dashed line has an *AU* of 0.99681 and a *SE* of 0.00318, which indicates that the cluster exists. However, within this cluster there are one cluster of palm oil and four clusters of *G.arborea*. For three of the 12 *PCs* of *G.arborea* it was not possible to separate the palm oil. The rest can be separated into four different *G.arborea* clusters. There are no spectral confusions between forest and mangroves, however, they are very close in the dendrogram. The mangrove formed 8 *PCs* spectra in one cluster and the remaining 4 (of a total of 12) *PCs* are in two clusters.

In a superior height, there is a cluster formed by 4 *PCs* of mangrove and two of forest, in this case there is a spectral confusion between these two land covers. For two of the *PCs* of forest it is possible to form a cluster at the height of 6. This indicates that if these two *PCs* are clustered, then the forest, mangrove, grass, palm and *G.arborea* also have to be clustered. There is no problem clustering the water and the soils.

Figure 4.5 (b) presents another way to see and detect how many clusters could be present in the *hy_14* *PCs* spectra. As it is shown in this figure, after approximately 8 clusters the within groups sum of squares tends to be constant.

Figure 4.9 illustrates the results from cluster analysis applied to the *PCs* spectral signatures of the scene *hy_14* spectra presented in Figure 4.2 and the spectra over the entire spectral range from 474 ηm to 2476 ηm . The horizontal line indicates

the cut height where the five best k clusters could be cut. In this case, this height cut the dendrogram into 6 clusters, one more than the *hy_15* scene. Again, if this cut height is considered as the best number of k based on the Figure 4.9 (b), then there are spectral confusions or it is not possible to separate the *G.arborea* from the palm. Also, the mangroves and the forest presents spectral confusions at this reference height. However, there is good separability from soils, clouds and grass. In this figure, for example, the first cluster has 23 *PCs* spectra with an average silhouette width of 0.78. This cluster is very similar to the one for the *hy_15*, just without the three forest spectra. If we follow the decision key presented in Table 3.6, a strong structure has been found for this cluster arrangement.

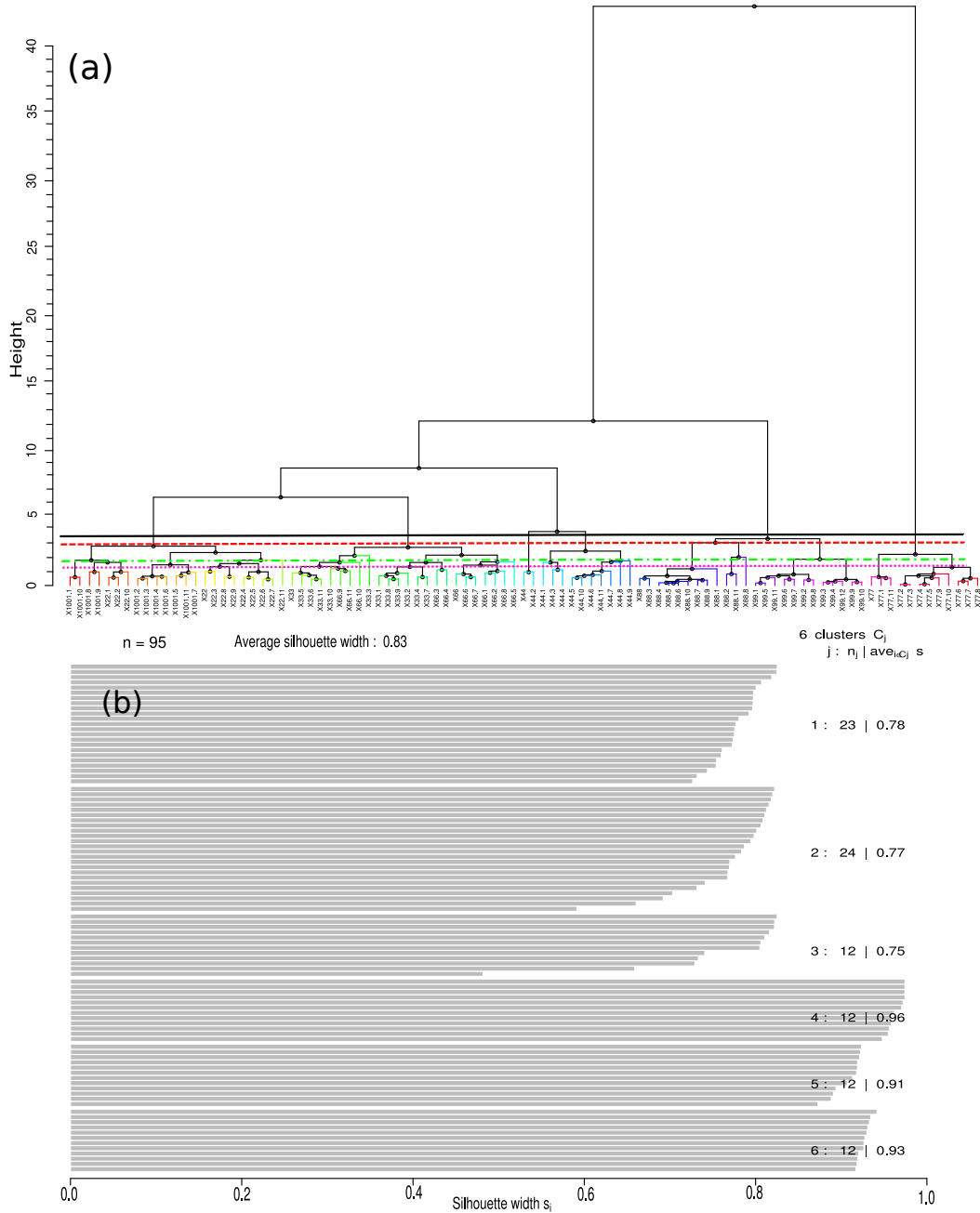


Figure 4.9: Hierarchical cluster analysis and silhouette graph of 95 PCs spectral signatures of hy_{14} image

(a) Dendrogram obtained from 95 PCs spectral signatures in the hy_{14} scene. The black horizontal line indicates the number of clusters with the SC maximum. The dashed red line indicates the original number of land covers (8). The green dot-dashed line is an arbitrary cut height. (b) Silhouette graph for the $k=6$ with the best average silhouette width for the hy_{14} scene.

In Figure 4.10 the results from of the uncertainty in the hierarchical cluster analysis applied to the *PCs* spectral signatures of the scene *hy_14* are shown. With this approach, there are no spectral confusions with *G.arborea* when the cluster with the dashed line (the X1000 spectra) in this figure is selected. This cluster within the dashed line has an *AU* of almost 1 and a low *SE*. This indicates that the cluster exists. There is very good separability, also for the forest (see the dotted line in the figure), forming a cluster with three clusters of forest at an inferior height.

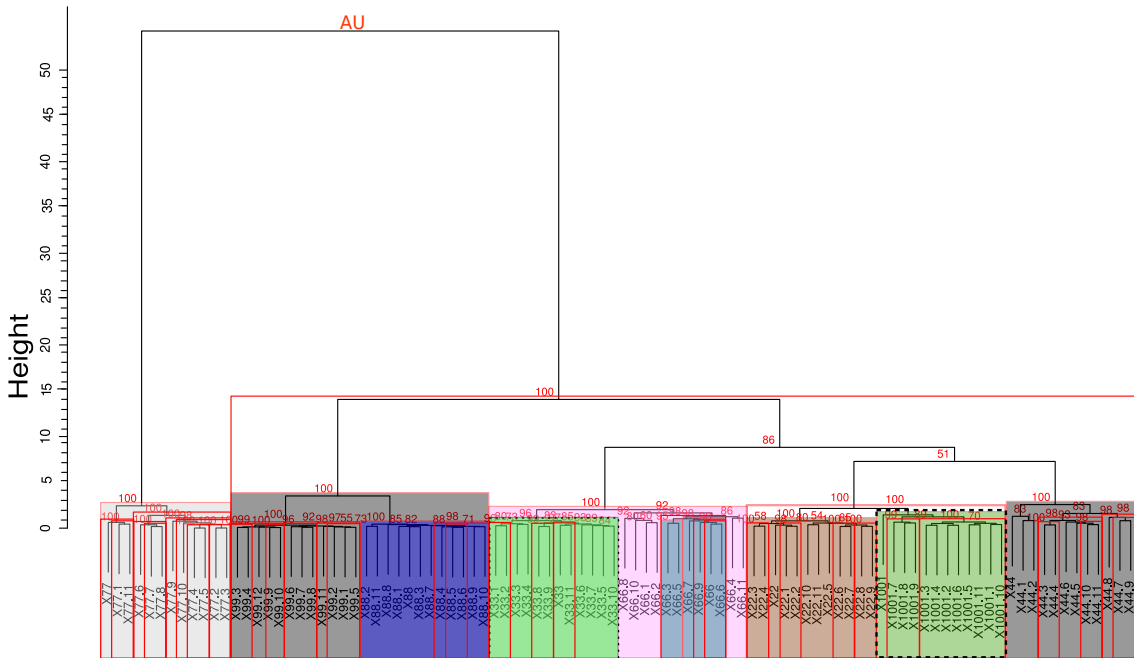


Figure 4.10: Approximately Unbiased (AU)*p-value* with the *manhattan* distance and *average* as the cluster method for 95 *PCs* spectral signatures of *hy_14* image.

4.1.4 The eo1_pz scene

The first part of the clustering process for the *PCs* spectral signatures of the *eo1_pz* scene are presented in Figure 4.11. Clustering behaves differently with respect to the number of types of *ROIs* that were selected for this image. The

Table 4.3: Statistics for the Approximately Unbiased p -values and the SE for the PCs of the hy_{-14} scene data

Statistic	AU	SE.AU
Min.	0.5149	0.000000
1st Qu.	0.8237	0.0008406
Median	0.9543	0.0033567
Mean	0.8813	:0.0041172
3rd Qu.	0.9955	0.0076124
Max.	1.0000	0.0103644

agglomerative coefficient and cophenetic index are 0.99 and 0.93 respectively (see Table 4.1). This degree of similarity is superior to that obtained for the hy_{-15} and hy_{-14} .

For the mangrove, there are spectral confusions with forest for 6 PCs , indicating that it is not possible to totally separate these covers. For the rest of the covers there are no spectral confusion. The covers of shadows, water, clouds, grass and soils have a good separability for such an uncertain dendrogram arrangement.

Figure 4.11 shows the results from cluster analysis applied to the PCs spectral signatures of the scene $eo1_{-pz}$ presented in Figure 4.3 and entire spectral range from 427 ηm to 2395 ηm . The horizontal line indicates the cut height where the four best k clusters could be cut. Again, if this cut height is considered as the best number of k based on Figure 4.11 (c), there are spectral confusion or it is not possible to separate within vegetation. Also the grass presents spectral confusion at this reference height. However, there is good separability from soils, clouds, water, and shadows. In this figure we can see where the problems are. For example, the first cluster has 49 PCs spectra with an average silhouette width of 0.70. If the decision key presented in Table 3.6 is followed, a reasonable structure has been found for this target cluster.

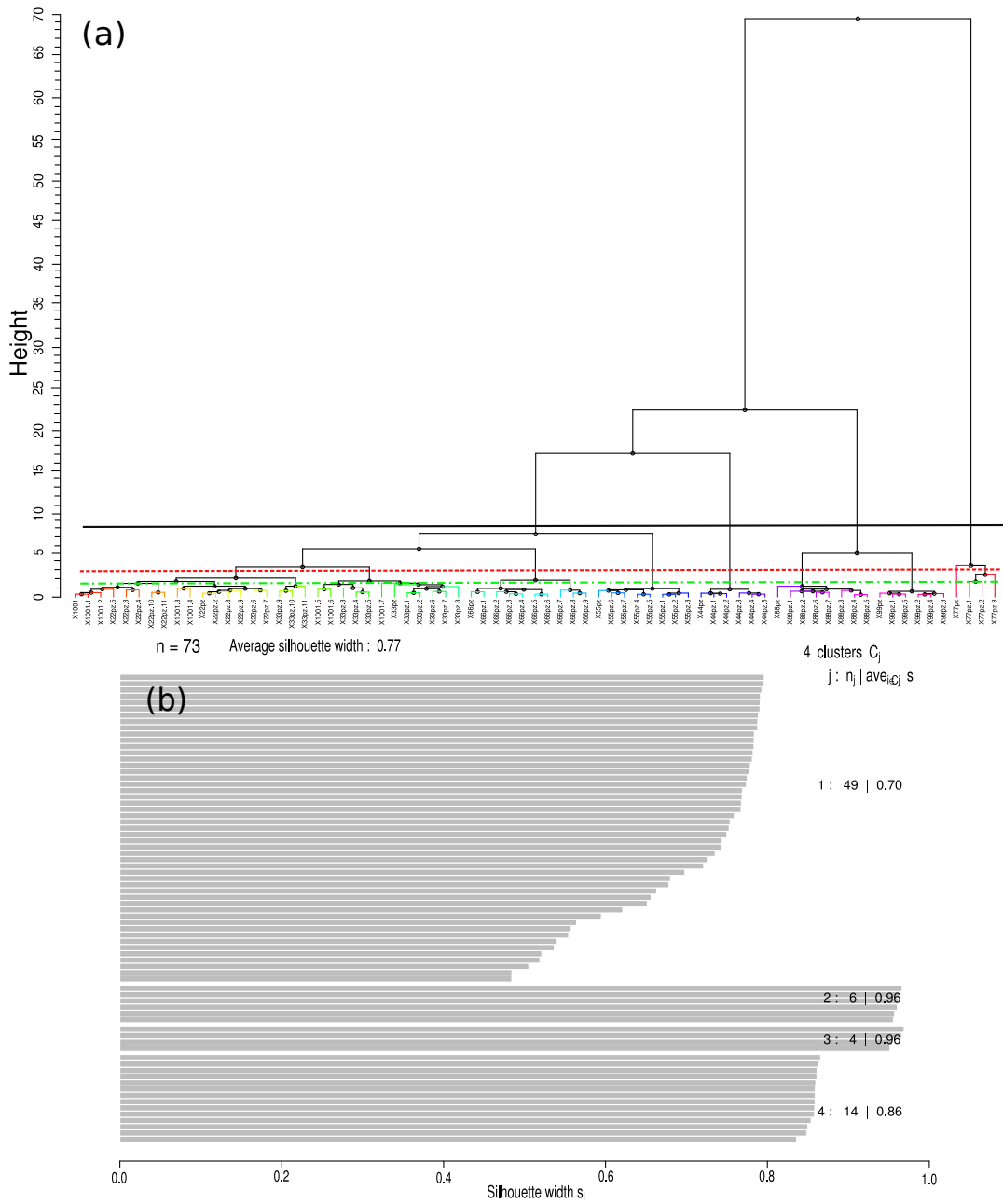


Figure 4.11: Hierarchical cluster analysis and silhouette graph of 73 *PC* spectral signatures of *eo1_pz* scene.

(a) Dendrogram obtained from 73 *PCs* spectral signatures in the *eo1_pz* scene. The black horizontal line indicates the number of clusters with the *SC* maximum. The dashed red line indicates the original number of land covers (9). The green dot-dashed line is an arbitrary cut height. (b) Silhouette graph for the $k=4$ with the best average silhouette width for the *eo1_pz* scene.

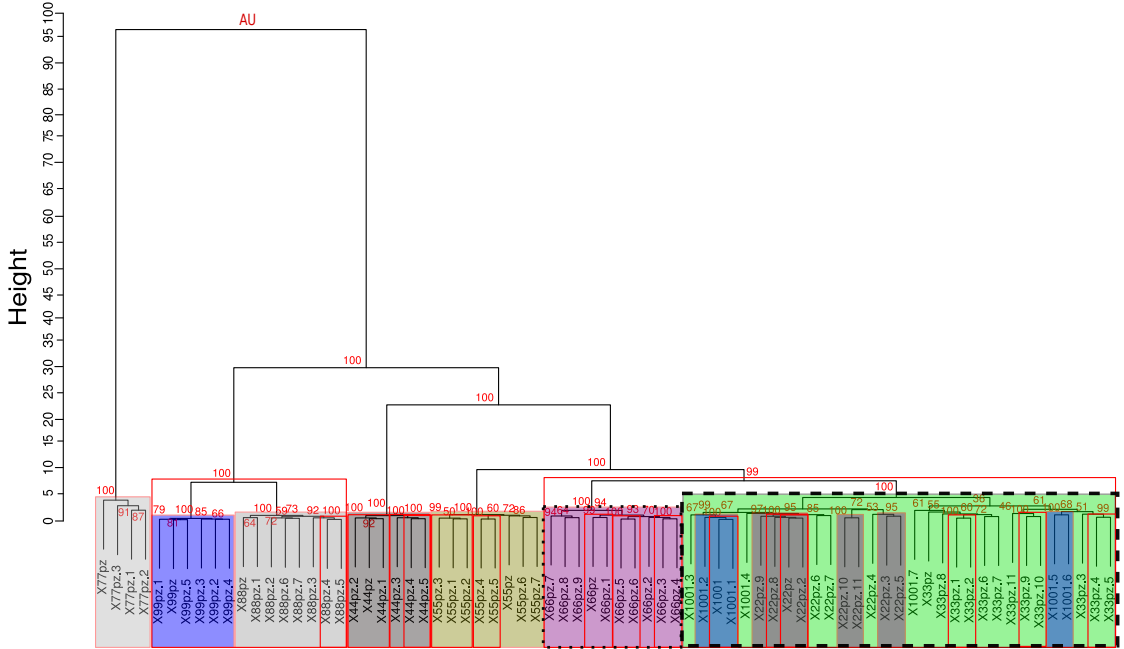


Figure 4.12: Approximately unbiased (AU) p -value with the *manhattan* distance and *average* as the cluster method for 73 *PCs* spectral signatures of *eo1_pz* scene.

4.1.5 The *eo1_kw* scene

The clustering process (first part) for the *PCs* spectral signatures of the *eo1_kw* scene is presented in Figure 4.13. This clustering behaves differently with respect to the number of types of *ROIs* selected for this image (a total of 9 land covers). The agglomerative coefficient and cophenetic index are 0.99 and 0.92 respectively (see Table 4.1). The degree of similarity is superior to that obtained for the *hy_15* and *hy_14* scenes, but also very similar to the *eo1_pz* scene.

Figure 4.13 shows the results from cluster analysis applied to the *PCs* spectral signatures of the scene *eo1_pz* spectra presented in Figure 4.4 with the same spectral range from 427 ηm to 2395 ηm as the *eo1_pz*. The horizontal line defines the seven best k clusters that could be cut, which in this case means a cut of the dendrogram into 7 clusters. If this cut height is considered as the best number of

Table 4.4: Statistics for the Approximately unbiased p -values and the SE for the PCs of the $eo1_pz$ scene data

Statistic	AU	SE.AU
Min.	0.3559	0.000000
1st Qu.	0.6747	0.0003279
Median	0.9249	0.0041088
Mean	0.8337	0.0047465
3rd Qu.	0.9990	0.0088480
Max.	1.0000	0.0116888

k clusters (based on the Figure 4.13 (a)), there are spectral confusions or it is not possible to separate between *G.arborea*, teak and forest. Mangrove and grass did not present spectral confusions at this reference height. As in other dendrograms, there is a good separability of soils, clouds, water, and shadows. The first cluster has 18 PCs spectra with an average silhouette width of 0.65. This cluster is very similar to the first cluster in Figure 4.9 (b) of the hy_14 dendrogram with *G.arborea* and palm oil in one single cluster. However, the decision key says that just one reasonable structure has been found for this target arrangement.

The minimum value of all AU for the four scenes corresponds to the PCs of the $eo1_pz$ (0.3559 in Tab. 4.4).

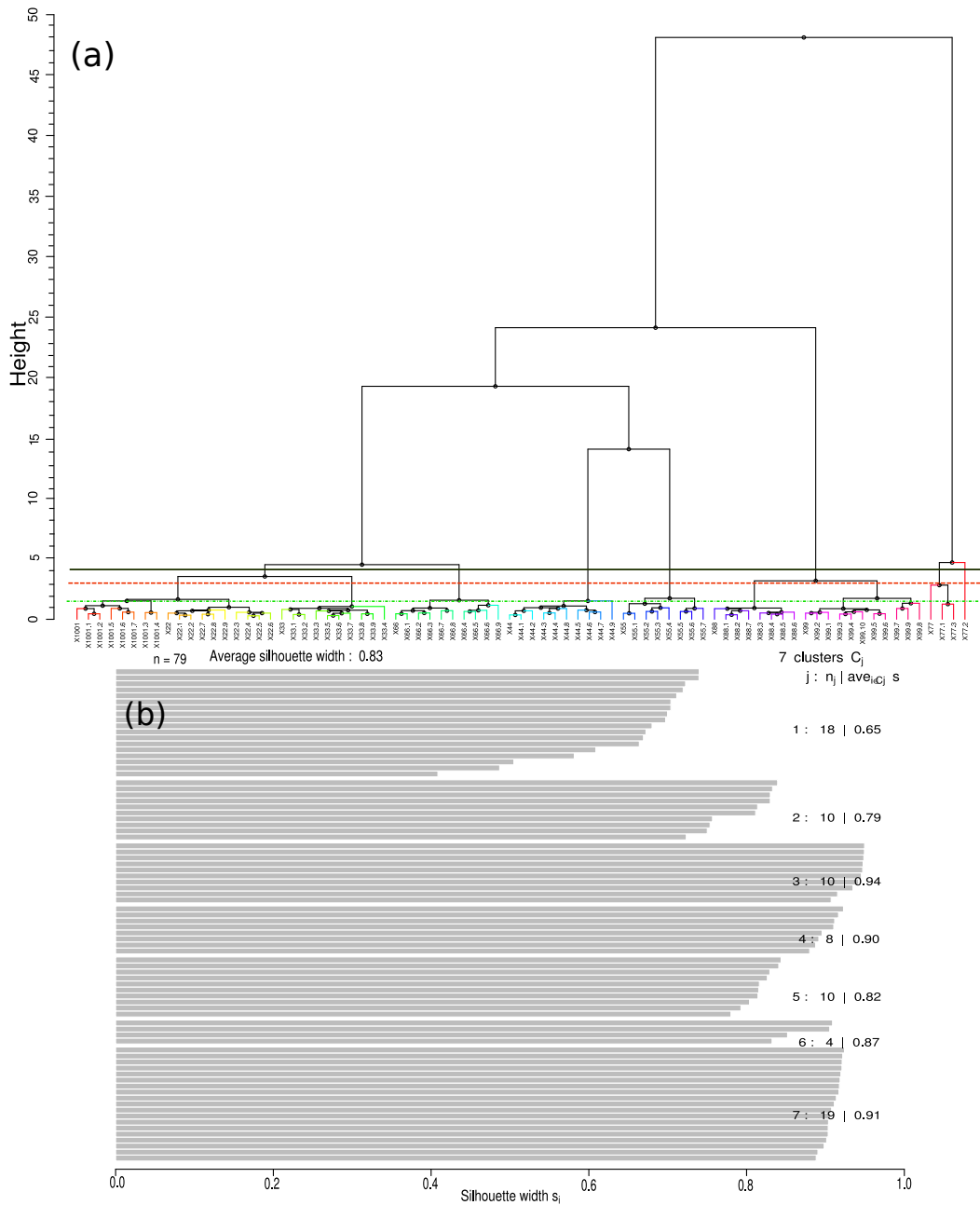


Figure 4.13: Hierarchical cluster analysis and silhouette graph of 79 PCs spectral signatures of $eo1_kw$ scene.

(a) Dendrogram obtained from 79 PCs spectral signatures in the $eo1_kw$ scene. The black horizontal line indicates the number of clusters with the SC maximum. The dashed red line indicates the original number of land covers (9). The green dot-dashed line is an arbitrary cut height. (b) Silhouette graph for the $k=7$ with the best average silhouette width for the $eo1_kw$ scene.

The perfect matches for the number of *ROIs* that were selected are in the uncertainty hierarchical cluster presented in the Figure 4.14. Only *G. arborea* presented two separate clusters, as is shown in the dashed box in this figure. This is one cluster with three *PCs* and another one with 5 *PCs*, for a total of 8. This indicates that all the clusters in boxes in this figure did not present spectral confusion in respect to the selected *ROIs*. However, the spectra of *G. arborea* are close from the palm oil spectra.

The minimum value of *AU* is 0.5164 (see Tab. 4.5) which is similar to the minimum of the *hy_15* and *hy_14* spectra. The median is the highest of all arrangements with *AU* values of 0.9849.

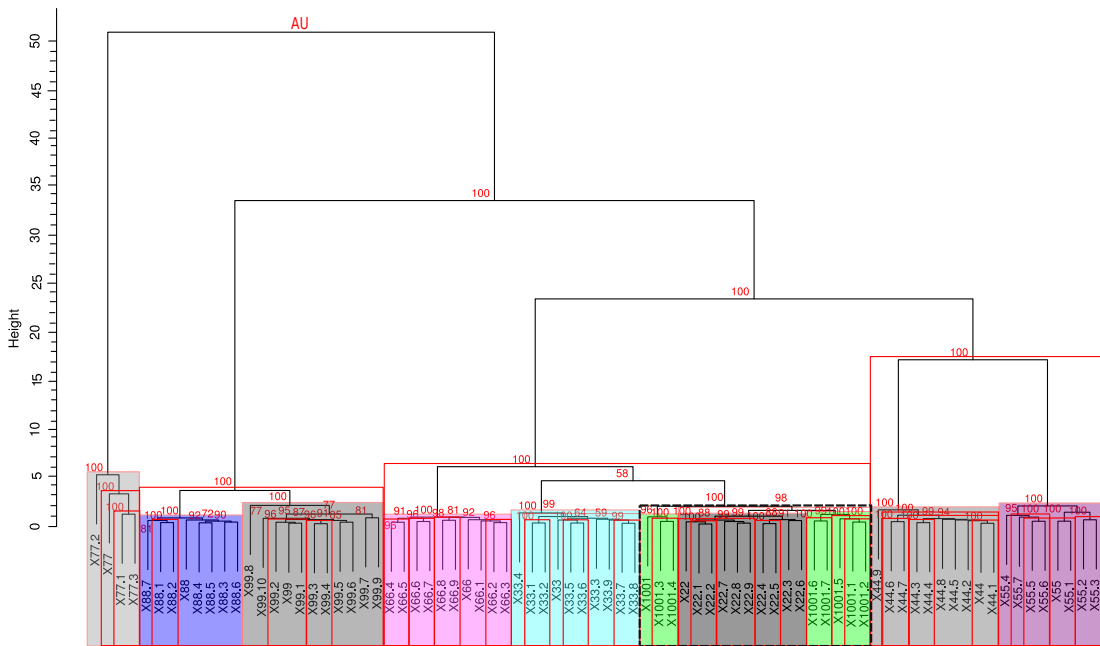


Figure 4.14: Approximately unbiased (AU)*p*-value with the *manhattan* distance and *average* as the cluster method for 79 *PCs* spectral signatures of the *eo1_kw* scene.

Table 4.5: Statistics for the approximately unbiased p -values and the SE for the PCs of the *eo1_kw* scene data

Statistic	AU	SE.AU
Min.	0.5164	0.0000000
1st Qu.	0.8733	0.0000084
Median	0.9849	0.0002636
Mean	0.9137	0.0003425
3rd Qu.	1.0000	0.0088480
Max.	1.0000	0.0010850

4.1.6 Specific clusters error evaluation

Figure 4.15 shows the number of clusters with the highest standard error for the four scenes. The highest value corresponds to the cluster 53 of the *hy_15* scene. This value is 0.03187 (see Tab.4.6) which is consistent with the inference that the true AU p -value is roughly estimated to exist in between $AU - 2 * SE$ and $AU + 2 * SE$. The confidence interval for this cluster is 0.92592 and 1.0534 or ≈ 1.0 . However, if the SE is too high, and considering that the AU p -value is defined to be between 0 and 1, this inference has no meaning.

The next interesting cluster is the number 24 in the scene *eo1_pz*, with an SE of 0.00909 (see Tab. 4.6). The confidence interval of where the true p -value may exist is between 0.9811 and ≈ 1.0 . As the α of 0.05 was defined, the existence of this cluster at this significance level is concluded.

The remaining clusters in all of the graphs in Figure 4.15 seems to have errors smaller than or around 0.01. Table 4.6 shows as an example some of these values which are marked in the Figure 4.15 with their respective *edge* number. The *edge* number is an identifier that permits the identification of every single cluster in one arrangement.

Table 4.6: Approximately unbiased p -values for the two highest standard error per scene.

Scene	Edge's number	AU	SE.AU
<i>hy_15</i>	24	0.60839	0.01001
	53	0.98966	0.03187
<i>hy_14</i>	52	0.57851	0.01036
	93	0.85807	0.00985
<i>eo1_pz</i>	63	0.35594	0.00999
	61	0.60811	0.01169
<i>eo1_kw</i>	44	0.54102	0.01085
	24	0.99928	0.00909

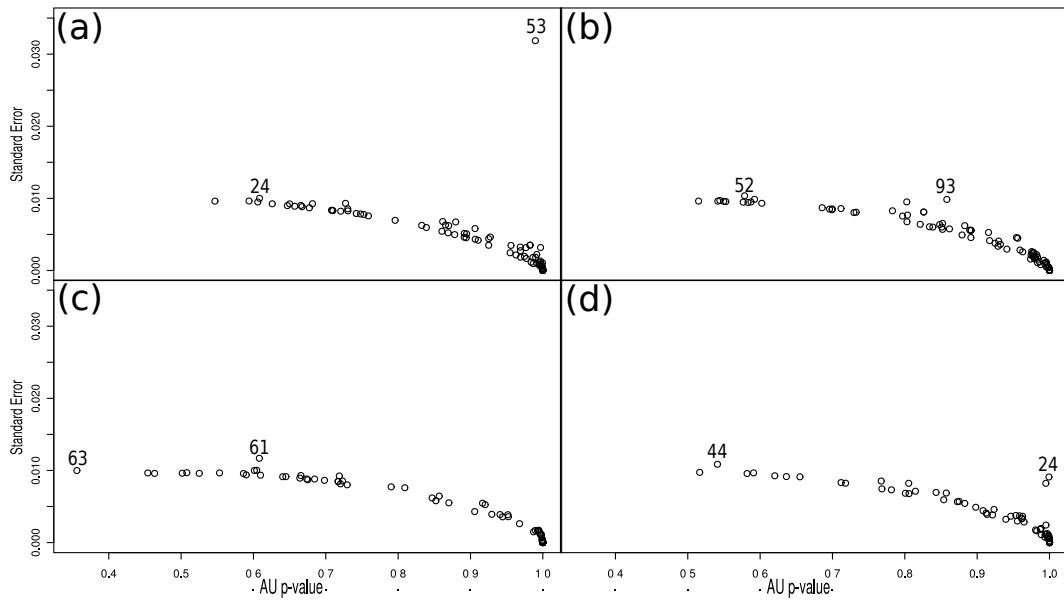


Figure 4.15: Approximately unbiased (AU) p -value and standard error for the four images

(a) *hy_15*, (b) *hy_14*, (c) *eo1_pz* and (d) *eo1_kw*

In Appendix C.1,C.2,C.3 and C.4, I presented the different arrangements of clusters solutions with their *edge* number. Thus, it is possible to identify each *edge* and the different AU and the respective SE .

4.2 USB650 Field Spectrometer

In this section the results for the *USB650* field spectrometer are described. These results have the same structure as those of the hyper-spectral scenes process.

In Figure 4.16 the mean and the standard deviation of the spectra of *G.arborea* for the ages of 6, 8 and 18 years in a range of wavelength from 366-1022 ηm is presented. In (a) the shaded area (standard deviation) is overlapping with the lines of the samples of the *G.arborea* of 6 and 18 years in practically the whole wavelength range. The least variability can be seen in the *SD* of the 6 years samples in the wavelength range of 500-700 ηm , but, in this section, there is an overlap with the 8 year *G.arborea* samples.

In the region of 900-1022 ηm there is no overlap between the *SD* of the mean *G.arborea* spectra of 6 and 18 years.

The best average silhouette width for the measured spectra with the *USB650* field spectrometer is 41, a value is possible with $k=14$. That differs with respect to the number of groups of spectra that were originally selected. The agglomerative coefficient and cophenetic index are 0.86 and 0.82 respectively. The agglomerative coefficient and the cophenetic index values are lower than all the clustering solutions of the *PCs* spectra of the hyper-spectral scenes.

Figure 4.17 (a) shows that after more than 12 clusters the within groups sum of squares are constant. The low value of the average silhouette width (0.41 for 14 clusters) indicates that the spectra measured with the *USB650* field spectrometer present high variability within the spectra of the different group of covers.

In Figure 4.18 the horizontal line indicates the cut height, where 14 k clusters could be cut. If this cut height is taken as the best number of k clusters, there are spectral confusions, or it is not possible to separate the *G.arborea* from the palm oil, teak, mangrove. The *edge* 49 (see App. C.5) forms a cluster with one spectrum of palm oil and one of *G.arborea*, indicating spectral confusion at this reference height. However, in general, there is good separability in the teak and *G.arborea* spectra.

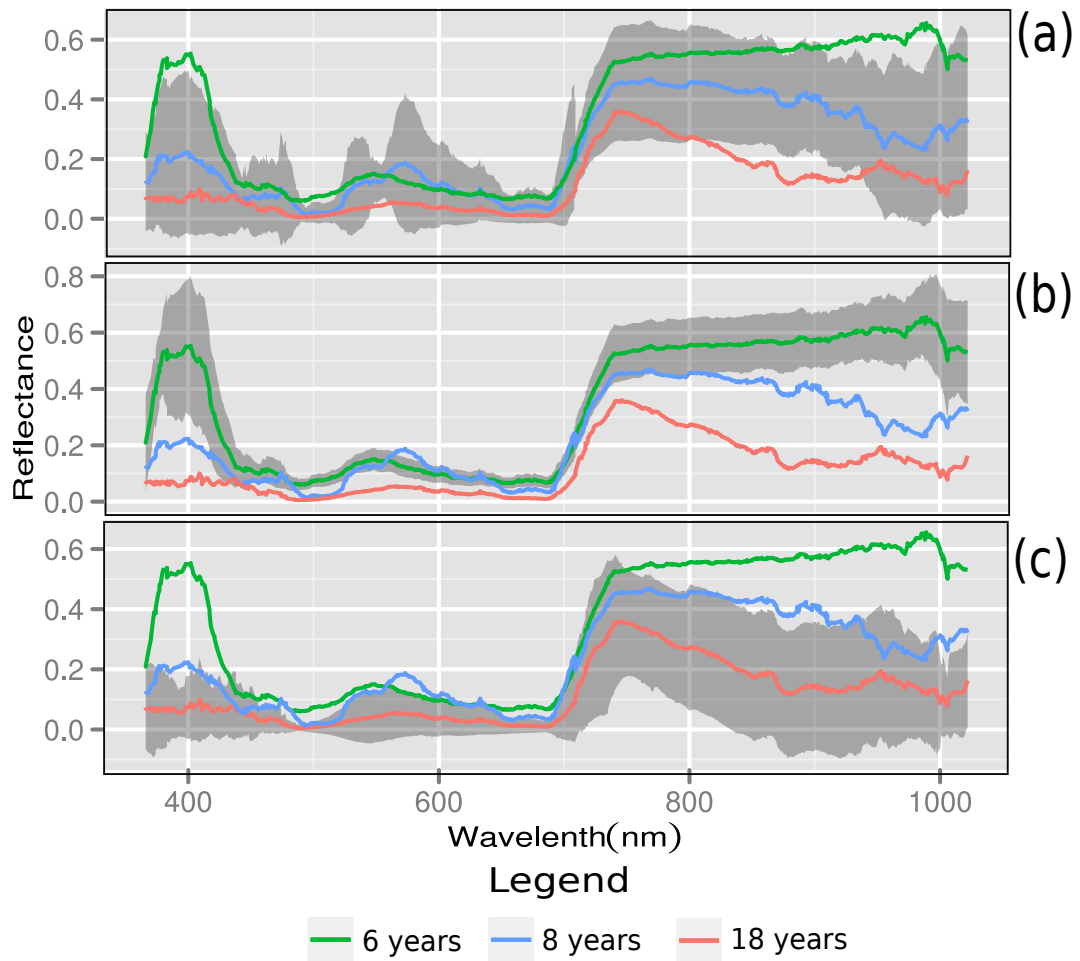


Figure 4.16: Mean \pm standard deviation (shaded area) of reflectance for leaves of three different *edges* of *G.arborea* plantations in the south of Costa Rica.

(a) Shaded area is for leaves in plantation of 8 years

(b) Shaded area is for leaves in plantation of 6 years

(c) Shaded area is for leaves in plantation of 18 years

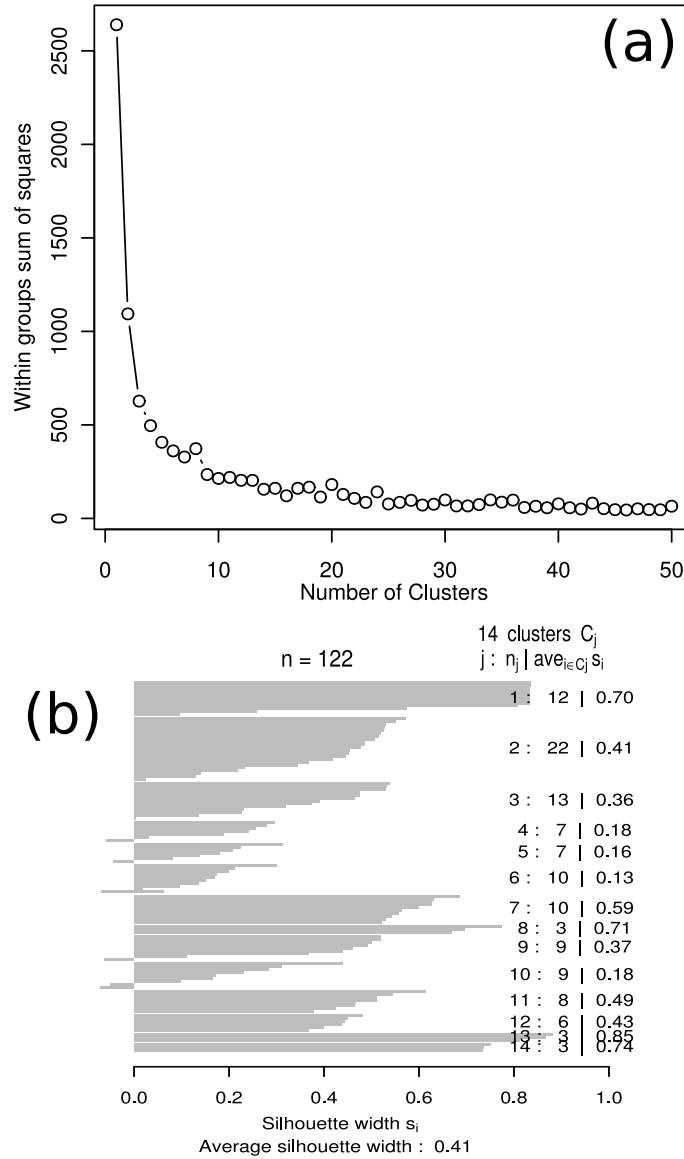


Figure 4.17: (a) Silhouette plot to identify possible groupings of the variables. (b) Silhouettes of a clustering with $k=14$ of 122 spectral signatures measured from the *USB650* field spectrometer

Figure 4.19 shows the relationship between the different *G.arborea* ages and the other reference spectra according to the hierarchical cluster analysis with the *AU* based method. The clusters which have more than 95% of *AU* and more than three spectra are marked in shaded boxes. In these shaded clusters there are no

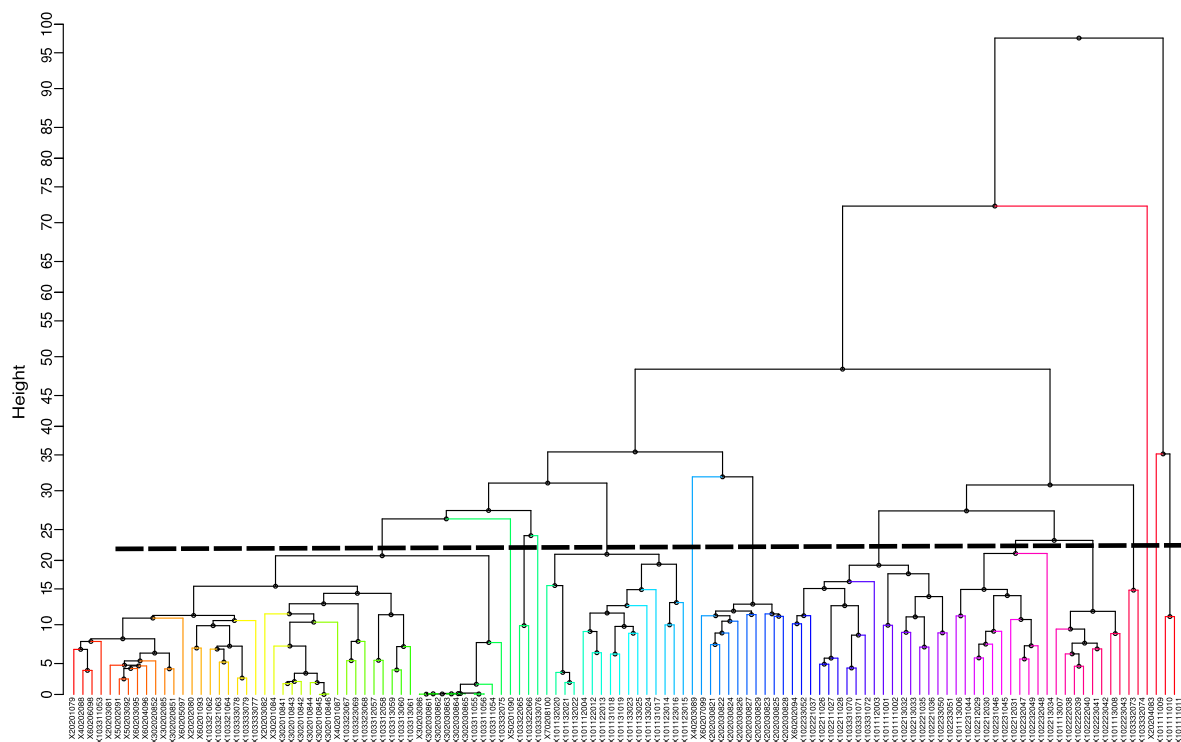


Figure 4.18: Dendrogram with $k=14$ cut height of 122 spectral signatures measured from the *USB650* field spectrometer

spectral confusions with *G.arborea* and spectra of other covers. However, there is a cluster of two spectra that mix *G.arborea* and palm oil.

The most important spectral confusion is between 4 spectra of 18 year old *G.arborea* and 6 of mangrove. Just one spectra of teak is not clustered with similar teak spectra, and one cluster is formed by palm oil and a 8 year *G.arborea* spectrum.

The mean of the *AU p-values* of the computed multi-scale bootstrap resampling of the *USB650* is 0.8420 and its median is 0.9271. The mean *SE* is 0.004965 and the median of the *SE* is 0.005854.

Figure 4.20 shows the number of clusters with the highest standard error for the *USB650* field spectrometer. The highest values (in both axes) correspond to the cluster 89 followed by the number 84. For *edge* 89 this value is 0.01526 (see Tab.4.8). Following the same rule applied to the *PCs* scenes spectra ($AU - 2 * SE$

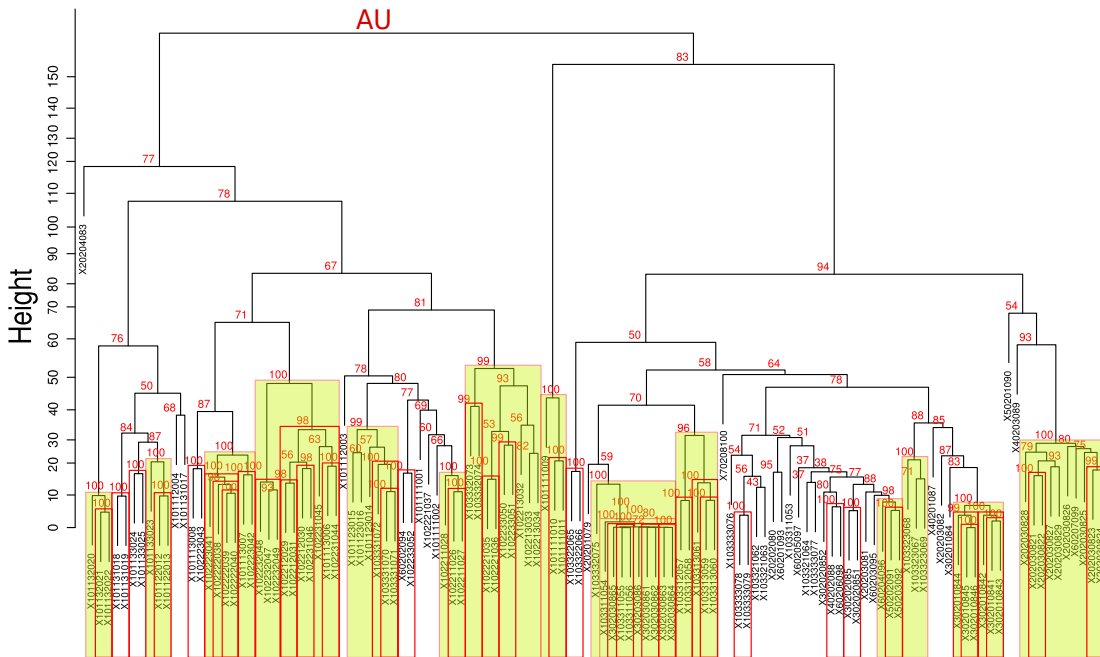


Figure 4.19: Approximately unbiased (AU) p -value with the *manhattan* distance and *average* as the cluster method for 122 spectral signatures of *USB650* field spectrometer.

and $AU + 2 * SE$), the confidence interval for this *edge* is between 0.95439 and 1.01543 or ≈ 1.0 . Thus, the existence of these clusters at the level of significance is concluded.

The next *edge* is number 84 in the scene *eo1_pz*, with an *SE* of 0.00986 (see Tab. 4.8). The confidence interval where the true p -value may exist is between 0.93599 and 0.97543. As the α of 0.05 was defined, this cluster cannot exist at this significance level.

The rest of the clusters in all of the graphs in Figure 4.20 seems to have errors smaller or around 0.01 as well as too low *AU* values. As an example, table 4.8 shows one value which is marked in Figure 4.20 with its respective *edge* number.

Table 4.7: Statistics for the approximately unbiased p -values and the SE for the *USB650* field spectrometer data

Statistic	AU	SE.AU
Min.	0.3708	0.000000
1st Qu.	0.7197	0.000000
Median	0.9271	0.005854
Mean	0.8420	0.004965
3rd Qu.	1.0000	1.00000
Max.	1.0000	0.015264

Table 4.8: Approximately unbiased p -values for the three *edges* with the highest standard error in the *USB650* field spectrometer data

Sensor	Edge's number	AU	SE.AU
<i>USB650</i>	89	0.98491	0.01526
	84	0.95571	0.00986
	95	0.86593	0.00792

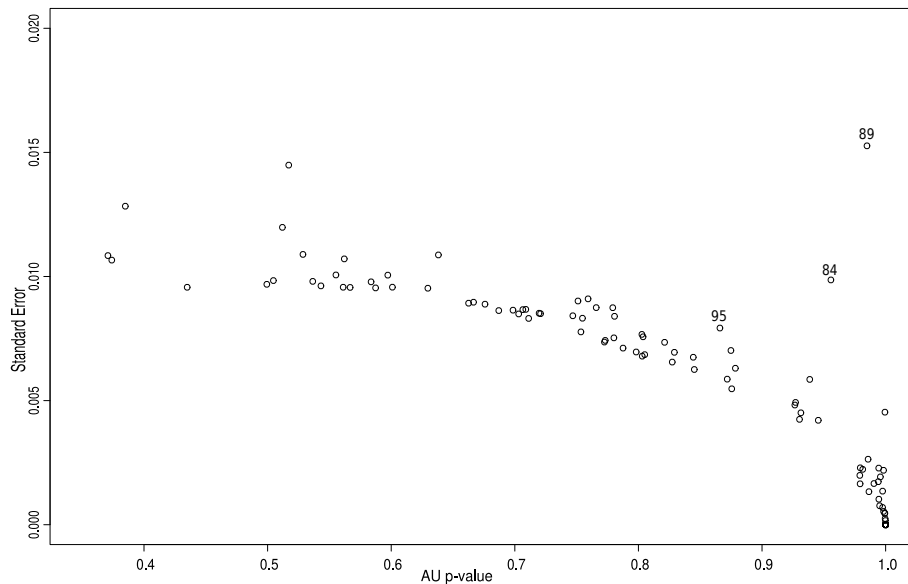


Figure 4.20: Approximately unbiased (AU) p -value and the standard error for the *USB650* spectral signatures

4.3 TerraSAR-X Processing and Land Cover Analysis

After a multi-looking process, a more pleasing appearance of the segment of TerraSAR-X is presented (see Figure 4.21). This process makes the features more interpretable, however, the cost is in the spatial resolution. For example, the *HH* co-registered images pass from 1.93 mean *GR* square pixel from 1 range look and 1 azimuth look to 9.31 mean *GR* square pixel with 4 range looks and 6 azimuth looks. With this process the images are degraded in order to enhance features which are easier to be classified as well as interpreted. The mean *GR* square pixel is the average of the range and azimuth pixel spacings in the multi-looked image. It was computed based on the number of range looks, the number of azimuth looks and the source image pixel spacings [ARRAY-ESA, 2011].

The mean and the *SD* of different *ROIs* in the *HH* example is shown in Appendix E.

Table 4.9 is a selection of bands which give some important information in order to analyse the separability of the covers in the study area as well as for the unsupervised classification. Not all the calculated bands are listed here because some bands did not, for example, add more information. The dates of the *HH* and *VV* polarized images are the same as it is shown in Table 3.8.

The images of *VV* polarization with high incidence angle $\approx 23^\circ$ exhibit substantial problems such as foreshortening. For that reason, they were excluded in the calculation of a mean of this polarization (as in B1).

In Figure 4.22, the *x*-axis describes the vector with the *LDA* was created, which encompasses all of the variables that were used and assigns importance to each of them. The *y*-axis represents the proportion of plots along each vector. Here a trend is shown, and it seems to separate the (*ROI*) forest (in flat terrain with dominance of raffia) into different bands.

All conditional density plots except Figure 4.24 are presented in Appendix D. These are Figures from D.1 to D.10. These plots represent the distributions

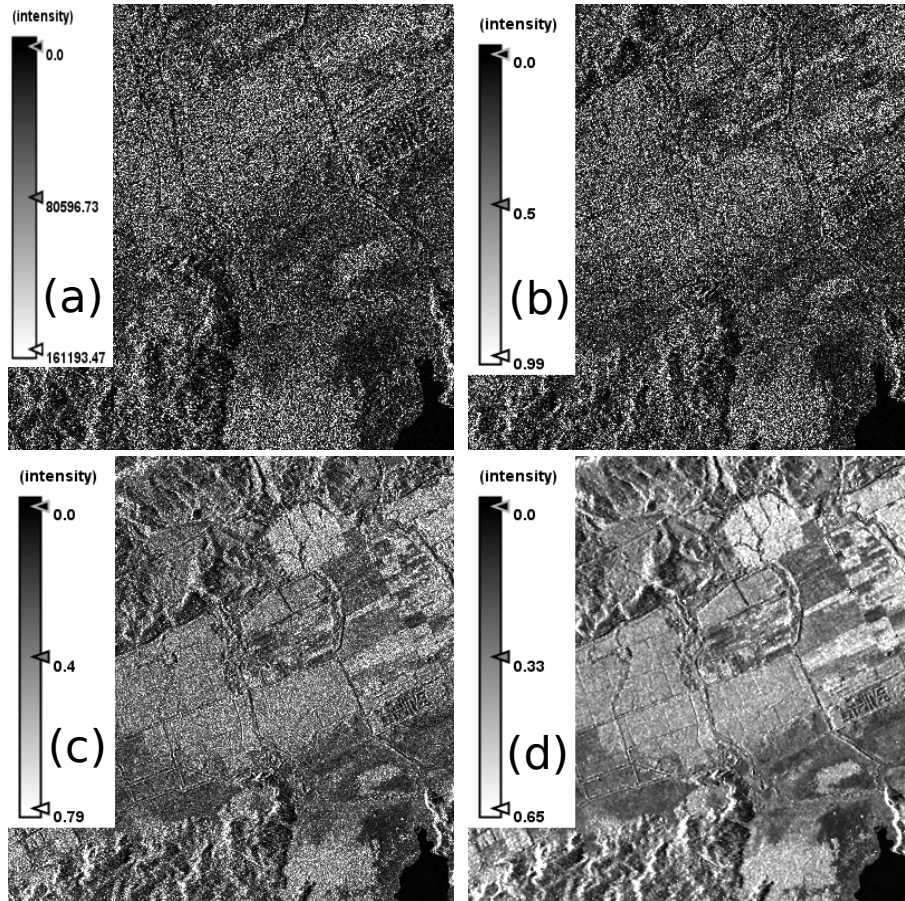


Figure 4.21: Segment of the TerraSAR-X *HH* process in the south of Costa Rica:
 (a) intensity raw image
 (b) calibrated image
 (c) calibrated and multi-looked image
 (d) calibrated, multi-looked, filtered and speckled image

of the different land covers (*ROIs*) (see Tab. 3.10 for the description) that were selected. With these plots it is possible to superimpose the kernel density plots of two or more land covers per band. Specifically, the posterior discussion is focused on the land covers of forest in flat terrain (*raffia*), forest in undulating terrain and rolling slopes, as well as the confusion with foreshortened and shadowing areas and palm oil (old and young plantations) land covers.

In Appendix E, the mean and the standard deviation of the different land covers per band is presented. The mean varies from band to band. For example, for the

Table 4.9: TerraSAR-X selected bands.

Band	Description
B1	Mean of the tree <i>HH</i> processed images
B2	<i>HH</i> polarization (2010.04.10)
B3	<i>HH</i> polarization (2010.05.01)
B4	<i>HH</i> polarization (2010.05.02)
B5	<i>VV</i> polarization (2010-06-02)
B6	<i>VV</i> polarization (2010-05-03)
B7	Mean of <i>HH</i> and <i>VV</i> polarization images
B8	Mean <i>HH</i> polarization <i>PCs</i>
B9	Mean <i>VV</i> polarization <i>PCs</i>

land cover *forest* in flat terrain with raffia's dominance, the bands B3, B4, B5 and B6 have a similar mean (81.63, 81.75, 71.29, and 75.40 respectively) as opposed to the bands B2, B7, B8 and B9 (123.79, 119.95, 130.20 and 117.90 respectively) (see Appendix E for the values). In this land cover, band B1 presents a multi-modal distribution with a high value of *SD*. Here it is difficult to interpret the mean value of 18.42, because of the influence of the skewness. As is shown in Figure 4.24, all the bands except the band B1 have a normal distribution which makes it possible to compare the means.

For the land cover *forest* in gently undulating and undulating terrains, again B1 exhibits a multi-modal distribution (see Fig. D.1), which makes it difficult to interpret the mean of 49.23. All the other bands present a right-skewed distribution which covers the whole range of the 8-bit intensity values. As the degrees of freedom determine the shape of the distribution, with an increase of n this distribution becomes more and more symmetric [Reimann et al., 2008]. However, the *outliers* are still there. This is an inherent property of the foreshortened and shadowing areas in the TerraSAR-X imagery. All the objects oriented to the antenna beam (fore-slope) will present high values in intensity, on the other hand the back-slope will present darker areas with low intensity values. In the unsupervised classification performed, the foreshortened areas account for 10.9% of the study area.

An extreme case of foreshortening is called image layover [Jensen, 2007]. The

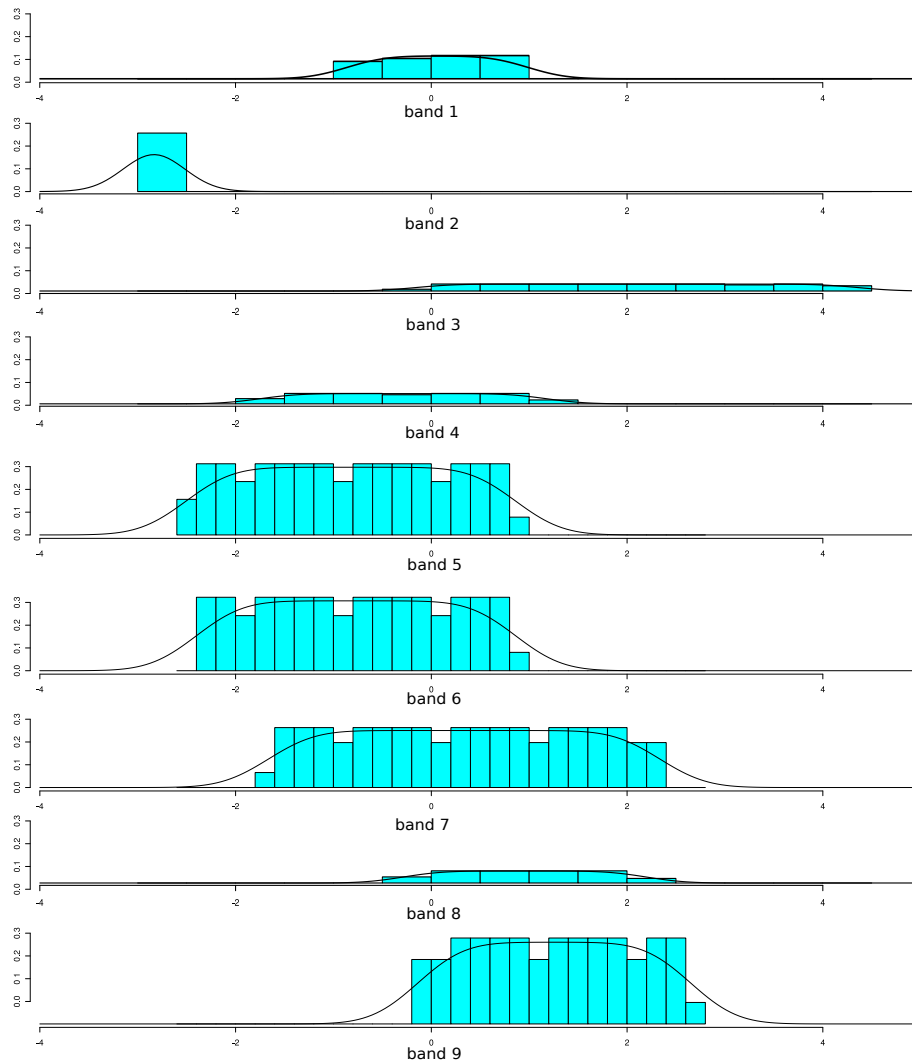


Figure 4.22: Linear density graphs for the *ROIs* of forest in undulation and rolling slopes for the 9 bands

result is presented in Figure D.1. The land cover *forest* in gently undulating and undulating terrains does not exceed the 12.3° of slope. However, there are many extreme values around 255 of the 8-bit intensity bands. The band that does not present the right-skewed distribution is band B1, however, it presents two peaks and some very low 8-bit intensity values and a *SD* of 26.21.

Figure D.2 shows the extreme cases of layover, again the band B1 does not present the right-skewed distribution but has two peaks in it, also with many values with

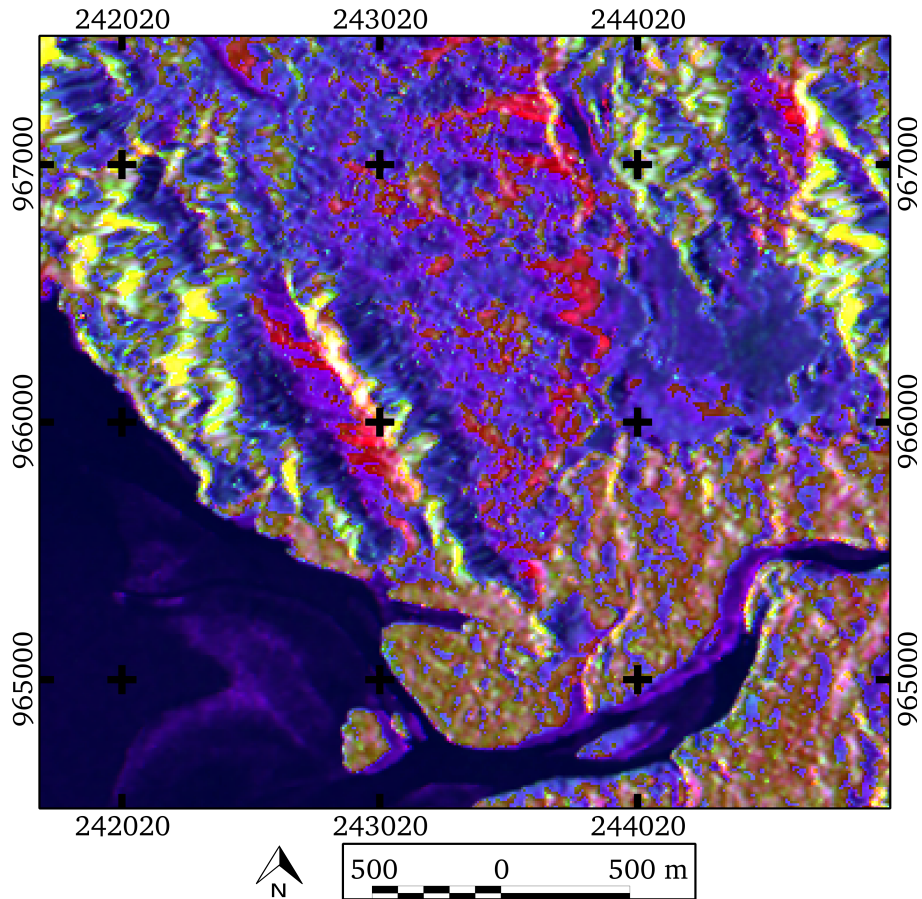


Figure 4.23: Example of TerraSAR-X Red (band B8), Green (band B5) and Blue (band B1) composite in a segment of the study area in the south of Costa Rica. (The coordinates UTM Zone 17 north, datum WGS84)

low intensity. This band also presents the low SD with 25.85 but its mean of 41.48 is difficult to interpret because of the shape of the distribution. The $ROIs$ were selected for this land cover in the slope category of *rolling* in the Table 1.1 which has a range between 12.3° and 17.6° . Especially the bands B2, B7 and B8 are very sensitive to the layover presenting the high values of SD (80.53, 50.08 and 62.11 respectively).

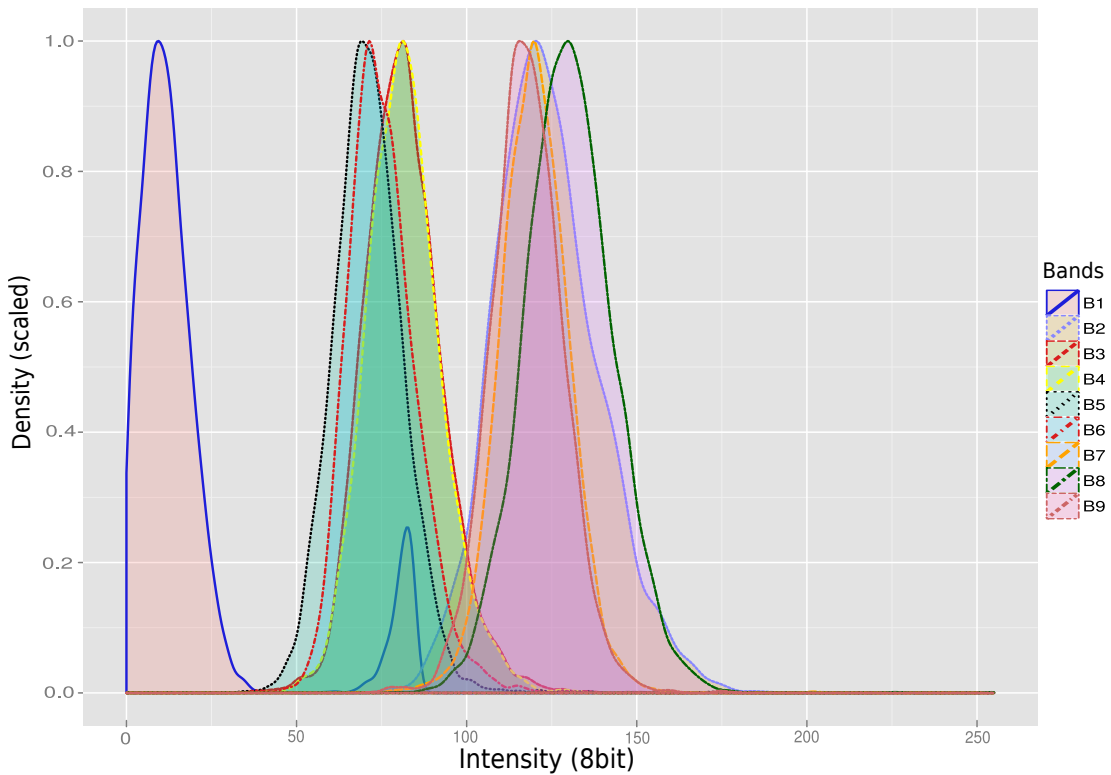


Figure 4.24: Conditional density plot for all bands for the land cover *forest* in flat terrain (*raffia*).

For the land cover *herbland*, all bands present a multi-modal distribution with more than one peak. In this case, all measures of central tendency will be located in different places (on x-axis of density of the plot, for example). Again, it is difficult to interpret the means. However, the *SD* values are lower than the *SD* values of the land cover *forest* in flat terrains (*raffia*).

A different behaviour from the rest of the land covers is shown in Figure D.4. Normally the man-made structures act as corner reflectors that send much of the incidence energy back toward the antenna [Jensen, 2007]. These are the highest mean values in the bands B2, B7 and B8 (252.94, 226.30, 244.30 respectively). However, there are very low values for the bands B1, B3, B4, B5 and B6 (41.55, 78.58, 73.39, 54.64 and 72.27 respectively).

Except for the band B1, the *mangrove* land cover (see Fig. D.5) presents a symmetrical distribution, but has heavy tails with some extreme values in the hold range of the 8-bit quantization. The bands B7, B8 and B9 have the highest values of intensity with 134.35, 132.24 and 143.16, respectively. There is no big difference in the *SD* values, the highest value is for band B5 with 31.11 and the lowest value is for band B2 with 20.79.

Similar to the distributions of the land cover *forest* in flat terrain (*raffia*), Figure D.6 presents a normal distributions in all (except B1) bands, however, they are clearly divided into two main groups. First, the bands B3, B4, B5 and B6 with intensity values of 93.18, 83.25, 73.83 and 77.13, respectively. The second group is formed by the bands B2, B7, B8 and B9, with values of 141.69, 132.80, 146.96 and 124.34, respectively. These groups are the same as the ones in Figure 4.24. In all the cases, the means for all the bands and the standard deviation are a little higher in the *palm oil* (old plantations), but keeping the same proportions as *forest* in flat terrain (*raffia*). In Figure D.7 the land cover *palm oil* (young plantation) is similar in Figures D.6 and 4.24.

For the land cover *pasture*, a normal distribution for all bands is shown, but with some small peaks. Contrary to previous Figures, band B1 has a similar distribution to all the rest of bands and also a similar mean (45.97) and *SD*.

All bands in Figure D.9, have a multi-modal like distribution (except B2) and together with the *forest* in gently undulating and undulating terrains and *forest* in rolling terrain, they present the highest values of *SD*, except for the band B2 with 10.56.

Figure 4.25 is a subset of three different bands of the process of TerraSAR-X imagery. Here, the *PDR* index (a) is presented. This result was excluded from the land cover analysis but was included in the unsupervised classification.

The *PDR* band is highly sensitive to the moisture content of the soils. The different surfaces or materials conduct the electricity better than others. One measure of the material's electrical characteristics is the complex dielectric constant which is defined as a measure of the ability of a material to conduct electrical energy Jensen [2007]. This property is important to understand the *SAR* backscatter.

In general, moist soils reflect more energy than dry soils [Jensen \[2007\]](#). However, Figure 4.25(b), is more useful than (a), if the interest is in soil moisture but in (a) there is doubt about the dissimilarities between the bare soils and vegetated soils.

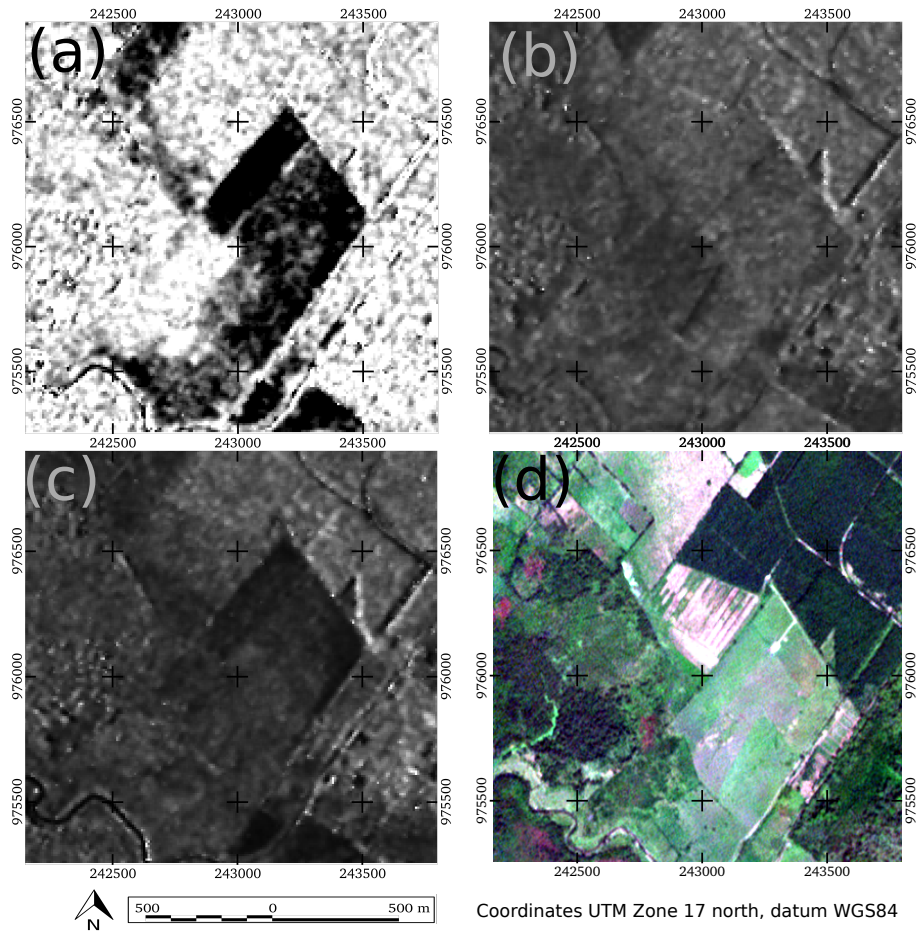


Figure 4.25: Example of TerraSAR-X penetration in soils in a segment of the study area in the south of Costa Rica:

- (a) *PDR* image (see Eq. 3.10)
- (b) *HH* polarization (2010.05.01)
- (c) *VV* polarization (2010.05.02)
- (d) True colour composite of RapidEye optical imagery

The land cover *water* presents the lowest *SD* values, band B3 has a value of 3.92, the remaining bands are around 2 or less. Band B9 has the highest mean value

with 53.08, but, as other bands, also (see Fig.D.10) presents several peaks in the distribution (multi-modal).

In Figures F.1 to F.9, the conditional density plots for all land covers per band are presented.

For band B1 (see Fig.F.1) the land covers that present the lower variability are; *herbland*, *pasture* and *water* (*SD* of 9.39, 9.13 and 1.37 respectively), the rest of the land covers have an *SD* > 17. However, the land cover *palm* which has a *SD* of 13.13 is an exception. It is not possible to discriminate the land covers *forest* in rolling slopes and *mangrove*, *forest* in gently undulating and undulating terrain from the others. In this Figure, all bands presented a multi-modal distributions and covers a range from 0 to 95 values of intensity. Except for the *forest* in rolling terrains which has values up to 255 of intensity.

In Figure F.2 the distributions are normal except for the land cover of *forest* in rolling slopes and the *forest* in gently undulating and undulating terrain which is right-skewed. For this band the lowest variabilities are for *water*, *pasture*, *soil* and *infrastructure* (1.46, 7.62, 10.56 and 8.77). The *infrastructure* is leaf-skewed with two small peaks.

Band B3 shows another behaviour from band B2. For example, *infrastructure* is right-skewed (see Fig. F.3) and its mean is 78.58 values of intensity (*SD* of 55.43). In contrast, the mean value of intensity in the band B2 is 252.94. In this band, there is no separability between the land covers of *soil*, *mangrove*, *palm* (old and young plantations) and *forest* (flat terrain with raffia). The *forest* (rolling slopes) covers the whole range of intensity values but with a right-skewed shape.

For band B4 the shape of the density plot of *forest* (flat terrain with raffia) and *palm* (old plantation) are similar, with mean values of 81.75 and 83.25 respectively. *Infrastructure* presents a right-skewed curve with a very low mean of intensity (73.39) and very high *SD* (42.88).

In Figure F.5, land covers *herbland* and *pasture* present a similar shape with means values of intensity of 33.99 and 41.51 respectively. However, the *pasture* is right-skewed at the lower part of the density. Here, the land covers of *soil*,

mangrove, *palm* (old and young plantations) and *forest* (flat terrain with raffia) present better separability than band B4.

In band B6 (see Figure F.6) the *palm* (young plantation) is slightly detached from the *palm* (old plantation) and *forest* (flat terrain with raffia). These land covers have a normal distribution with mean values of 92.69, 77.13 and 75.40 respectively. The standard deviations are 12.97, 12.83 and 10.25 respectively which are the small ones, if the land cover *water* is excluded.

Contrary to band B6, the *palm* (old plantation) is slightly separated from the *forest* (flat terrain with raffia) and *palm* (young plantation) in band B7 (see Figure F.7). These three land covers present a normal distribution with close means for *forest* (flat terrain with raffia) and *palm* (young plantation) (119.95 and 123.20 respectively). The *infrastructure* presents a leaf-skewed distribution trough out the range of the distribution. Band B8 is similar to band B7 (see Fig.F.8).

In band B9 the shape of the curve of the land covers of *herbland* and *pasture* are very similar, except that *pasture* is right-skewed and *herbland* has a small peak. In this band the *water* has the highest mean value and an *SD* with 53.08 and 2.13 respectively.

Figure 4.26 shows the unsupervised classification of all bands presented in Table 4.9 plus the *PDR* band.

Results (see Table 4.11) from the *isodata* classification indicate an acceptable overall accuracy of 86 percent. Visual inspection of the cross table confirms that 12 of the misclassification are attributed to the foreshortening areas.

With this exploratory classification the producer's accuracy of forest is (4752) 90.4 percent while the user accuracy is (4762) 75.8 percent. It is expect that roughly 76% of all pixels classified as forest are indeed forest on the reference data. However, as a producer 90.4% of all the forest pixels as such were classified.

There are pixel confusions, especially with two blocks of palm oil plantations and mangroves. This problems is shown in Figure 4.26 in the class palm oil with the orange colour.

Table 4.10: Summary area and percentage per class of the unsupervised classification of the TerraSAR-X imagery in the study area

Land cover	Area (ha)	Percentage
water	2006.56	3.8
soil-grass	5591.75	10.7
grass-soil	4300.50	8.2
herblands	7894.28	15.1
forest	14367.05	27.4
forest-raffia	5803.18	11.1
palm-oil-raffia	4372.87	8.4
palm-oil	2335.42	4.5
foreshortening	5687.55	10.9
Total	52359.18	100.0

Table 4.11: Error matrix of forest/non-forest classes for the classification of TerraSAR-X imagery in south of Costa Rica.

		Reference truth		Total (classified)	User's accuracy
		forest	no forest		
Classified map	forest	47	15	62	75.8%
	no forest	5	76	81	93.8%
Total (ground truth)		52	91	143	Overall accuracy
Producer's accuracy		90.4%	83.5%		86.0%

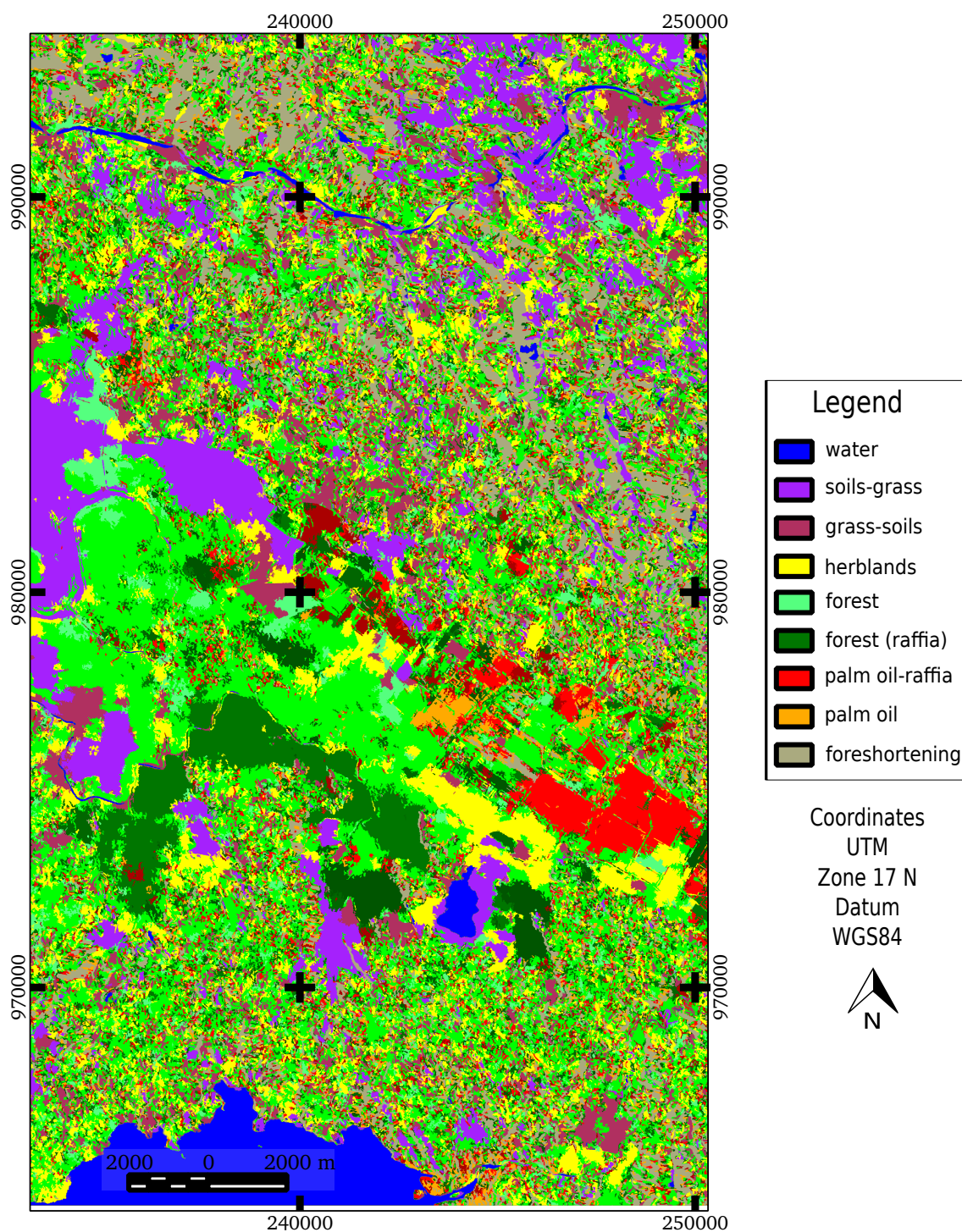


Figure 4.26: Unsupervised Classification of the TerraSAR-X imagery in the study area.

Chapter 5

Discussion

The discussion of the hyper-spectral and TerraSAR-X processing begins in Chapter 5. In order to analyse the clustering results obtained, the hyper-spectral results are subdivided in the two sections. First, the analysis of the hyper-spectral scenes is done by comparing the scenes (within sensors) based on seasonality and atmospheric conditions. Second, the field spectrometer spectra at the leaf level is an analysis of the variability within *G. arborea* to illustrate the spectral confusion of the scene based spectra.

Finally, on the one hand, the analysis of the TerraSAR-X processing and forest/non-forest classification is presented. On the other hand, a General discussion about remote sensing and seasonality of vegetation based on the precipitation, potential evapo-transpiration and normalized difference vegetation index is presented.

5.1 Hyper-Spectral Sensors and General Considerations on the Clustering Process

Regarding the hyper-spectral pre-processing, it is not possible to automatically perform elementary operations on the imagery using the *FLAASH* module for the atmospheric correction. Therefore, the necessary manual pre-processing is

a cumbersome task (calibration, BIL-to-BIP conversion, scaling, writing wavelengths to header). This also means that batch processing is not possible, making the atmospheric correction of several images at time practically impossible.

A special consideration is needed for the *ROI* selection. As a method based on visual inspection it has several severe disadvantages. First, the selection or the availability of the reference high spatial resolution photographs is constrain and these are not always from the same month as the imagery. The *ROI* selection is also inherently subjective as it relies heavily on the researcher's visual capabilities and experience at the time when inspecting the photographs. Secondly, the features determining different covers have different conditions depending on the quality of the photographs. Finally, visual inspection and selection is limited by the variation of the sensitivity of the human eye over the visible spectral range, and by previous and ancillary information available. Nevertheless, visual inspection and *ROI* selection are widely used. These are, however, not exempt from human errors in terms of interpreting the key correctly.

In addition to the previous considerations, all the derived *ROIs'* spectra are in the category of *flat* and undulating slopes (see the ranges of Table 1.1) which could present topographic effects. Despite this, the topographic normalization was not applied to the four scenes.

The first approach of *HCA* results shows that there are several spectral confusions between the vegetation land covers in all the *PC* spectra of the four scenes. But the degree of confusion depends on the cut height of the dendrogram. There are some obvious variations in the reflectance patterns of the four scenes. These variations, for example, can come in the shape of shaded areas due to clouds which have a lower reflectance. However, these shaded areas can also appear as a result of lower incidence angles as well as the topographic effects.

It is important to know if a partition obtained by a clustering algorithm is good or not. It is possible to compare the partition with other partitions obtained by other clustering algorithms. The better the agreement is among these partitions, the more condensed the obtained partition. Lack of agreement among these partitions indicates either that the cluster structure is difficult to detect or that

some clustering algorithms do not perform properly for the specific data set. An example of this situation is the difference in the partitions of the *eo1_kw* and the *USB650* field spectrometer with values of 0.92 and 0.82 for the cophenetic index respectively. In this situation, a strength index that measures the separation among clusters of the data set is useful to indicate which partition is better for the specific data set. The larger the minimum separation a partition has, the better the partition is.

If in the clustering evaluation phase, the final arrangement of a group of clusters produces very different spectra, it can be suspected that there are genuinely no clusters in some of the spectra, or that the chosen distance measure was not the best. The *manhattan* distance was the best one tested for all the hyper-spectral data including the data from the *USB6650* field spectrometer.

The cophenetic index does not require a priori selection of an optimal number of clusters and provides a general view of the partition obtained. Thus, this index facilitates comparison of the results for the different clustering solutions.

The overall average silhouette width of the entire plot is the average of all objects in the data frame. The suggestion given by [Rousseeuw \[1987\]](#) of choosing the number of clusters when the *SC* is as large as possible was followed. However, in the case of the arrangements of *eo1_pz* and *hy_14* this criterion was not applied. The highest *SC* for these two arrangements is a *k* of 2. This is not the optimal case because the number of coverages for the images is 9.

The basic indicator for a well fitting cluster is that the variance within the cluster is small while variance between clusters is large.

In the first clustering approach, the height at which the dendrogram is cut defines the objects that are classified. For example, in [Figure 4.13 \(a\)](#), there are two clusters defined for the cover of clouds. There is a significant difference in the heights that define these two clusters indicating that the types of clouds are very heterogeneous.

For all the clustering solutions the initial number of land covers was known. This means that if the *SD* value did not correspond with this number, the cluster

solution showed the spectral confusion for some of the clusters. For example, the black line in Figure 4.4 is the cut height for the highest SD , but the original number of land covers in this case was 9. The dashed red line represents these 9 clusters. However, this should be seen with caution because of the within $ROIs$ variability. In other words, it is possible to capture two ages of one land cover in one ROI .

As an example of the idea described above, the green dot-dashed line in the same figure is an arbitrary cut height which gives the separability between all selected land covers. However, there are more than 9 clusters at this cut height (15 clusters), which would be an indicator of the variability of the spectra. For example, the clouds (X77) at this reference height are divided in three clusters from a total of four PCs spectra. If there is no previous knowledge about the possible number of clusters, the average silhouette width is a good indicator of the possible number of clusters.

In the case of the *eo1_kw* scene, it is possible to classify all 9 land covers using PCA as a data reduction technique. However, for the scene *eo1_pz* it is not possible to establish a cut height where there is a clear separability among *G.arborea*, palm oil, forest and mangrove. Comparing these two scenes, it is clearly seen that seasons influence the success of the clustering process. This result confirms the importance of seasonality for the classification of vegetation.

The same behaviour, as described above, in terms of differences in the clustering process, was displayed by the two scenes from HyMap. Here, the green dot-dashed lines from figures 4.1 and 4.2 represent an attempt to get more separability but at the expense of getting a lot of clusters. In a specific dendrogram, if there is the needed to draw a cut line at a short distance (near 0), the result is that the number of clusters is too high, which is not ideal for classifying the objects in small classes or categories. Contrary to the *eo1* scenes, the differences are attributed due to the atmospheric conditions from one day to another.

In addition to the lack of multi-temporal hyper-spectral images, they also have limited coverage. For example, in the above mentioned *CARTA* missions only 70% of Costa Rica was successfully acquired with flight lines of ≈ 7 km width.

On the other hand, EO-1 Hyperion has 7.7 km width which is considerably small compared with *Lansat* which has a width of 186 km. However, the potential of these optical sensors is still being explored. One important milestone is not only the free availability of the EO-1 imagery [Fagan and DeFries, 2009], but also the projected satellite-borne missions like the German *EnMap*, which will increase the availability of hyper-spectral data.

5.2 The USB650 Field Spectrometer

Sampling in trees is always a challenge. For example, Cochrane [2000] studied the vegetation reflectance variability for species level classification using hyper-spectral data. Regarding the sampling for this study, 325 spectra from various mahogany foliage and other 190 spectra from foliage of the other 10 species were collected by the authors. However, despite the effort, there is no mention of the sampling technique applied to collect the data.

In situ measurements are likely to have a human-method-produced error as well as measurement device calibration problems [Jensen, 2007]. Despite the possible errors, a method of using spectral libraries to classify satellite images is needed [Rao, 2008]. This author used both *in situ* hyper-spectral measurements and pixel based spectra for an automatic classification in agricultural crop varieties. This interesting approach integrates the spectral libraries into the classification scheme and uses them as the training data.

The most important determinant of the spectral properties of a leaf, particularly in the *VIS*, is the chemical composition of the leaf [Jones and Vaughan, 2010]. From this starting point, the considerations made by Acquaye [1964]; Agyeman and Safo [1997] who made a wide analysis of foliar techniques applied to cocoa crops and *Terminalia ivorensis*, are applied to a spectral signature analysis. These authors pointed out some useful key strategies for foliar analysis in nutrition studies. These factors should be taken into account when taking samples for analysis and for interpreting results of the leaves' spectral properties.

As leaves are very different within the same tree [Jones and Vaughan, 2010], and

depend on the position in the canopy, considerations regarding the timing of the sampling of leaves have to be taken into account [Acquaye, 1964]. For example, the time of the day, the lighting conditions, number of leaves per strata and per tree, the position of the leaves in the canopy, etc. have to be considered.

The *USB650* spectral signature presented is a very simplified view of the interaction of the radiation with leaves. However, the variability of spectral response is high within a species, between ages of the same species (*G.arborea*) and within a single tree. If high spectral resolution data will be used, the variability will also be high. Thus, this variability needs to be incorporated into the classification schemes.

Some of the major clusters are indicated in Figure 4.19. However, there are also clusters formed by two spectra, for example, the edge 54 (see App. C.5) which forms a cluster with spectra of the X101113008 and X102223043. Although this cluster is very close to the edge 69 on the dendrogram, it does not form a cluster with it. This shows precisely how well the *AU* method fits in comparison to the first Hierarchical Clustering applied in this study. What matters in the first *HCA* approach is the height at which the dendrogram is cut and the cutting height set by the number of clusters. Thus, if the *AU*'s in Figure 4.19 are ignored and the dendrogram is cut at the level of approximately 75, the edges 54 and 69 remain in the same cluster. By this time the *AU* is 71% and it can be concluded that these two edges are not supported by the data to form a cluster. It is noteworthy that edge 54 presents ages of 6 and 8 years, while edge 69 has only spectra of 8 years.

Another cluster that mixes one spectrum of 6 years with 9 spectra of 8 years is defined by edge 104. But more critical is the cluster defined by edge 109, where two spectra are combined with 7 other spectra of 8 years. Another cluster that exhibits this problem is defined by edge 90 where 3 of 6 spectra are of 6 years and the another 3 of 18 years old *G.arborea*.

The cluster defined by edge 60, (cluster shown in the spectra of X103331070, 103331071 and X103331072 of 18 years), is at a smaller height in the y-axis in the dendrogram than the one defined by the edge 90. However, the spectra of

X101123014, X101123016 and X101123015 could not be in another cluster that is not defined by the edge 90. Therefore, it can be said that there is a spectral confusion for coverage of *G.arborea* between 6 and 18 years.

There is a total of 55 clusters with an *AU* greater than 95%. This brings it to 97 spectra out of a total of 122, or 79.5%. From the remaining (25 spectra) only 9 spectra belongs to the *G.arborea*. From these 9 spectra, 5 are from the edge of 18 years.

But what does it mean that a remaining 20.4% are not in any cluster that meets that *AU p-values* > 0.95. It is the separability. In other words, these spectra are close to some other spectra but not close enough in terms of the distance to form a cluster.

As described in the previous paragraphs the within variability of the *G.arborea* is evident. However, it also suggests, that there are some strong differences within the three ages (6, 8 and 18 years). Taking into account the within variability, a future study could be done in order to test the within variability of *G.arborea* in different ages but with scene based spectra.

Contrary to the results obtained from the *PCs* scenes spectra where there was spectral confusion, in general, between palm oil, forest and *G.arborea*, the spectral confusion in the *USB650* spectra were found between spectra of leaves of *G.arborea* (18 years) and leaves spectra of mangrove. Specifically, the most important spectral confusion was between 4 spectra of 18 year old *G.arborea* and 6 spectra of mangrove. Also 6 spectra of 18 years were not clustered at all. There were spectral confusions between the spectra of *G.arborea* leaves of 6, 8 and 18 years.

On the other side, 9 of 14 teak spectra, thus 64% of the spectra, are in the cluster defined by the edge 81. However, there is one palm oil spectrum within this cluster. Four of the other spectra of teak are two individual clusters and the remaining (X20203082) does not correspond to any cluster but is very close to form a cluster along with several spectra of a spectrum of mangrove. However, the analysis of the teak and mangrove spectra just describes the general behaviour and suggest the possible spectral confusion with palm oil and *G.arborea* spectra. An

intensive sampling scheme is needed in order to detect and compare the spectral confusion with *G.arborea*.

In tropical forest environments, [Clark et al. \[2005\]](#) found that "the leaf spectral variability among individuals of a certain species, or even within a single crown, was attributed to differences in internal leaf structure and biochemistry (e.g., water, chlorophyll content, epiphyll cover and herbivory)". As some differences could be found within spectra of a single crown, the spectra based analysis could help to understand the correlations between the relations of plant-soil-atmosphere carbon fluxes if it exist. A very interesting approach could be archiving by measuring spectra and chlorophyll content of leaves at the same time and height of CO₂ measures in bio-climatic towers.

5.3 TerraSAR-X Processing and classification

One of the most important steps in the processing of *RADAR* is the speckle reduction. This process tries to remove the grainy salt-and-pepper pattern in the images that is present due to the coherent nature of the radar wave [[Jensen, 2007](#)].

The desired feature of a speckle filter is to smooth out speckles while preserving the useful information. However, there is always a trade-off between these two requirements [[Shi and Fung, 1994](#)]. In [Figure 4.21](#) the more pleasing appearance segments of the TerraSAR-X image are presented. As the speckle filtering mostly remains an estimation problem, the objective is to retrieve the scene reflectivity from the observed speckled *SAR* measurement [[Touzi et al., 2004](#)]. The final choice of the multiple looks that were applied took into account the objective of the whole process which is the ultimate forest/not-forest classification.

After the multi looking and the filtering process, there are spatial distortions in terrain slopes facing a side-looking TerraSAR-X illumination, which are mapped as having a compressed range scale as compared to their appearance if the same terrain was levelled. Foreshortening is a special case of elevation displacement. The effect is more pronounced for steeper slopes (producing the layover as an

extreme form of elevation displacement or foreshortening), and for images with steeper incidence angles. For that reason, the images with an incidence angle of $\approx 23^\circ$ were excluded. The opposite effect occurs for slopes that face away from the radar illumination. This produces the shadows in the images. The shadows are regions hidden behind elevated features such as man-made structures as well as mountains which in the image appear out of the beam of the *SAR* system.

In the visual inspection performed on the classification results presented in Figure 4.26, some small shadowed areas were classified in the same class as water.

Figure D.4 is very instructive because normally the buildings, bridges, etc. are corner reflectors. That means that much of the backscatter is recorded by the antenna. This was the case for the highest values of the bands B2, B7 and B8. However, the mean values are high and there are many values with low intensity. An explanation of this phenomenon is that the *ROIs* capture shadowing areas of the man-made structures. But it is the rest of the bands (B1, B3, B4, B5 and B6) which are critical because of the very low mean intensity values which indicate that every band provides some unique information about the different land covers. As Liao et al. [2009] pointed out, *SAR* has several disadvantages in urban areas which are related to the wavelength and oblique viewing geometry of *SAR* systems.

In order to discriminate the human-made structures, another possibility, not implemented in this study, is to use the coherence information to separate between vegetation and artificial lands as well as to use texture parameters with larger kernels which are quite sensitive to the heterogeneity of urban areas.

Another example is the land cover of *water* which is theoretically a smooth surface. This means that calm and pure water acts as a specular reflector, deflecting most of the energy sent by the antenna which then cannot be recorded. However, as shown in Figure D.10, there are several differences between the different bands. A plausible explanation of this is that water bodies are not always motionless since the wind produces small waves. Also, water bodies have different sediment contents but these do not directly impact the signal. In Figure 4.23 (low leaf part) there is a spot of sediments which has been captured as a water

producing variability between the bands.

TerraSAR-X has a frequency of 9.65 GHz (≈ 3 cm wavelength). At this wavelength, the penetration of the signal mainly takes place in the canopy components and not as much in the stems or branches. As an example of the complexity of the signal backscatter, the land cover *forest* in flat terrain (with dominance of *raffia*) and *palm oil* (old plantation) are very similar. As described earlier, forest in flat terrain (see Section 1.3.3) contains the *Raphia taedigera* (*raffia*) which is a palm typically found in natural swampy areas in tropical lowlands [Kappelle et al., 2002]. Both *raffia* and *palm oil* plants are from the *Arecaceae* family, but in the case of the *SAR* backscatter, the chemical components are not as important as in the passive remote sensing. In the microwave domain, the roughness, dielectric properties which are related to the canopy water content, and the structure of the biomass canopy [Jensen, 2007], play an important role in the amount of energy recorded by the antenna.

However, as showed earlier, depending on the polarization and the incidence angle, the returned backscatter could be very different (see Fig.4.25) for mapping purposes. The raw TerraSAR-X images that were used in this study transmit either H or V polarized pulses and receive in H or V polarization. The resulting product was just in one polarimetric channel, in one of the *HH* and *VV* combinations. However, other acquisition parameters such as *HV* or *VH*, ascending and descending pass directions may be used. Through the use of more acquisition modes, the potential of TerraSAR-X is to be exploited, especially in non flat terrains.

It is clear that the use of more acquisition modes increases the cost of any mapping project. Further theoretical and experimental investigations using fusion techniques, derived band combinations with different polarization *SAR* data are needed for: a better comprehension of the physical behaviour of the different land covers and the optimization and careful selection of the acquisition parameters for specific regions of ecosystems. The cost would be greatly reduced with additional investigation related to regional selection of acquisition parameters.

For accessing TerraSAR-X imagery, in the context of *REDD+* projects, for ex-

ample, a good option is provided by the German Aerospace Centre (DLR) that can also provide the user segment information for scientific applications. Commercial data acquisition is done through the Infoterra GmbH [Roth et al., 2007], which nowadays it is the Astrium GEO-Information Services. In this context, the international donors play an important role providing the necessary founding to acquire the imagery.

In the case of *REDD+*, however, the sponsorship of donors could take care of the data costs; at least as long the market is not developed. In addition, deploying a Forest Carbon Partnership Facility (FCPF) as well as bilateral agreements can address the need of capacity building and training. With more than US\$150 million committed, the *FCPF* is a multilateral World Bank initiative which helps developing countries to prepare for the *REDD+* process [Davis et al., 2010].

Also in the context of *REDD+* as a result-based initiative, where monitoring deforestation and degradation is a challenge, tests have shown that TerraSAR-X can be used for detecting changes in forest cover and to map tropical forest [Scheuchl et al., 2009]. In another approach Kuntz et al. [2011], using a multi-temporal TerraSAR-X High Resolution SpotLight image (1m resolution), showed a selectively logged stand in the Amazon rainforest/Brazil. The logging of individual trees was detected and automatically identified combining three images from different dates [Kuntz et al., 2011].

Related to the issue of capacity building and training, the use of open source software is increasing in popularity. This is also the case for *SAR* processing, specifically with the open source software called *NEST* [ARRAY-ESA, 2011], with which all the TerraSAR-X image processing was tested and done. There are other open source *SAR* software possibilities, for example, the *RAT*¹(Radar Tools) and I.D.I.O.T. (InSAR Deformation Inspection and Observation Tool) [Reigber et al., 2007] among others. These tools open the opportunities for giving and receiving *SAR* courses at very low cost in developing countries.

However, the capacity and training could be focused on fusion techniques, derived band combinations, and leaving all pre-processing process to the data providers.

¹<http://radartools.berlios.de/>

In this way, an initiative or project can focus on just obtaining interesting features that could be for direct use and for integration with different initiatives.

As was discussed earlier for the hyper-spectral approach, a special consideration is needed for the *ROI* selection. In the case of TerraSAR-X applications, one polygon (*ROI*) could capture the true intensity values for one land cover but at the same time this polygon could be capturing another thing. For example, in the case of *infrastructure*, one *ROI* in one band can also capture shadows, and in another band just the *infrastructure* itself could be captured.

This last consideration can be added to the sampling procedure, just as if it were for a pixel based classification or for an exploratory analysis that is valid for optical and for *SAR* image processing. Thus, to provide a complete answer, the following questions have to be taken in account [Congalton and Green, 2009]:

- What are the thematic map classes to be assessed and how are they distributed across the landscape ?
- What is the appropriate sample unit?
- How many samples should be taken?
- How should the samples be chosen?

There are some guidelines which help determine the optimal selection of the *ROIs* [van Genderen et al., 1978]. But the optimal sample size depends on how abundant and fragmented the landscape is and on the classification objectives. In the practice of allocating sampling and training sites, the site location constraints and minimum mapping unit sizes have to be taken into account. For example, in the study site, the infrastructure was in most of the cases small constructions in the middle of big farms, which easily led to some very bright pixels as well as some shadows being present in the same *ROI*. In such a case, the principle that the classification scheme should be mutually exclusive and totally exhaustive would be violated [Congalton and Green, 2009].

An object-oriented approach takes the form, textures and spectral information into account. Its classification phase starts with the crucial initial step of grouping neighboring pixels into meaningful areas which can be handled in a later step

of classification. In this procedure, the grouping capabilities increase with the number of bands. This is valid for optical and *SAR* imaging processing. In the optical domain, the use of indexes is a very common procedure, for example, [Fuchs et al. \[2009\]](#) give a description and references for the satellite image bands and transformations which are commonly used and are very well tested. However, this is not the case of the *SAR* remote sensing for two main reasons: first, the availability of the *SAR* imagery is a constrain due to the high prices as well as the need of highly skilled software. Nevertheless, as was discussed earlier, this not an excuse. Second, the knowledge needed for processing *SAR* imagery is also very different to the one for the optical domain.

The *PDR* is an example of the use of valuable information through one index. This index is a band ratio between the *HH* and *VV* polarizations. [Figure 4.25](#) shows a band index that is very useful to discriminate bare soils based on the differences in the moisture content. The surface soil moisture is a key parameter that influences numerous environmental processes that occur over a large range of spatial and temporal scales [[Barrett et al., 2009](#)]. It is also an important link between the land surface and the atmosphere [[Owe et al., 2001](#)].

If an object oriented classification is being done, the information contained in the *PDR* band is important because it can discriminate the texture of the bare soils. The final objective of any classification is to avoid class confusions. Here, any additional information that helps to discriminate the classes is useful.

Finally, there are several considerations related to the error matrix. First, the reference data such as RapidEye mosaics or aerial photographs are not exempt of errors and are not always representative of the entire classification because of the cloud coverage. Second, the reference data and classified images are not perfectly co-registered. For that reason, the actual accuracy of the classification is unknown because it is impossible to perfectly assess the true class of every pixel. To overcome the mentioned aspects, the "expert analysis based" approach was needed based on our expertise gained during our fieldwork and previous knowledge of the study site.

The classification key presented in [Figure 4.26](#) could be used to develop an ac-

curate forest/non-forest classification. For example, if each class becomes one stratum, there is more control over how accurately the user accuracy is estimated. With this technique, one can ensure that there are enough samples for smaller classes. It is also possible to define some hard to get to strata. For example, spectral confusion were found in mangroves and palm oil plantations. The above strategy is useful for both the training phase as well as the evaluation of the accuracy.

Nowadays on the microwave domain, there is a continuous development of new *SAR* processing tools. More sensors such as the Italian CosmoSkyMed and the Canadian RADARSAT-2, increase the opportunities for improving the analysis procedures of monitoring of tropical ecosystems.

5.4 Climate, Vegetation and Remote Sensing in Costa Rica

For this subsection, the following variables: precipitation (*PRE*), potential evapotranspiration (*PET*) and normalized difference vegetation index (*NDVI*) were used to analyse the seasonality and vegetation dynamics of the study area.

The amount of *PET* as a demand of the atmosphere [Tsakiris et al., 2007] represents water consuming activities. However, it is a hypothetical quantity which cannot be directly measured Holdridge [1962]. There are several methods to estimate *PET* in the literature which are based on the mean air temperature only. Most of them are intended for temperate zones; hence a correction is needed for low latitudes as in the case of Costa Rica. Information regarding *PET* can be used as a baseline for climate change studies in which potential distributions of flora and fauna can be predicted under the impact of climate change. From such information, priority areas for biodiversity conservation can be identified. Furthermore, the difference between the *PRE* and *PET* can be easily calculated on a monthly basis. This provides a simple measure of water surplus or deficit. One issue that makes *PET* important is that it can be used to calculate the actual evapo-transpiration. With this quantity it is possible to derive a soil wa-

ter balance measure which is the ultimate information needed to establish the relationship between ecosystems, the water balance and the service of the water regulator.

NDVI time series provide relevant information of the state of vegetation. Therefore, this information is very valuable when it is linked with climatic data for the purpose of addressing climate change, land use as well as land use and land cover (LULC), or agriculture. As an example of these relationships [Shisanya et al. \[2011\]](#), found that the trends of biomass in the south-east of Kenya are explained by changes in *LULC* but also by rainfall variability.

The accumulated *NDVI* integrated over the growing season is related to the Net Primary Production (NPP) [[Zhao and Running, 2008](#)]. Annual *NPP* is a direct measure of the quantity of carbon fixed by the vegetation through photosynthesis in a year since it is linked with the photosynthetic activity of the vegetation. However, caution has to be taken when *NDVI* is applied in tropical forest with higher biomass because of the rapid saturation of the *NDVI* values [[Wang et al., 2005](#)].

The *NDVI* time series and climatic data could also serve as guide lines for a better understanding of some spectral confusions on hyper-spectral data. Recently, [Somers and Asner \[2012\]](#) used 4 years EO1 Hyperion time series spectra to analyse the seasonal dynamics of the canopy of four tree species of the Hawaiian ecosystems. The authors found that native and invasive trees are unique in their reflectance and derivative spectral properties. Multi-temporal studies highlight the fact that photosynthetic activity and functional properties in different land covers directly express themselves in spectral differences. *NDVI* could serve as an indicator for imagery selection as well as planning hyper-spectral airborne missions. For example, incorporating the reflection behaviour of plants in the near infra-red spectrum, the *NDVI* can be used as an indicator for healthy/non-healthy vegetation and its condition at a specific point of time.

Gathering and exploitation of information delivered by airborne or satellite sensors can be performed in a variety of different ways depending on the user's intention and the field of research. One major issue is how to handle the data in

an appropriate manner. As the atmospheric conditions are changing over time and location, the solar irradiance is never a constant. Therefore, the calculation of special spectral indices derived from remotely sensed data is a suitable way. However, even then such results should be considered carefully because the indices can either be obtained from a sensor voltage output, digital numbers and radiance, or reflectance values producing different results [Jackson and Huete, 1991].

Many studies show the (seasonal) relation between *NDVI* and rainfall. But in the case of tropical countries such as Costa Rica, there is only a small variation in the monthly mean temperatures. For that reason, it is also crucial that the precipitation and potential evapo-transpiration are taken into account in order to have a better understanding of the interactions and relationships of climate and vegetation.

Another issue worth mentioning is the subsequent reaction or growth of living vegetation after rainfall. This circumstance can be seen in Figure 1.2. For example, in the month of October in the south of Costa Rica, the *NDVI* is lowest when the *PRE* is at its highest, then the *PRE* decreases to a minimum in the month of February. On the other hand, the *NDVI* has the highest values in the months of January and February. But after January the curve of *NDVI* decreases until the month of April, where the highest value of *PET* is observed. In the months of January to April there are several changes in ecosystems in the south of Costa Rica.

The dates of acquisition of the imagery that were used are in this period of climate and vegetation activity. For example, one Hyperion scene is from February 23, 2006 and the second one from March 15, 2005. In the case of the HyMap scenes, one is from March 14, and the second one from March 15, 2005. In the case of the TerraSAR-X one image is from the month of April, where for example, there is a larger difference in terms of *PRE* than in the rest of the TerraSAR-X images that were acquired from the months of May and June. In *SAR* multi-temporal data analysis is very common to separate classes, which can be a function of water content and vegetation response according to changes of the di-electric constant.

Regarding the vegetation water content, [Danson et al. \[1992\]](#) found high correlations between the wavelengths corresponding to the slopes on the edges of the water absorption bands and the leaf water content (but insensitive to differences in leaf structure) in a laboratory measurements of the spectral reflectance of leaves from a variety of species. If the relation between spectral reflectance and leaf water content is important, the acquisition dates within a year are also important.

[Rao \[2008\]](#), suggest that matching the spectra in order to label unknown spectra could be a useful method for automatic identification and extraction of various crop species and may also lead to proper selection of training pixels for mapping of various crop varieties in an area. If spectral libraries will provide a source of reference spectra that can aid the interpretation of hyper-spectral and multi-spectral images [[Rao, 2008](#)], they have to take into account the intra-species variability in various types of soil, mineral and vegetation species [[Price, 1994](#)]. If valuable information from spectral libraries is developed, some factors such as time, year, growth stage, type of variety, soil and climate have to be considered.

An alternative approach is to use a signature derived from training data and a matching image from another period and apply this to the images for which no training data are available [[Laborte et al., 2010](#)].

As [Rao \[2008\]](#) pointed out after an extensive literature review, there are no systematic studies carried out for the development of an agriculture-specific spectral library. If this is the situation for agricultural environments, then it is even more complicated to develop a systematic spectral library for trees in a forest or in a forest plantation following the principles of sampling.

Normally, in the optical domain of remote sensing, the acquisition and the availability of cloud free images are a challenge. In the hyper-spectral field it's even more challenging. The cluster analysis of the hyper-spectral scenes and the field based spectra that were performed could serve as an exploratory data analysis to identify the similarities as well as the dissimilarities among different land covers. This previous step is desirable prior to designing a sampling procedure even if the spectra are taken from the pixels in the imagery or from the *in situ* measurements.

In a supervised classification for example, the inputs are the training data and a set of reference data. The reference data (truth) are needed in order to evaluate the classification error. This error is the discrepancy between the situation depicted on the thematic map and reality [Foody, 2002]. If a statistically sound accuracy assessment is desirable, both sets of data have to be as high in quality and quantity as possible.

To fill the gaps in the training data sets, the use of spectral libraries could be an alternative, especially in classification schemes where the larger regions comprise multiple images or when images from different years (but also months) are used. However, the use is limited when the imagery uses broad-bands. Thus, the finer spectral differences coming from the spectral libraries are not detectable. Further research is needed using spectral libraries classifying multi-derived *SAR* bands.

In large projects where the use of remote sensing is needed or in the frame of *REDD+* projects, forest area maps are required. To provide the spatial and the large spatial view, the use of spectral signatures could help to better understand the specific land cover as well as to identify the spectral confusion such as that detected in this study among *G.arborea*, the palm oil plantations and the forest in flat slopes with the dominance of *raffia*. A further investigation could be the use of local *in situ* spectral signatures and their variation during the year to determine how correlated these spectra are with the *NDVI* time series.

The use of spectral libraries requires a very thorough-full data calibration of both the reference spectra and the remote sensing data. A further investigation is required on the calibration of the remote sensing data using more accurate *DEMs*. Related to the very high resolution TanDEM-X mission, Riegler et al. [2011] pointed out that "in 2014, a homogeneous elevation model will be available for the Earth's land surface featuring a vertical accuracy of 2-4 meters (relative) and 10 meters (absolute), within a horizontal raster of 12 meters" [Riegler et al., 2011]. This mission will open the opportunity to calibrate the *SAR* imagery even better and thus reduce the shadow problems even more.

Since 1972 remote sensing has been providing accurate measurements for monitoring the global forest. However, there are several challenges that the global

community has to take into account if remote sensing is to play a significant role in forest monitoring. There are several issues that have to be considered to guide this discussion and this is a matter the scientific community should be concerned with as well. At the same time, the political and social sectors are relevant in achieving the necessary consensus. Also, in the framework of the *REDD+* initiative, operational approaches are needed.

The availability of satellite imagery and its methods must be guided by the best accuracy that is possible and follow scientific principles in accordance with the ideas of monitoring, reporting and verification.

Many remote sensing instruments have been built in many different satellites and platforms over the years. The purposes have been very different as many different sensors exist. For example, a sample of this variability is presented in the Appendix 3 of [Jones and Vaughan \[2010\]](#) and there are some websites which provide up-to-date lists, for example in [Union of Concerned Scientist \(UCS\) \[2012\]](#). The problem is that many of the sensors are built for a specific time which might be too short for vegetation monitoring purposes. However, there are other sensors that have been there for many years, and for longer time than they had been designed for, the TM5 instrument for example was in operation until November 2011, many years after its planned design life.

There are many airborne-sensors but the data are not available after the specific programs or laboratories have used the data to their own purposes. This also poses limitations if a vegetation monitoring system is desirable.

Despite of the data availability and continuity of recent or past sensors, there are several upcoming and promising new sensors such as Sentinel-2, SPOT continuity through Pleiades system, SPOT 6 and 7 and Landsat Continuity Mission. In the microwave domain TerraSAR-X2 will be launched in 2014 [[Knabe, 2010](#)] and Sentinel C-band will be launched around 2013 [[Torres et al., 2012](#)].

Chapter 6

Conclusions

6.1 Hyper-spectral Conclusions

The analysis of hyper-spectral optical data is limited not only because of restrictions and limitations in the availability of imagery, but also and most importantly, due to the limitations of the temporal resolution and the coverage of on single scene. The temporal resolution could be solved by the use of aerial platforms. However, in the case of the two Costa Rican missions, one that took place in 2003 and the second in 2005, 100% of the 51.000 km² of the Costa Rican surface was not acquired due to adverse weather conditions and cloud coverage, demonstrating the restriction of these enterprises.

Nevertheless, space-borne hyper-spectral imagery could scan the same area at regular time intervals making the monitoring of vegetation and their spectral confusion possible while taking into account the vegetation phenology. Still, the availability of the imagery continues to be a constraint due to the cloud coverage and the time resolution.

Because of the phenology variability, removing the atmospheric effects is a mandatory task since it permits inter-sensor comparisons. However, these are incomplete atmosphere models. A special challenge for this research was precisely atmospheric corrections, because it was not possible to automatically perform

elementary operations on the imagery using the *FLAASH* module.

The ability to obtain more spectral and spatial details comes at the expense of a very high data volume which also poses computational and methodological challenges in the processing of the imagery with high spectral dimensionality.

Principal Component Analysis was needed as a data reduction technique because of the amount of hyper-spectral information. However, the amount of useful information that was discarded is ignored.

The highest scene based spectral variability was in the *NIR* portion of the spectra. This suggests that further analysis extracting spectra from some specific regions could solve some of the spectral confusion encountered.

Regarding the clustering process, this technique is not exempt from some problems. For example, the effectiveness of the method depends on the definition of a *distance*. The result of the clustering can be interpreted in different ways, as well as the cut height, which defines the number of clusters. For instance, multiple sensor and measurement conditions are needed in order to establish spectral libraries. One of the most important issues in cluster analysis is the evaluation of clustering results to find the partitioning that best fits the underlying data.

HCA analyses showed that cluster solutions of the *PCs* spectra from the two sensors present good separability at a global scale. However, they were subject to systematic differences; only the *eo1_kw* and the *hy_14* *PCs* spectra did not present spectral confusion among *G.arborea*, palm oil and the forest. The first approach of *HCA* defines the number of clusters depending on the cut height. At an arbitrary cut height, only the dendrograms of the *eo1_kw* and *hy_14* scene did not present spectral confusion at all. At the height of 0.6 *G.arborea* presented spectral confusion with palm oil plantations. These results suggested that hyperspectral imagery should not be acquired at a very high spatial resolution to provide adequate discrimination of land covers.

With the Approximately Unbiased *HCA*, the spectral detail provided by hyper-spectral analysis was successful and allowed to identify where exactly the spectral confusion was among different vegetation land covers. For example, the *AU*

method applied to *hy_15* scene presents spectral confusion only in three spectra of *G.arborea* and palm oil.

Each data set has its variability. These variabilities are determined by the sensor characteristics, angle of measure, light intensity, water content, and others. The same sensor under different conditions will give different spectra and different cluster results. Furthermore, data collection techniques need to acquire time series spectral signatures.

Regarding the field based spectra, there was spectral confusion in the majority of 18 years of leaves of *G.arborea* and mangrove. Specifically, the most important spectral confusion was between 4 spectra of an 18 year old *G.arborea* and 6 spectra of mangrove. Also 6 spectra of 18 years were not clustered at all. There was spectral confusion between the spectra of *G.arborea* leaves of 6, 8 and 18 years.

An extensive amount of variability was found in the spectral response of foliage within a species and even within the ages of leaves of *G.arborea*. Due to the non-unique nature of spectral responses of the leaves of *G.arborea*, separation of vegetation at the species level is never likely to be perfect, but it can be a useful tool for the development of specific spectral libraries as well as a tool to explain the variability of spectra derived from hyper-spectral imagery.

In addition, robust methods are needed to address the optimal data reduction and fusion (not implemented here) as well as the classification of multi-source data in conjunction with hyper-spectral images, *SAR* imagery; and variable spectral-spatial resolution. If valuable information from spectral libraries is developed, some factors such as time, year, growth stage, type of vegetation variety, soil and climate have to be considered.

In this study, the field based spectra that clustered *G.arborea* and mangrove reflectance should be interpreted with caution. A more realistic result of land covers separability is to use data from airborne or space-borne spectrometers, with all of the associated variability introduced by poor radiometric calibration, positive and negative scattering, angle of views and angle of illumination.

6.2 TerraSAR-X Conclusions

More than 50% of the study area has slopes that vary from rolling to highly steep ones. However, with the methodology followed, only 10.9% of the area was classified as foreshortening. On one hand, the use of ascending and descending orbits and the calibration based on very high resolution *DEMs* can reduce even more the shadow and foreshortening problems. On the other hand, if a pixel base multi-band supervised classification is done, the training as well as the validation points or polygons have to take care of the high differences within the artificial *SAR* bands.

The object-oriented approach takes into account textures and spectral information. Its classification phase starts with the crucial initial step of grouping neighboring pixels into meaningful areas, which can be handled in the later step of classification. In this procedure the usefulness of more bands is shown as this increases the grouping capabilities. This is valid for optical and *SAR* imaging processing. Specifically, in this study, a simple fast and reliable approach to combine the polarized TerraSAR-X imagery is presented. More band combinations, fusions and indexes are needed especially for *SAR* sensors like TerraSAR-X. With more acquisition parameters, the potential of TerraSAR-X is to be exploited.

As shown in section 4.3, the use of several derived bands is necessary to obtain optimal results for hybrid texture classification. The overall accuracy for forest and non-forest classification for the study site was 86%. The proven band combinations and fusions are an adequate foundation for today's classification tasks using TerraSAR-X data. However, the results obtained are highly dependent on the reference data which were not optimal because of the cloud coverage of the actual RapidEye imagery and the differences in the acquisition dates from the other auxiliary data.

Nowadays, the available *SAR* systems have the potential to develop powerful tools to map and monitor regions with frequent cloud coverage. The accessible open source software provides new opportunities to process *RADAR* images at low cost.

More collaboration and partnerships are needed between agencies and organizations that have agreements with governmental institutions in developing countries in order to use and observe the forest changes.

6.3 Remote Sensing and Climate

Global data sets that are available free of charge for science and for developing countries, provide a starting point to run models at a local scale. However, we need to understand and close the information gaps in order to develop methods to identify relationships between climate and vegetation at a local scale.

In order to apply *NDVI* time series analysis, it is necessary to collect data on a regular basis and only remote sensing is able to provide such data. In the long term, vegetation modelled as *NDVI* will provide valuable information about climate change. In addition, these time-series data are necessary to build or adjust ecosystem maps as well as to understand vegetation phenological differences.

The use of remote sensing data has the potential to view the ecosystems at a global scale in terms of time, cost and cover. However, the interpretation requires an understanding of how the energy interacts with the different land covers and how to extract the information of interest from the different remote sensing systems.

A crucial step to extract valuable information from remote sensing data is the atmospheric correction. In this study the imagery used (including the free VEGE-TATION products), the atmospheric effects are still present because the methods used to remove these effects are model-based and the information used to archive this models are based on general conditions. Moreover, for comparison purposes, further considerations related to *NDVI* such as soil moisture must be taken into account, especially in areas of evergreen lowlands swamp forests.

Soil water storage quantification is needed in order to calculate the effect of water movements on the *NDVI*, but also, for multi-seasonal remote sensing data analysis. For that reason, the quantification of the actual evapo-transpiration is needed.

Through this study and analysis we wish to make a contribution to improving approaches for remote sensing based on forest cover estimations. Such information may also be useful for monitoring purposes and could help to develop a national standard definition using remote sensing approaches.

At a country level, the efforts to measure forests using remote sensing technologies face several technical challenges. Ground truth data must be acquired across diverse ecosystems or using different sources. Creating national forest maps with high, regionally consistent accuracy is demanding. Choices are necessary in selecting imagery with the appropriate trade-off in cost and different resolutions (spatial, spectral, time and radio-metrical). Challenges also remain to process time series imagery in automated or semi-automated manner, which increases the amount of data and the information to accurately monitor forest over large areas.

Finally, emerging and in-development sensors either passive or active such as the global TanDEM-X, EnMAP, ALOS-2, SPOT6 and 7, Landsat Data Continuity Mission and PAZ, etc. will provide the needed sustainability and operational use for *REDD+* purposes.

Bibliography

- Abonyi, J. and Feil, B. F. (2007). *Cluster Analysis for Data Mining and System Identification*. Birkhuser Verlag AG. [36](#)
- Acquaye, D. K. (1964). Foliar analysis as a diagnostic technique in cocoa nutrition. Sampling Procedure and Analytical Methods. *ournal of the Science of Food and Agriculture*, 15:855–863. [111](#), [112](#)
- Adler-Golden, S., Matthewa, M., Bernsteina, L., Levinea, R., Berka, A., Richtsmeiera, S., Acharyaa, P., Anderson, G., Feldeb, G., Gardner, J., Hoke, M., Jeong, L., Pukall, B., Mellob, J. Ratkowski, A., and Burkec, H. (1999). Atmospheric Correction for Short wave Spectral Imagery Based on MODTRAN4. *SPIE Proceeding, Imaging Spectrometry*, 3753. [25](#), [26](#)
- Agyeman, V. K. and Safo, E. Y. (1997). Mineral nutrition and die-back in Terminalia ivorensis A. Chev. in Ghana. *Journal of Tropical Ecology*, 13:317–335. [111](#)
- Albregtsen, F. (2008). Statistical Texture Measures Computed from Gray Level Coocurrence Matrices. *on line*, on line:14. [65](#)
- ARRAY-ESA (2011). NEST: Next ESA SAR Toolbox. Array Systems Computing-contributors and European Space Agency. [59](#), [61](#), [62](#), [95](#), [117](#)
- ARRAY-ESA (.N.d.). *NEST: Next ESA toolbox*. European Space Agency. Array Systems Computing-contributors and European Space Agency. [63](#)

- Asner, G. P. (1998). Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sens. Environ*, 64:234–253. [22](#)
- Asner, G. P., Jones, M. O., Martin, R. E., Knapp, D. E., and Hughes, F. R. (2008). Remote sensing of native and invasive species in Hawaiian forests. *Remote Sensing of Environment*, 112(5):1912–1926. [Earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue](#). [15](#)
- Asner, G. P., Treuhaft, R. N., and Law, B. E. (2000). Vegetation Structure from Quantitative Fusion of Hyperspectral Optical and Radar Interferometric Remote Sensing. In *Summaries of the 9th JPL Airborne Earth Science Workshop, Pasadena, CA, USA*. [12](#)
- Barrett, B. W., Dwyer, E., and Whelan, P. (2009). Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques. *Remote Sensing*, 1. [30](#), [119](#)
- Beck, R. (2003). *EO-1 User Guide v. 2.3*. USGS Earth Resources Observation Systems Data Center (EDC), 2,3 edition. [19](#), [40](#), [50](#)
- Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988). *The new S language: a programming environment for data analysis and graphics*. Wadsworth & Brooks/Cole Advanced Books & Software. [52](#)
- Becker, S. and Plumbley, M. (1996). Unsupervised neural network learning procedures for feature extraction and classification. *Applied Intelligence*, 6:185–203. [10.1007/BF00126625](#). [34](#)
- Beleites, C. (2009). *Package hyperspec*. University of Trieste. [51](#)
- Borengasser, M., Hungate, W. S., and Watkins, R. (2008). *Hyperspectral Remote Sensing: Principles and Applications*. CRC press. [25](#)
- Breidenbach, J., Ortiz, S. M., and Reich, M. (2009). Forest monitoring with TerraSAR-X: first results. *Eur J Forest Res*. [38](#)
- Castro-Esau, K. L., Snchez-Azofeifa, G. A., Rivard, B., Wright, S. J., and Quesada, M. (2006). Variability In Leaf Optical Properties Of Mesoamerican Trees

- And The Potential For Species Classification. *American Journal of Botany*, 93(4):517-530. [21](#), [31](#)
- Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, 1(2):245–276. [51](#)
- Chen, G., Jaradat, S.-A., Banerjee, N., Tanaka, T., Ko, M. S. H., and Zhang, M. Q. (2002). Evaluation and Comparison of Clustering Algorithms in Analyzing ES Cell Gene Expression Data. *Statistica Sinica*, 12:241–262. [35](#)
- Cheng, Q., Varshney, P. K., and Arora, M. K. (2006). Logistic Regression for Feature Selection and Soft Classification of Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, 3(4):491–494. [32](#)
- Clark, D. A., Brown, S., Kicklighter, D. W., Chambers, J. Q., Thomlinson, J. R., and Ni, J. (2001). Measuring Net Primary Production in Forest: Concepts and Field Methods. *Ecological Applications*, 11(2):356–370. [12](#)
- Clark, M. L., Roberts, D. A., and Clark, D. B. (2005). Hyperspectral discrimination of tropical rainforest trees species at leaf to crown scales. *Remote Sensing of Environment*, 96(34):375–398. [21](#), [114](#)
- Clausi, D. (2002). An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1):45–62. [65](#)
- Cochrane, M. A. (2000). Using vegetation reflectance variability for species level classification of hyperspectral data. *int. j. remote sensing*, 21(10):2075–2087. [111](#)
- Cocks, T., Jensen, R., Stewart, A., Wilson, I., and Shields, T. (1998). The HYMAP Airborne Hyperspectral Sensor: The System, Calibration and Performance. In *1st EARSEL Workshop on Imaging Spectroscopy, Zurich, October 1998*. [43](#), [50](#)
- Cohen, W. B. and Goward, S. N. (2004). Landsat's Role in Ecological Applications of Remote Sensing. *BioScience*, 54(6):535–545. [12](#)

- Commaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603619. [35](#)
- Congalton, R. G. and Green, K. (2009). *Assessing the Accuracy of Remotely Sensed Data Principles and Practices. Second edition*. CRC press. Taylor and Francis Group, LLC. [66](#), [118](#)
- Danson, F. M., Steven, M. D., Malthus, T. J., and Clark, J. A. (1992). High-spectral resolution data for determining leaf water content. *International Journal of Remote Sensing*, 13:461–470. [123](#)
- Daughtry, C. and Walthall, C. (1998). Spectral Discrimination of Cannabis sativa L. Leaves and Canopies. *Remote Sens. Environ*, 64:192–201. [16](#), [22](#)
- Davis, C., Williams, A., Goers, L., Daviet, F., and Lupberger, S. (2010). Getting Ready with Forest Governance: A Review of the World Bank Forest Carbon Partnership Facility Readiness Preparation Proposals. Working Paper Version 1.4, World Resources Institute. [117](#)
- Dennison, P. E. and Roberts, D. A. (2003). The Effects of Vegetation Phenology on Endmember Selection and Species Mapping in Southern California Chaparral. *Remote Sensing of Environment*, 87:295309. [21](#), [44](#)
- DLR (2007). TerraSAR-X Ground Segment Lever 1b Product format specification. Technical report, DLR. [58](#)
- Duin, R., de Ridder, D., and Tax, D. (1997). Experiments with a featureless approach to pattern recognition. *Pattern Recognition Letters*, 18:1159–1166. [52](#)
- Elach, C. and van Zyl, J. (2006). *Introduction to the Physics and Techniques of Remote Sensing*. A John Wiley and Sons, Inc, second edition edition. Jin Au Kong (ed). [22](#), [23](#), [24](#)
- Fagan, M. and DeFries, R. (2009). Measurement and Monitoring of the Worlds Forests: A Review and Summary of Remote Sensing Technical Capability, 2009-2015. Rff report, Columbia University. [2](#), [3](#), [15](#), [16](#), [22](#), [111](#)

- Faller, N. and Weber, M. (2007). TerraSAR-X and TanDEM-X: Revolution in Spaceborne Radar. In *Geoscience and Remote Sensing Symposium, IGARSS 2007. IEEE International*. 58
- FAO (1948). Forest resources of the world. *Unasylva*, 2(4). FAO (Food and Agriculture Organization of the United Nations). 12
- FAO (2001). Global Forest Resources Assessment 2000 - FAO Main Report. FAO Forestry Paper 140, FAO (Food and Agriculture Organization). 2
- FAO (2006). Global Forest Resources Assessment 2005. Technical Report FAO Forestry Paper 147, FAO (Food and Agriculture Organization of the United Nations), Rome. 2
- Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2):179–188. 37
- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80:185–201. 124
- Fried, R., Schettlinger, K., and Borowski, M. (2011). *robfilter: Robust Time Series Filters*. R package version 3.0. 49
- Frost, V. S., Stiles, J. A., Shanmugan, K. S., and Holtzman, J. C. (1982). A model for radar images and its application to adaptive digital filtering of multiplicative noise. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-4(2):157–166. 45, 62
- Fuchs, H., Magdon, P., Kleinn, C., and Flessa, H. (2009). Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory. *Remote Sensing of Environment*, 113:518–531. 17, 119
- Gade, M. and Stelzer, K. (2000). On the Joint Use of Microwave and Optical Remote Sensing Sensors for the Observation of Dry-Fallen Intertidal Flats in the German Bight. In *28th International Symposium on Remote Sensing of Environment, March 27-31, 2000, Cape Town, South Africa*. 30

- Gao, B.-C., Davis, C. O., and Goetz, A. F. H. (2006). A Review of Atmospheric Correction Techniques for Hyperspectral Remote Sensing of Land Surfaces and Ocean Color. In of Electrical, T. I. and Electronics Engineers, I., editors, *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on Digital Object Identifier*. 0-7803-9510-7. [26](#)
- Gao, B.-C., Heidebrecht, K. B., and Goetz, A. F. (1993). Derivation of scaled surface reflectances from AVIRIS data. *Remote Sensing of Environment*, 44(2-3):165–178. [24](#), [43](#)
- Glenn, E. P., Huete, A. R., Nagler, P. L., and Nelson, S. G. (2008). Relationship Between Remotely-sensed Vegetation Indices, Canopy Attributes and Plant Physiological Processes: What Vegetation Indices Can and Cannot Tell Us About the Landscape. *sensors*, 8:2136–2160. [17](#)
- Goetz, A. F. H., Kindel, B. C., Ferri, M., and Qu, Z. (2003). HATCH: Results From Simulated Radiances, AVIRIS and Hyperion. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6):1215–1222. [25](#)
- Goetz, A. F. H., Vane, G., Solomon, J. E., and Rock, B. N. (1985). Imaging Spectrometry for Earth Remote Sensing. *Science*, 228:1147–1153. [20](#)
- Grainger, A. (2008). Difficulties in tracking the long-term global trend in tropical forest area. *PNAS*, 105(2):818–823. [13](#)
- GRASS Development Team (2011). Geographic Resources Analysis Support System (GRASS) Software. Version 6.4.1. Open Source Geospatial Foundation Project.<http://grass.osgeo.org>. [152](#)
- Grieser, J., Gommers, R., and Bernardi, M. (2006). New LocClim - the Local Climate Estimator of FAO. *Geophysical Research Abstracts*, 8. [154](#)
- Haines-Young, R. and Potschin, M. (2007). The Ecosystem Concept and the Identification of Ecosystem Goods and Services in the English Policy Context. Review Paper to Defra, Project Code NR0107. Technical report, DEFRA, Department of Environment Food and Rural Affairs. 21 pp. [12](#)

- Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, 17:107–145. [34](#), [35](#), [36](#)
- Hamon, W. (1963). Computation of Direct Runoff Amounts From Storm Rainfall. *Int. Assoc. Sci. Hydrol. Pub.*, 63:52–62. [154](#)
- Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Systems, Man, and Cybernetics Society*, 3(6):610–621. [65](#)
- He, K. S., Rocchini, D., Neteler, M., and Nagendra, H. (2011). Benefits of hyperspectral remote sensing for tracking plant invasions. *Diversity and Distributions*, 17(3):381–392. [15](#)
- Hengl, T. (2009). *A Practical Guide to Geostatistical Mapping*. Office for Official Publications of the European Communities. [152](#), [153](#)
- Hoffbeck, J. P. and Landgrebe, D. (1995). Classification of High Dimensional Multispectral Data. Technical report, School of Electrical Engineering Purdue University. TR-EE 95-14. [34](#)
- Holdridge, L. R. (1962). The determination of atmospheric water movements. *Ecology*, 43 (1):1–9. [120](#), [154](#)
- Honaker, J., King, G., and Blackwell, M. (2011). *Amelia: Amelia II: A Program for Missing Data*. R package version 1.5-4. [152](#)
- Hornik, K. (2012). *clue: Cluster Eensembles*. R package version 0.3-44. [53](#)
- Houghton, R. A. (2005). Aboveground forest biomass and the global carbon balance, Global Change. *Biol*, 11:945958. [13](#)
- Infoterra (2008). Radiometric Calibration of TerraSAR-X Data. Technical report, Infoterra GmbH. [61](#)
- ITT Visual Information Solutions (2012). ENVI version 4.8. Boulder, Colorado. [42](#), [65](#)

- ITT Visual Information Solutions (N.d.). *Synthetic Aperture Radar and Sarscape*. 27
- Izenman, A. (2008). *Modern multivariate statistical techniques: regression, classification, and manifold learning*. Springer texts in statistics. Springer. 33, 51, 52, 53
- Jackson, D. R. and Huete, R. A. (1991). Interpreting Vegetation Indices. *Preventive Veterinary Medicine*, 11:185–200. 15, 17, 122
- Jacquemoud, S. and Ustin, S. (2003). Application of radiative transfer modes to moisture content estimation and burned land mapping. *In Proceedings of the 4th International Workshop on Remote Sensing and GIS applications to Forest Fire management, Innovative concepts and methods in fire danger estimation*. 22, 23
- Jacquemoud, S., Verhoef, W., Beret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., Francois, C., and Ustin, S. (2009). PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment*, 113:S56S66. 15
- Jain, A. K. and Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice-Hall. 34
- Jensen, J. R. (2007). *Remote Sensing of the Environment: An Earth Resource Perspective*. ISBN-10: 0131889508. Prentice Hall, 2/e edition. 18, 19, 21, 24, 26, 28, 29, 30, 58, 61, 97, 100, 101, 102, 111, 114, 116
- Ji, L. and Peters, A. J. (2003). Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices. *Remote Sensing of Environment*, 87(1):85–98. 17
- Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, 6 edition. Petra Recter (ed). 33, 35
- Jones, H. G. and Vaughan, R. A. (2010). *Remote Sensing of Vegetation: Principles, techniques, and applications*. Oxford University Press. 1, 15, 18, 19, 20, 22, 23, 24, 29, 43, 45, 111, 125

- Jung-Rothenhusler, F., Weichelt, H., and Pach, M. (2007). RapidEye A Novel Approach to Space Borne Geo-Information Solutions. In *Heipke, C C .; Jacobsen, K. and Gerke, M. (eds.): Proceedings of the ISPRS Hannover Workshop 2007: High-Resolution Earth Imaging for Geospatial Information. Hanover, Germany.* 50, 64
- Kappelle, M., Castro, M., Acevedo, H., Gonzalez, L., and Monge, H. (2002). *Ecosystems of the Osa Conservation Area (ACOSA)*. Editorial Inbio. 9, 10, 65, 116
- Kasischke, E. S. and Christensen-Jr, N. L. (1990). Connecting forest ecosystem and microwave backscatter models. *International Journal of Remote Sensing*, 11(7):1277–1298. 31
- Kaufman, L. and Rousseeuw, P. (2005). *Finding groups in data: an introduction to cluster analysis*. Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley. 34, 53, 54, 56
- Keshava, N. and Mustard, J. F. (2002). Spectral unmixing. *IEEE Signal Processing Magazine*, pages 44–57. 15
- Khan-Ullah, K. and Jian, Y. (2007). Polarimetric Synthetic Aperture Radar Image Classification by a Hybrid Method. *Tsinghua Science and Technology*, 12 (1):97–104. 33
- Kim, Y., Street, W., and Menczer, F. (2002). *Feature selection in data mining. In: Data mining: opportunities and challenges*. Hershey-Idea. Wang, J. 80-105. 65
- Klein, M. E., Aalderink, B. J., Padoan, R., de Bruin, G., and Steemers, T. A. (2008). Quantitative Hyperspectral Reflectance Imaging. *Sensors 2008*, 8, 5576-5618, 8:5576–5618. 32
- Kleinn, C., Corrales, L., and Morales, D. (2002). Forest Area in Costa Rica: A Comparative Study of Tropical Forest Cover Estimates over Time. *Environmental Monitoring and Assessment*, 73:17–40. 2, 14

- Knabe, S. (2010). TerraSAR-X Status and Future Plans. In *SAR oceanography workshop, SeaSAR*. On line at: <http://earth.eo.esa.int/workshops/seasar2010/>. Accessed 12.04.2012. 125
- Kruse, F. A. (2004). Comparison of ATREM, ACORN, AND FLAASH Atmospheric Corrections Using Low-Altitude AVIRIS Data of Boulder, CO. Technical report, Horizon GeoImaging, LLC, 4845 Pearl East Circle, Suite 101, Boulder, CO 80303-6113 USA. 25
- Kuntz, S. (2010). Potential of spaceborne SAR for monitoring the tropical environments. *Tropical Ecology*, 51 (1):3–10. 3, 16, 29, 58
- Kuntz, S., von Poncet, F., Baldauf, T., Plugge, D., Kenter, B., and Khl, M. (2011). A multi-stage inventory scheme for REDD inventories in tropical countries. In *Proceedings of 34th International Symposium for Remote Sensing of the Environment*. 117
- Laborte, A. G., Maunahan, A. A., and Hijmans, R. J. (2010). Spectral Signature Generalization and Expansion Can Improve the Accuracy of Satellite Image Classification. *Plos ONE*, 5(5). 123
- Liang, S. (2004). *Quantitative Remote Sensing of Land Surfaces*. Wiley. ISBN: 978-0-471-28166-5. 33, 34
- Liao, M., Balz, T., Zhang, L., Pei, Y., and Jiang, H. (2009). Analyzing TerraSAR-X and COSMO-SkyMed high-resolution SAR data of urban areas. In *ISPRS Hannover Workshop 2009 High-Resolution Earth Imaging for Geospatial Information*. 28, 115
- Lillesand, T., Kiefer, R., and Chipman, J. (2008). *Remote Sensing and Image Interpretation*. Wiley, 6 edition. 19
- Lorentz, M., Varjo, J., and Bahamondez, C. (2005). Part III Cross Cutting Issues In Sustainable Forest Management. In *Forests in the Global Balance-Changing Paradigms*. IUFRO World Series Volume 17, IUFRO, Helsinki. Mery, Gerardo and Alfaro, Ren and Kanninen, Markku and Lobovikov, Maxim (eds). 318 p. Access 05.03.2011. 1, 18

- Lu, D., Mausel, P., Brondizio, E., and Moran, E. (2002). Assessment of atmospheric correction methods for Landsat TM data applicable to Amazon basin LBA research. *int. j. remote sensing*, 23(13):2651–2671. [25](#)
- Lu, J., Sun, G., McNulty, S. G., and Amatya, D. M. (2005). A Comparison of Six Potential Evapotranspiration Methods for Regional Use in the Southeastern United States. *Journal of the American Water Resources Association*, 03175:621–633. [154](#)
- Maechler, M. (2012). *Package "cluster"*. Version 1.14.2. [52](#), [54](#), [55](#)
- Maindonald, J. H. (2004). Using R for Data Analysis and Graphics: Introduction, Code and Commentary. Technical report, Centre for Bioinformation Science, Australian National University. [34](#)
- Matthew, M. W., Adler-Golden, Steven, M., Berk, A., Richtsmeiera, S. C., Levinea, R. Y., Bernsteina, L. S., Achary, P. K., Gail, P., Anderson, G., Felde, W., Hoke, M. P., Ratkowski, A., Burke, H.-H., Kaiser, R. D., and Miller, D. P. (2000). Status of atmospheric correction using a MODTRAN4-based algorithm. *SPIE Proceeding*. [25](#), [26](#)
- McDonald, K. C., Zimmermann, R., and Kimball, J. S. (2002). Diurnal and Spatial Variation of Xylem Dielectric Constant in Norway Spruce (*Picea abies* [L.] Karst.) as Related to Microclimate, Xylem Sap Flow, and Xylem Chemistry. *IEEE Transaction on Geoscience and Remote Sensing*, 4(9):2063–2082. [31](#)
- Mernyi, E. (1998). Self-Organizing ANNs for Planetary Surface Composition Research. In *Proceedings of the 6th European Symposium on Artificial Neural Networks, ESANN98*, pages 197–202, Bruges-Belgium. [20](#)
- Mernyi, E. (1999). The Challenges in Spectral Image. Analysis: an Introduction, and Review of ANN Approaches. In *Proceedings of th 7th European Symposium of Artificial Neural Networks, ESANN99*, pages 93–98, Bruges-Belgium. [20](#), [32](#)
- Musa, M. and Hussin, Y. A. (2000). Multi-data Fusion for Sustainable Forest Management: A Case Study From Northern Part of Selangor, Malasia. *International Archives of Photogrammetry and Remote Sensing*, 33. [34](#)

- National Geographic Institute of Costa Rica (1979). Topographical map Rincon (3542 III). Scale: 1:50.000. Second edition. [63](#)
- National Geographic Institute of Costa Rica (1980). Topographical map Changuena (3542 IV). Scale: 1:50.000. Second edition. [62](#)
- Neteler, M. and Mitasova, H. (2008). *Open Source GIS: A GRASS GIS Approach*. Springer, third edition edition. eBook ISBN: 1-4020-8065-4 Print ISBN: 1-4020-8064-6. [152](#), [153](#)
- Ocean Optics, I. W. H. (2006). *USB650 Fiber Optic Spectrometer Installation and Operation Manual Document Number 170-00000-RT-02-1106*. Ocean Optics, Inc. World Headquarters. [45](#), [46](#), [47](#)
- Ollinger, S. V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist*, 189(2):375–394. [14](#), [15](#), [16](#), [22](#)
- Oren, R., Ewers, B., Todd, P., Phillips, N., and Katul, G. (1998). Water Balance Delineate the Soil Layer in which Moisture Affects Canopy Conductance. *Ecological Applications*, 8(4):990-1002. [31](#)
- OTB Development Team (2012). Orfeo Tool Box version OTB-3.11. On line. <http://orfeo-toolbox.org/otb/download.html>. [65](#)
- Owe, M., de Jeu, R., and Walker, J. (2001). A Methodology for Surface Soil Moisture and Vegetation Optical Depth Retrieval Using the Microwave Polarization Difference Index. *IEEE Transaction on Geoscience and Remote Sensing*, 39 (8):1643–1654. [119](#)
- Park, S. E., Kimb, D., Lee, H., Moon, W. M., and Wagner, W. (2010). Tidal Wetland Monitoring Using Polarimetric Synthetic Aperture Radar. In *ISPRS TC VII Symposium 100 Years ISPRS, Vienna, Austria, July 57,*, volume XXXVIII, Part 7A. [30](#)
- PASCO (N.d). *Using Ocean Optics Spectrometers with the PASCO Xplorer GLX Instruction Manual 012-09731B*. PASCO, instruction manual 012-09731b edition. [45](#)

- Pohl, C. and van Genderen, J. L. (1998). Multisensor image fusion in remote sensing: concepts, methods and applications. *int. j. remote sensing*, 19(5):823–854. [18](#), [33](#)
- Price, J. (1994). How unique are spectral signatures? *Remote Sensing of Environment*, 49:181–186. [16](#), [123](#)
- Pu, R. and Gong, P. (2004). Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping. *Remote Sensing of Environment*, 91:212–224. [33](#)
- Quantum GIS Development Team (2012). *Quantum GIS Geographic Information System. Open Source Geospatial Foundation Project*. Open Source Geospatial Foundation, <http://qgis.osgeo.org>. [44](#)
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. [51](#), [52](#), [53](#), [54](#), [56](#), [67](#), [152](#)
- Rao, N. R. (2008). Development of a cropspecific spectral library and discrimination of various agricultural crop varieties using hyperspectral imagery. *International Journal of Remote Sensing*, 29(1):131–144. [111](#), [123](#)
- Reigber, A., Erten, E., Guillaso, S., and Hellwich, O. (2007). I.D.I.O.T.: A Free and Easy to Use Software Tool for DINSAR Analyses. In *Envisat Symposium 2007 2007*, Montreux, Switzerland. [117](#)
- Reimann, C., Filzmoser, P., and Dutter, R. (2008). *Statistical Data Analysis Explained Applied Environmental Statistics with R*. John Wiley and Sons, Ltd. John Wiley & Sons Ltd. [37](#), [51](#), [97](#)
- Reynolds, A. P., Richards, G., De La Iglesia, B., and Rayward-Smith, V. J. (2006). Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms. *Journal of Mathematical Modelling and Algorithms*, 5:475504. [54](#)
- Rezaee, M. R., Lelieveldt, B., and Reiber, J. (1998). A new cluster validity index for the fuzzy c-mean. *Pattern Recognition Letters*, 19:237–246. [36](#), [37](#)

- Richards, J. A. (2009). *Remote Sensing with Imaging Radar*. Signals and Communication Technology. Springer-Verlag. Springer. [26](#), [27](#), [29](#), [61](#)
- Riegler, G., Teufel, F., Lang, O., Petrat, L., Koppe, W., and Herrmann, J. (2011). TerraSAR-X, TanDEM-X, TerraSAR-X2 and their applications. In *3rd International Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*. [124](#)
- Roberts, D., Dennison, P., Ustin, S., Reith, E., and Morais, M. (1999). Development of a regionally-specific library for the Santa Monica Mountains using high-resolution AVIRIS data. In *AVIRIS Workshop*. [20](#)
- Robila, S. A. (2005). Using spectral distances for speedup in hyperspectral image processing. *International Journal of Remote Sensing*, 26(24):5629–5650. [32](#)
- Rohlf, F. J. (1982). Consensus indices for comparing classifications. *Mathematical Biosciences*, 59(1):131–144. [53](#)
- Roth, A., Eineder, M., and Schttler, B. (2007). TerraSAR-X: A New Perspective for Application Requiring High Resolution Spaceborne SAR Data. In *Geoscience and Remote Sensing Symposium, 2006*. [58](#), [117](#)
- Roth, A., Huber, M., and Kosmann, D. (2004). Geocoding of TerraSAR-X Data. In *Proc. 20th ISPRS Congress, Istanbul, Comm. 7*. [58](#)
- Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, 20:53–65. [36](#), [53](#), [54](#), [109](#)
- Sadidya, J., Zeaiean, P., and Firouzabadi, A. E. (2004). The Use of RADARSAT and LANDSAT Image Fusion Algorithms and Different Supervised Classification Methods to Improve Landuse Map Accuracy Case Study: Sari Plain IRAN. In *The 5th International Symposium on Mobile Mapping Technology (MMT 07)*. [34](#)
- Sanchez-Azofeifa, G. A., Castro-Esau, K. L., Kurz, W. A., and Joyce, A. (2009). Monitoring carbon stocks in the tropics and the remote sensing operational

- limitations: from local to regional projects. *Ecological Applications*, 19 (2):480–494. [18](#), [21](#)
- Scheuchl, B., Koudogbo, F., Petrat, L., von Poncet, F., and Gonzalez, A. (2009). TerraSAR-X: Applications for Spaceborne High Resolutio Sar Data. In *Anais XIV Simpsio Brasileiro de Sensoriamento Remoto, Natal, Brasil, 2009, INPE*, pages 7457–7464. [58](#), [117](#)
- Schmidt, K. S. and Skidmore, A. K. (2001). Exploring spectral discrimination of grass species in African rangelands. *Int. J. Remote Sensing*, 22(17):342–3434. [22](#)
- Scholten, J., Klein, M., Steemers, T., and Bruin, G. (2005). Hyperspectral imaging a novel nondestructive analytical tool in paper and writing durability research. In Parisi, C., Buzzanca, G., and Paradisi, A., editors, *Proceedings of Art05 8th International Conference on Non Destructive Investigations and Microanalysis for the Diagnostics and Conservation of the Cultural and Environmental Heritage*, Lecce, Italy. [32](#)
- Shi, Z. and Fung, K. (1994). A comparison of digital speckle filters. *IEEE Xplore*, 4:2129 –2133 vol.4. [45](#), [114](#)
- Shimodaira, H. (2002). An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology*, 51(3):492–508. [36](#), [56](#)
- Shisanya, C. A., Recha, C., and Anyamba, A. (2011). Rainfall Variability and Its Impact on Normalized Difference Vegetation Index in Arid and Semi-Arid Lands of Kenya. *International Journal of Geosciences*. [17](#), [121](#)
- Singh, D. (2006). Scatterometer performance with polarization discrimination ratio approach to retrieve crop soybean parameter at X-band. *International Journal of Remote Sensing*, Vol. 27, No. 19,:41014115. [64](#)
- Smith, L. I. (2002). *A tutorial on Principal Components Analysis*. [20](#), [33](#)
- Sokal, R. R. and Rohlf, F. J. (1962). The Comparison of Dendrograms by Objective Methods. *Taxon*, 11(2):33–40. [53](#)

- Somers, B. and Asner, G. P. (2012). Hyperspectral time series analysis of native and invasive species in hawaiian rainforests. *Remote Sensing*, 4(9):2510–2529. [121](#)
- Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., and Macomber, S. A. (2001). Classification and Change Detection Using Landsat TM Data: When and How to Correct Atmospheric Effects? *Remote Sens. Environ.*, 75:230244. [24](#), [32](#)
- Suzuki, R. and Shimodaira, H. (2004). An application of multiscale bootstrap resampling to hierarchical clustering of microarray data: How accurate are these clusters? *The Fifteenth International Conference on Genome Informatics 2004, P034*. [56](#), [57](#)
- Suzuki, R. and Shimodaira, H. (2011). *pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling*. R package version 1.2-2. [56](#)
- Tansey, K. J., Luckman, A. J., Skinner, L., Balzter, H., Strozzi, T., and Wagner, W. (2004). Classification of forest volume resources using ers tandem coherence and jers backscatter data. *International Journal of Remote Sensing*, 25(4):751–768. [12](#)
- Theseira, M. A., Thomas, G., Taylor, J. C., Gemmell, F., and Varjo, J. (2003). Sensitivity of mixture modelling to endmember selection. *ternational Journal of Remote Sensing*, 24(7):1559 1575. [32](#)
- Torrecilla, E., Stramski, D., Reynolds, R. A., Milln-Nez, E., and Piera, J. (2011 In press). Cluster analysis of hyperspectral optical data for discriminating phytoplankton 6 pigment assemblages in the open ocean. *Remote Sensing of Environment*, doi:10.1016/j.rse.2011.05.014. [52](#), [54](#)
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L’Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., and Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120(0):9–24. [125](#)

- Touzi, R., Boerner, W., Lee, J., and Lueneburg, E. (2004). A review of polarimetry in the context of synthetic aperture radar: concepts and information extraction. *Can. J. Remote Sensing*, 30(3):380–407. [114](#)
- Tsakiris, G., Pangalou, D., and Vangelis, H. (2007). Regional drought assessment based on the Reconnaissance Drought Index (RDI). *Water Resour. Manage.*, 21:821–833. [120](#), [151](#), [154](#), [155](#)
- Tso, B. and Mather, P. M. (2009). *Classification Methods for Remotely Sensed Data*. CRC press, second edition. [24](#), [25](#), [31](#)
- Tuceryan, M. and Jain, A. K. (1988). *The Handbook of Pattern Recognition and Computer Vision*. World Scientific Publishing Co, 2nd edition. C. H. Chen, L. F. Pau, P. S. P. Wang (eds). [31](#), [32](#)
- UNESCO (2008). *Internationally Developed Data Analysis and Management Software Package (IDAMS). WinIDAMS 1.3 Reference Manual in PDF format*. UNESCO. [55](#), [56](#)
- Union of Concerned Scientist (UCS) (2012). UCS Satellite Database 1-1-12.xls. Technical report, Union of Concerned Scientist. [18](#), [27](#), [28](#), [33](#), [125](#)
- U.S Geological Survey (2010). Spectral Characteristics Viewer. Web.11-01-2012. Page Last Modified: 12-30-2010. [23](#)
- U.S. Geological Survey (2006). Hyperion Level 1GST (L1GST) Product Output Files Data Format Control Book (DFCB) Earth Observing-1 (EO-1) version 1.0. Technical report, Department of the Interior, U.S. Geological Survey. [42](#)
- USGS- U.S. Geological Survey (2006). Shuttle Radar Topography Mission. Publisher: Global Land Cover Facility, University of Maryland. 1 Arc Second scenes. Version 2.0. [63](#)
- Ustin, S., DiPietro, D., Olmstead, K., Underwood, E., and Scheer, G. (2002). Hyperspectral remote sensing for invasive species detection and mapping. *In Geoscience and Remote Sensing Symposium. IGARSS 02. 2002 IEEE International.*, Vol(3) ISBN: 0-7803-7536-X:1658–1660. [15](#)

- Ustin, S. L., Roberts, D. A., Gamon, J. A., Asner, G. P., and Green, R. O. (2004). Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. *BioScience*, 54 No. 6:523–534. [14](#)
- Ustin, S. L. and Santos, M. J. (2010). Spectral Identification of Native and Non-Native Plant Species. *Proceedings of ASD and IEEE GRS; Art, Science and Applications of Reflectance Spectroscopy Symposium*. [15](#)
- van Etten, R. J. H. . J. (2011). *raster: Geographic analysis and modeling with raster data*. R package version 1.9-27. [152](#)
- van Genderen, J., Lock, B., and VassP.A. (1978). RemoteSensing: Statistical Testing of Thematic Map Accuracy. *Remote Sensing of Environment*, (1):314. [118](#)
- van Gool, L., Dewaelea, P., and Oosterlinck, A. (1985). Survey Texture analysis Anno 1983. *Computer Vision, Graphics, and Image Processing*, 29 (3):336–357. [31](#)
- Vega-Araya, M., Fuchs, H., Aberle, H., and Kleinn, C. (2012). Modelling climate and vegetation global data to a regional scale, the case of Costa Rica, Central America. In Fehrman, L. and Kleinn, C., editors, *Proceedings of the second International DAAD Workshop, 1.-7. December 2012, Pietermaritzburg and Durban, South Africa*, pages 137–147. Cuvallier Verlag. [6](#), [7](#)
- Vermote, E., Tanr, D., Deuz, J. L., Herman, M., Morcrette, J. J., and Kotchenova, S. Y. (2006). *Second Simulation of a Satellite Signal in the Solar Spectrum - Vector (6SV) par 1*, version 3 edition. [24](#)
- Vidal, C., Lanz, A., Tomppo, Erkki Schadauer, K., Gschwantner, T., di Cosmo, L., and Robert, N. (2008). Establishing forest inventory reference definitions for forest and growing stock: a study towards common reporting. *Silva Fennica*, 42(2):247–266. [13](#), [14](#)
- Waggoner, P. (2009). Forest inventories: Discrepancies and uncertainties. Discussion paper 09-29. *Washington, DC: Resources for the Future*. [13](#)

- Wang, C., Qi, J., and Cochrane, M. (2005). Assessment of Tropical Forest Degradation with Canopy Fractional Cover from Landsat ETM+ and IKONOS Imagery. *Earth Interactions*, 9(22):1–18. [121](#)
- Weishou, S., Di, J., Hui, Z., Shouguang, Y., Haidong, L., and Naifeng, L. (2011). The Response Relation between Climate Change and NDVI over the Qinghai-Tibet plateau. *World Academy of Science, Engineering and Technology*, 59:2216–2222. [17](#)
- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York. [67](#)
- Wiebe, J. A. (1998). Texture Estimates Of Operational Forestry Parameters. Master’s thesis, Department of Geomatics Engineering: Calgary, Alberta University. [31](#)
- Zaitchik, B., Evans, J., Geerken, R., and Smith, R. (2007). Climate and vegetation in the Middle East: Interannual variability and drought feedbacks. *Journal of Climate*, 20:3924 – 3941. [17](#)
- Zhang, X., Jiao, L., Liu, Fang Bo, L., and Gong, M. (2008). Spectral clustering ensemble applied to sar image segmentation. *IEEE Transaction on Geoscience and Remote Sensing*, 46(7):2126–2136. [65](#)
- Zhao, M. and Running, S. W. (2008). *Remote Sensing of Terrestrial Primary Production and Carbon Cycle in Advances in Land Remote Sensing*. Springer. S. Liang (ed) Chapter 16. [17](#), [121](#)

Appendix A

Climate and Vegetation in Costa Rica

A.1 Normalized Difference Vegetation Index

I used climate variables like precipitation and potential evapo-transpiration as well as the Reconnaissance Drought Index (RDI) which is also comparable with the *FAO's* aridity index [Tsakiris et al., 2007]. To correlate the *NDVI* data, we combined several techniques like interpolation and linear regressions in order to compute Costa Rica's surface in pixels of mean monthly values.

We extracted subsets of downloaded SPOT-VEGETATION D10 *NDVI* tiles using VGTEExtract 1.4.2. Three composites per month were used, on the first day of the month, the 11th and on the 21st day. The *NDVI* is derived from spectral (in reflectance, ρ) data as shown in Equation 1:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (A.1)$$

Where *RED* is the red portion of the electromagnetic spectrum (0.6-0.7 ηm) and *NIR* is the near infrared portion of the electromagnetic spectrum (0.75-1.5 ηm). Then we imported the imagery into R [R Development Core Team, 2011] with the *raster* package [van Etten, 2011].

High cloud cover occurs frequently in Central America leading to large regions with missing optical sensor observations even in the *NDVI* 10 days synthesis products. Therefore, it was necessary to estimate the missing sensor observations using a spatio-temporal imputation method using the Expectation Maximization (EM) algorithm of the *Amelia* R package [Honaker et al., 2011]. The average monthly values of *NDVI* for the period from 2001 to 2010 were calculated based on these filled data.

A.2 Precipitation

Spatially irregularly distributed monthly means of meteorological stations were interpolated to a regular grid using the Regularized Spline with Tension (RST) interpolation method implemented as *v.surf.rst* in GRASS [GRASS Development Team, 2011]. A spline is a type of piecewise polynomial, which is preferable to a simple polynomial interpolation [Hengl, 2009]. In our study we specifically defined the amount of smoothness semi-norm, a tension parameter which controls the behaviour of the resulting surface from a thin membrane to a stiff steel plate and segment processing [Neteler and Mitasova, 2008]: as the precipitation is highly variable in time and space we adjusted the smoothing, tension and the number of segments to the spline interpolation process for each month in order to control the variance of the residuals over the data set. Hereby, we wish to avoid overshoots. The segmentation process in the *RST* uses a decomposition of the studied region into rectangular segments with variable size, dependent on the density of the data points [Neteler and Mitasova, 2008]. The use of the segmented processing was the key factor in the present application. In equation 2 we present the general idea of a regularized spline with tension and smoothing [Neteler and Mitasova,

2008].

$$z(r) = T(r) + \sum_{i=1}^N \lambda_j \times R(r, r^{[j]}) \quad (\text{A.2})$$

Where z is the studied phenomenon (in this case precipitation), (r) is an un-sampled location (x, y) , λ is a coefficient and the trend function $T(r)$ is described in the next equation:

$$T(r) = \sum_{i=1}^M a_i f_i(r) \quad (\text{A.3})$$

Where $f_i(r)$ is a set of linearly independent monomial functions and $R(r, r^{[j]})$ is a radial basis function with an explicit form which depends on the choice of weights for partial derivatives in the smooth semi-norm [Neteler and Mitasova, 2008]. A complete description of the variables or equations can also be found in Hengl [2009].

A.3 Temperature

We used a linear regression model for each climatic region and for every month based on the variation of the temperature and elevation from the New LocClim data base. These parameters were then applied to the *SRTM* digital elevation model to get the mean air temperature of Costa Rica.

$$T_m = \beta_0 - \beta_1 \times Elev \quad (\text{A.4})$$

Where T_m is the mean monthly air temperature in °C; $Elev$ is the elevation in meters. B_0 and B_1 are the linear parameters that represent the intercept and the slope, respectively.

A.4 Potential Evapo-transpiration

We calculated the Potential Evapotranspiration (PET) based on Hamon [1963] and the formulae presented in Lu et al. [2005] as follows:

$$PET = 0.1651 \times Ld \times RHOSAT \times kPEC \quad (A.5)$$

Where PET is in mm, Ld is the daytime length, which is the time from sunrise to sunset in multiples of 12 hours; $RHOSAT$ is the saturated vapor density (g/m^3) at the daily mean air temperature (T_m estimated with equation 7.4). $kPEC$ is the calibration coefficient which was 2.2 in the study. Specifically $RHOSAT$ is:

$$RHOSAT = \frac{216.7 \times ESAT}{T_m + 273.3} \quad (A.6)$$

$ESAT$ is the saturated vapour pressure (mb) at the given T_m :

$$ESAT = 6.108 \times \exp\left(\frac{17.26939 \times T_m}{T_m + 273.3}\right) \quad (A.7)$$

We also calculated PET based on four other methods such as Holdridge ([Holdridge, 1962], Thornthwaithe [Lu et al., 2005], Lu-Hamon [Lu et al., 2005]) and the *FAO* Penman-Monteith equation as in New LocClim (version.1.1) [Grieser et al., 2006]. The purpose was to compare the methods and to decide which method and which calibration coefficient to choose for our study.

A.5 Reconnaissance Drought Index

The aridity index used was proposed by Tsakiris et al. [2007] as described below:

$$a_0 = \frac{PRE_i}{PET_i} \quad (A.8)$$

Where PRE is the mean precipitation and PET is the potential evapo-transpiration of the i^{th} month for the analysed period. The second expression is called the Normalized Reconnaissance Drought Index (RDI) [Tsakiris et al., 2007], and the monthly calculation is given by the following equation:

$$RDI_n = \frac{a_0}{\bar{a}_0} - 1 \quad (\text{A.9})$$

Appendix B

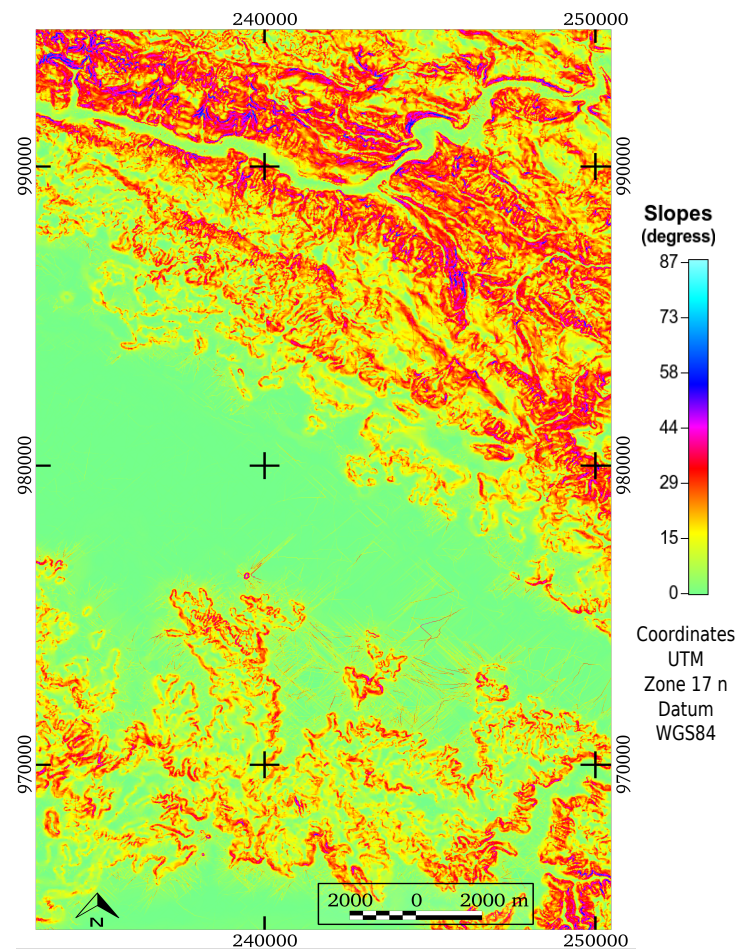


Figure B.1: Slopes of the TerraSAR-X study area.

Appendix C

List of the Clusters and Edges

The appendixes [C.1](#),[C.2](#),[C.3](#),[C.4](#) and [C.5](#) are the text-printed dendrograms for the all hyper-spectral data.

Each of these appendixes presents all the information contained in the dendrograms, however, they contain the edge number of each cluster arrangement. This information is not presented in the Figures [4.8](#),[4.10](#),[4.12](#),[4.14](#) and [4.19](#) because of the space needed to print the edge numbers and the *AU* values in each dendrogram.

C.1 Clusters and Edge Number of the Dendrogram of the *hy_15 PCs* Spectral Signatures.

```
cluster X99.10, X99.11(edge2)
cluster X22.8, X22.9(edge4)
cluster X99.5, X99.6 X99.7, X99.8 (edge5)
cluster X99.1, X99.9(edge7)
cluster X99.5, X99.6, X99.7, X99.8, X99.10, X99.11(edge8)
cluster X99.2, X99.3, X99.12(edge10)
cluster X1001.1, X1001.2(edge11)
cluster X55.8, X55.9(edge13)
```

Clusters and Edge Number of the Dendrogram of the *hy_14* PCs Spectral Signatures. (List of the Clusters and Edges)

cluster X22.3, X22.4(edge18)
cluster X66.3, X66.6 X33.10, X33.11(edge19)
cluster X22.6, X22.7, X22.11(edge22)
cluster X55.3, X55.4(edge28)
cluster X22, X22.6, X22.7, X22.11(edge29)
cluster X1001.7, X1001.8, X1001.9(edge30)
cluster X44, X44.1(edge31)
cluster X44.2 X1001.3, X1001.4(edge32)
cluster X1001.10, X1001.11(edge33)
cluster X22.2, X22.3, X22.4 X44.3(edge34)
cluster X44.4, X44.5, X44.11(edge36)
cluster X66.1, X66.2(edge37)
cluster X99.1, X99.2, X99.3, X99.4, X99.5, X99.6, X99.7, X99.8, X99.9, X99.10, X99.11, X99.12(edge38)
cluster X55.3, X55.4, X55.5(edge42)
cluster X44.6, X44.10(edge50)
cluster X33, X33.1(edge53)
cluster X33.5, X33.7(edge55)
cluster X66, X66.4, X66.5, X66.7, X66.8, X66.9, X66.10, X66.11(edge56)
cluster X44.7, X44.8, X44.9(edge60) X33.6, X33.10, X33.11(edge61)
cluster X33.3, X33.9, X66.1, X66.2, X66.3, X66.6(edge64)
cluster X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11(edge66)
cluster X55, X55.1, X55.2, X55.3, X55.4, X55.5, X55.6, X55.7, X55.8, X55.9, X55.10, X55.11(edge67)
cluster X44, X44.1, X44.2, X44.3, X44.4, X44.5, X44.6, X44.10, X44.11(edge68)
cluster X33.5, X33.7, X66, X66.4, X66.5, X66.7, X66.8, X66.9, X66.10, X66.11(edge69)
cluster X33, X33.1, X33.2(edge71)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X1001.8, X1001.9, X1001.10, X1001.11, X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11(edge75)
cluster X44, X44.1, X44.2, X44.3, X44.4, X44.5, X44.6, X44.7, X44.8, X44.9, X44.10, X44.11(edge76)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X1001.8, X1001.9, X1001.10, X1001.11, X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11, X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X33.10, X33.11, X55, X55.1, X55.2, X55.3, X55.4, X55.5, X55.6, X55.7, X55.8, X55.9, X55.10, X55.11, X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9, X66.10, X66.11(edge81)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X1001.8, X1001.9, X1001.10, X1001.11, X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11, X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X33.10, X33.11, X44, X44.1, X44.2, X44.3, X44.4, X44.5, X44.6, X44.7, X44.8, X44.9, X44.10, X44.11, X55, X55.1, X55.2, X55.3, X55.4, X55.5, X55.6, X55.7, X55.8, X55.9, X55.10, X55.11, X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9, X66.10, X66.11(edge82)

C.2 Clusters and Edge Number of the Dendrogram of the *hy_14* PCs Spectral Signatures.

cluster X77.4, X77.5(edge1)
cluster X77.2, X77.3(edge2)
cluster X99.9, X99.10(edge3)
cluster X99.3, X99.4(edge4)
cluster X88.5, X88.6(edge5)

Clusters and Edge Number of the Dendrogram of the *hy_14* PCs Spectral Signatures. (List of the Clusters and Edges)

cluster X99.1, X99.2, X99.5(edge9)
cluster X99.3, X99.4, X99.9, X99.10, X99.12(edge10)
cluster X99.6, X99.7, X99.8(edge12)
cluster X77.7, X77.8(edge15)
cluster X33.8, X33.9(edge16)
cluster X22.6, X22.7(edge17)
cluster X99.1, X99.2, X99.5, X99.11(edge19)
cluster X88.4, X88.5, X88.6, X88.9, X88.10(edge20)
cluster X22.8, X22.9(edge21)
cluster X22.1, X22.2(edge23)
cluster X22.3, X22.4(edge24)
cluster X33.1, X33.2(edge28)
cluster X33, X33.11(edge31)
cluster X99.1, X99.2, X99.3, X99.4, X99.5, X99.6, X99.7, X99.8, X99.9, X99.10, X99.11, X99.12(edge35)
cluster X44.3, X44.4(edge37) X77.6, X77.7, X77.8(edge40) X44.10, X44.11(edge42)
cluster X88, X88.1, X88.2, X88.3, X88.4, X88.5, X88.6, X88.7, X88.8, X88.9, X88.10, X88.11(edge47)
cluster X77.1, X77.11(edge48) X66, X66.6 (edge49) X66.3, X66.5(edge56)
cluster X77.2, X77.3, X77.4, X77.5, X77.10(edge61) X66, X66.6, X66.9(edge63)
cluster X77, X77.1, X77.11(edge64) X1001.7, X1001.8, X1001.9(edge65)
cluster X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.10, X1001.11(edge67)
cluster X77.2, X77.3, X77.4, X77.5, X77.9, X77.10(edge68) X66, X66.6, X66.7, X66.9(edge69)
cluster X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11(edge70)
cluster X44.7, X44.9(edge71)
cluster X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X33.10, X33.11(edge72)
cluster X44.3, X44.4, X44.5, X44.6, X44.10, X44.11(edge74)
cluster X66, X66.3, X66.5, X66.6, X66.7, X66.9(edge76) X44.7, X44.8, X44.9(edge79)
cluster X77.2, X77.3, X77.4, X77.5, X77.6, X77.7, X77.8, X77.9, X77.10(edge83)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X1001.8, X1001.9, X1001.10, X1001.11(edge84)
cluster X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X33.10, X33.11, X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9, X66.10, X66.11(edge87)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X1001.8, X1001.9, X1001.10, X1001.11, X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11(edge88)
cluster X77, X77.1, X77.2, X77.3, X77.4, X77.5, X77.6, X77.7, X77.8, X77.9, X77.10, X77.11(edge89)
cluster X44, X44.1, X44.2, X44.3, X44.4, X44.5, X44.6, X44.7, X44.8, X44.9, X44.10, X44.11(edge90)
cluster X88, X88.1, X88.2, X88.3, X88.4, X88.5, X88.6, X88.7, X88.8, X88.9, X88.10, X88.11, X99.1, X99.2, X99.3, X99.4, X99.5, X99.6, X99.7, X99.8, X99.9, X99.10, X99.11, X99.12(edge91)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X1001.8, X1001.9, X1001.10, X1001.11, X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11, X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X33.10, X33.11, X44, X44.1, X44.2, X44.3, X44.4, X44.5, X44.6, X44.7, X44.8, X44.9, X44.10, X44.11, X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9, X66.10, X66.11, X88, X88.1, X88.2, X88.3, X88.4, X88.5, X88.6, X88.7, X88.8, X88.9, X88.10, X88.11, X99.1, X99.2, X99.3, X99.4, X99.5, X99.6, X99.7, X99.8, X99.9, X99.10, X99.11, X99.12(edge94)

C.3 Clusters and Edge Number of the dendrogram of the *eo1_pz PCs* Spectral Signatures.

cluster X88.4, X88.5(edge2) X1001, X1001.1(edge3)
cluster X66.5, X66.6(edge7)
cluster X55.4, X55.5(edge8)
cluster X66.3, X66.4(edge10)
cluster X44.4, X44.5(edge11)
cluster X99, X99.1, X99.2, X99.3, X99.4, X99.5(edge13)
cluster X55.1, X55.2, X55.3(edge15)
cluster X33.1, X33.2(edge16)
cluster X22, X22.2(edge17)
cluster X44, X44.1, X44.2(edge18)
cluster X22.10, X22.11(edge21)
cluster X44.3, X44.4, X44.5(edge22)
cluster X66, X66.1(edge26)
cluster X1001, X1001.1, X1001.2(edge27)
cluster X22, X22.2, X22.8(edge29)
cluster X33.4, X33.5(edge30)
cluster X33.9, X33.10(edge32)
cluster X22, X22.2, X22.8, X22.9(edge34)
cluster X22.3, X22.5(edge36)
cluster X44, X44.1, X44.2, X44.3, X44.4, X44.5(edge39)
cluster X88, X88.1, X88.2, X88.3, X88.4, X88.5, X88.6, X88.7(edge41)
cluster X1001.5, X1001.6(edge45)
cluster X55, X55.1, X55.2, X55.3, X55.4, X55.5, X55.6, X55.7(edge48)
cluster X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9(edge58)
cluster X77, X77.1, X77.2, X77.3(edge65)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X22, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11, X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X33.10, X33.11(edge66)
cluster X88, X88.1, X88.2, X88.3, X88.4, X88.5, X88.6, X88.7, X99, X99.1, X99.2, X99.3, X99.4, X99.5(edge67)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X22, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11, X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X33.10, X33.11, X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9(edge68)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X22, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11, X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X33.10, X33.11, X55, X55.1, X55.2, X55.3, X55.4, X55.5, X55.6, X55.7, X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9(edge69)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X22, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11, X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X33.10, X33.11, X44, X44.1, X44.2, X44.3, X44.4, X44.5, X55, X55.1, X55.2, X55.3, X55.4, X55.5, X55.6, X55.7, X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9(edge70)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X22, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X22.10, X22.11, X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X33.10, X33.11, X44, X44.1, X44.2, X44.3, X44.4, X44.5, X55, X55.1, X55.2, X55.3, X55.4, X55.5, X55.6, X55.7, X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9, X88, X88.1, X88.2, X88.3, X88.4, X88.5, X88.6, X88.7, X99, X99.1, X99.2, X99.3, X99.4, X99.5(edge71)

C.4 Clusters and Edge Number of the Dendrogram of the *eo1_kw PCs* Spectral Signatures.

cluster X22.1, X22.2(edge1)
cluster X99.3, X99.4(edge2)
cluster X22.4, X22.5(edge3)
cluster X33.7, X33.8(edge4)
cluster X44, X44.1(edge5)
cluster X33.5, X33.6(edge8)
cluster X33.1, X33.2(edge9)
cluster X44.3, X44.4(edge11)
cluster X66.2, X66.3(edge12)
cluster X88.1, X88.2(edge13)
cluster X99, X99.1, X99.2(edge14)
cluster X66.4, X66.5(edge15)
cluster X1001.1, X1001.2(edge18) X22, X22.1, X22.2(edge19)
cluster X44.6, X44.7(edge20)
cluster X22.7, X22.8, X22.9(edge21)
cluster X66.6, X66.7(edge24)
cluster X55.5, X55.6 X55, X55.1(edge25)
cluster X1001.6, X1001.7(edge28)
cluster X66.8, X66.9(edge33)
cluster X55.2, X55.3(edge39)
cluster X1001.1, X1001.2, X1001.5(edge45)
cluster X44, X44.1, X44.2, X44.3, X44.4, X44.5, X44.8(edge46)
cluster X88, X88.1, X88.2, X88.3, X88.4, X88.5, X88.6, X88.7(edge50)
cluster X55.5, X55.6, X55.7(edge53)
cluster X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9(edge54)
cluster X1001, X1001.3, X1001.4(edge55)
cluster X33, X33.1, X33.2, X33.3, X33.5, X33.6, X33.7, X33.8, X33.9(edge56)
cluster X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9(edge57)
cluster X44, X44.1, X44.2, X44.3, X44.4, X44.5, X44.6, X44.7, X44.8(edge59)
cluster X1001.1, X1001.2, X1001.5, X1001.6, X1001.7(edge60) X77.1, X77.3(edge61)
cluster X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9(edge63)
cluster X1001.1, X1001.2, X1001.5, X1001.6, X1001.7, X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9(edge65)
cluster X44, X44.1, X44.2, X44.3, X44.4, X44.5, X44.6, X44.7, X44.8, X44.9(edge66)
cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9(edge67)
cluster X55, X55.1, X55.2, X55.3, X55.4, X55.5, X55.6, X55.7(edge68)
cluster X99, X99.1, X99.2, X99.3, X99.4, X99.5, X99.6, X99.7, X99.8, X99.9, X99.10(edge69)
cluster X77, X77.1, X77.3(edge70)
cluster X88, X88.1, X88.2, X88.3, X88.4, X88.5, X88.6, X88.7, X99, X99.1, X99.2, X99.3, X99.4, X99.5, X99.6, X99.7, X99.8, X99.9, X99.10(edge71)
cluster X77, X77.1, X77.2, X77.3(edge73)

Clusters and Edge Number of the Dendrogram of the USB650 Spectral Signatures. (List of the Clusters and Edges)

cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9(edge74)

cluster X44, X44.1, X44.2, X44.3, X44.4, X44.5, X44.6, X44.7, X44.8, X44.9, X55, X55.1, X55.2, X55.3, X55.4, X55.5, X55.6, X55.7(edge75)

cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X44, X44.1, X44.2, X44.3, X44.4, X44.5, X44.6, X44.7, X44.8, X44.9, X55, X55.1, X55.2, X55.3, X55.4, X55.5, X55.6, X55.7, X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9(edge76)

cluster X1001, X1001.1, X1001.2, X1001.3, X1001.4, X1001.5, X1001.6, X1001.7, X22, X22.1, X22.2, X22.3, X22.4, X22.5, X22.6, X22.7, X22.8, X22.9, X33, X33.1, X33.2, X33.3, X33.4, X33.5, X33.6, X33.7, X33.8, X33.9, X44, X44.1, X44.2, X44.3, X44.4, X44.5, X44.6, X44.7, X44.8, X44.9, X55, X55.1, X55.2, X55.3, X55.4, X55.5, X55.6, X55.7, X66, X66.1, X66.2, X66.3, X66.4, X66.5, X66.6, X66.7, X66.8, X66.9, X88, X88.1, X88.2, X88.3, X88.4, X88.5, X88.6, X88.7, X99, X99.1, X99.2, X99.3, X99.4, X99.5, X99.6, X99.7, X99.8, X99.9, X99.10(edge77)

C.5 Clusters and Edge Number of the Dendrogram of the USB650 Spectral Signatures.

cluster X103311055, X103311056(edge1)

cluster X302010845, X302010846(edge2)

cluster X302030861, X302030862(edge3)

cluster X302030863, X302030864 (edge4)

cluster X30203086, X103311055, X103311056, X302030861, X302030862, X302030863, X302030864, X302030865 (edge8)

cluster X30203086, X103311054, X103311055, X103311056, X302030861, X302030862, X302030863, X302030864, X302030865(edge9)

cluster X302010841, X302010843 X302010844, X302010845, X302010846(edge11)

cluster X302010841, X302010842, X302010843(edge12)

cluster X103333078, X103333079(edge13)

cluster X101132021, X101132022(edge14)

cluster X30202085, X302020851(edge15)

cluster X50202091, X50203092(edge16)

cluster X40202088, X60206098(edge17)

cluster X302010841, X302010842, X302010843, X302010844, X302010845, X302010846(edge18)

cluster X50202091, X50203092, X60204096(edge19)

cluster X103312057, X103312058(edge20)

cluster X103313059, X103313060 (edge21)

cluster X101131018, X101131019(edge22)

cluster X101122012, X101122013(edge23)

cluster X101132020, X101132021, X101132022(edge24)

cluster X102222039, X102222040(edge26)

cluster X102211026, X102211027(edge29)

cluster X103331070, X103331071(edge30)

cluster X102222038, X102222039, X102222040(edge33)

cluster X30203086, X103311054, X103311055, X103311056, X103332075, X302030861, X302030862, X302030863, X302030864, X302030865(edge35)

cluster X102212029, X102212031(edge36)

cluster X103313059, X103313060, X103313061(edge37)

Clusters and Edge Number of the Dendrogram of the USB650 Spectral Signatures. (List of the Clusters and Edges)

cluster X102232047, X102232048, X102232049(edge39)
cluster X102221035, X102221036(edge40)
cluster X102222038, X102222039, X102222040, X102223041(edge42)
cluster X202030821, X202030822(edge45)
cluster X102211026, X102211027, X102211028(edge46)
cluster X101113007, X102223042(edge47)
cluster X101113024, X101113025(edge48)
cluster X60202094, X102233052(edge49)
cluster X103322065, X103322066(edge52)
cluster X202030823, X202030824(edge53)
cluster X101113008, X102223043(edge54)
cluster X102212030, X102231046(edge55)
cluster X101113006, X102231044(edge59)
cluster X103331070, X103331071, X103331072(edge60)
cluster X101122012, X101122013, X101133023(edge63)
cluster X101111010, X101111011(edge65)
cluster X103323067, X103323068, X103323069(edge66)
cluster X101113007, X102222038, X102222039, X102222040, X102223041, X102223042(edge69)
cluster X102233050, X102233051(edge80)
cluster X60207099, X202030821, X202030822, X202030823, X202030824, X202030825, X202030826, X202030827, X202030828, X202030829(edge81)
cluster X103312057, X103312058, X103313059, X103313060, X103313061(edge84)
cluster X101113006, X102212029, X102212030, X102212031, X102231044, X102231045, X102231046(edge89)
cluster X101123014, X101123015, X101123016, X103331070, X103331071, X103331072(edge90)
cluster X103332073, X103332074(edge97)
cluster X101111009, X101111010, X101111011(edge100)
cluster X101113006, X102212029, X102212030, X102212031, X102231044, X102231045, X102231046, X102232047, X102232048, X102232049(edge104)
cluster X102213032, X102213033, X102213034, X102221035, X102221036, X102233050, X102233051, X103332073, X103332074(edge109)

Appendix D

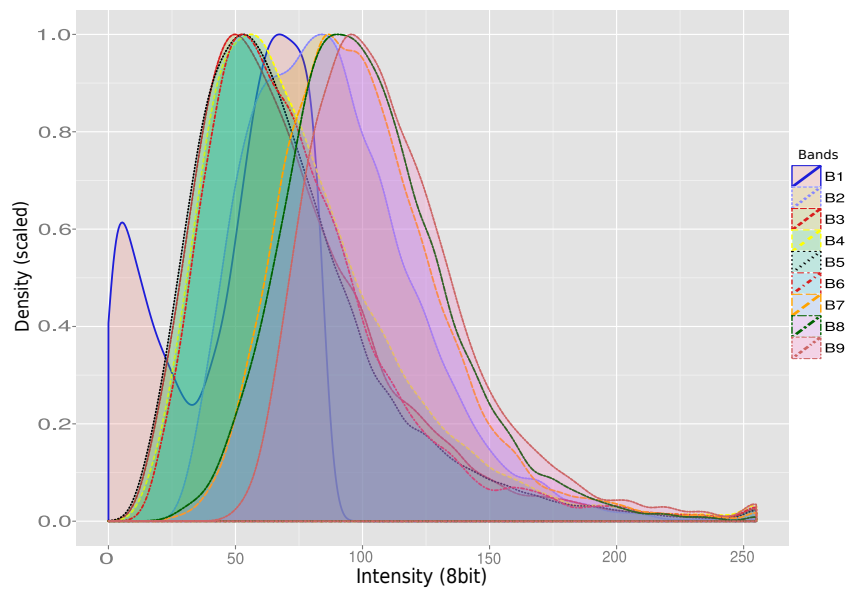


Figure D.1: Conditional density plot all bands for the land cover *forest* in gently undulating slopes.

(Conditional Density Plots for All Bands and the Different Land Covers)

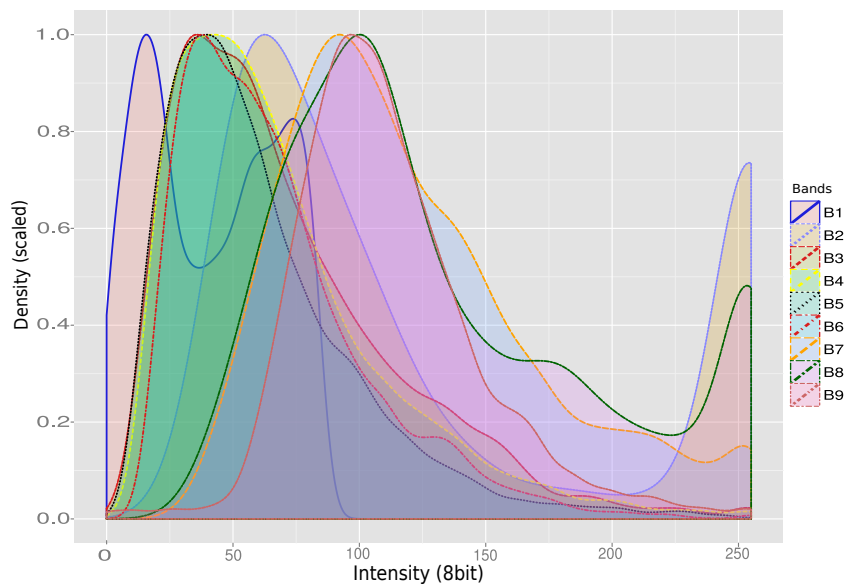


Figure D.2: Conditional density plot for all bands the land cover *forest* in rolling slopes.

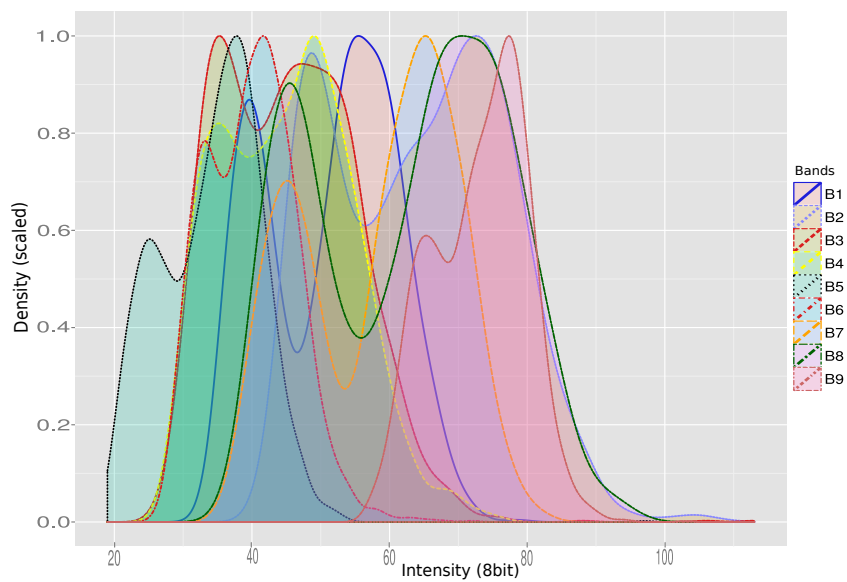


Figure D.3: Conditional density plot for all bands for the land cover *herbland* in flat slopes.

(Conditional Density Plots for All Bands and the Different Land Covers)

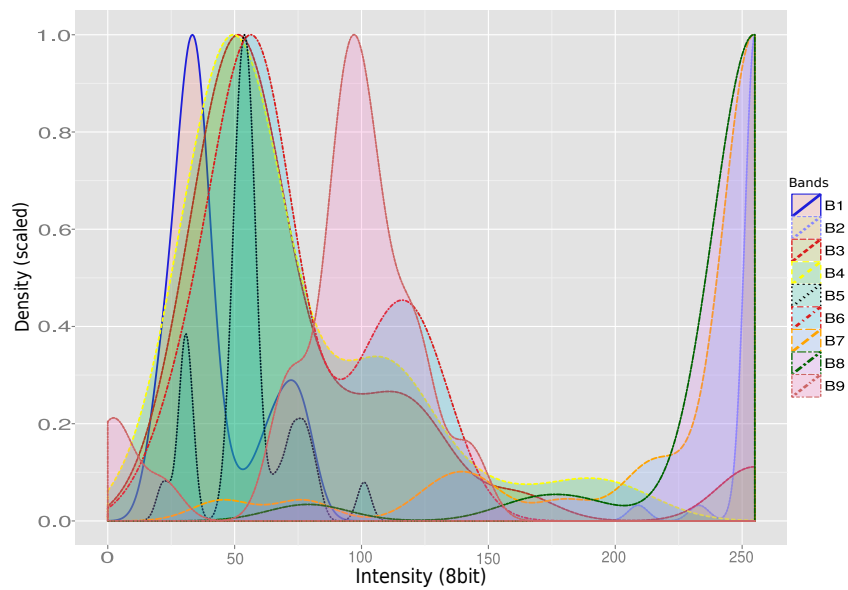


Figure D.4: Conditional density plot for all bands for the land cover *infrastructure*.

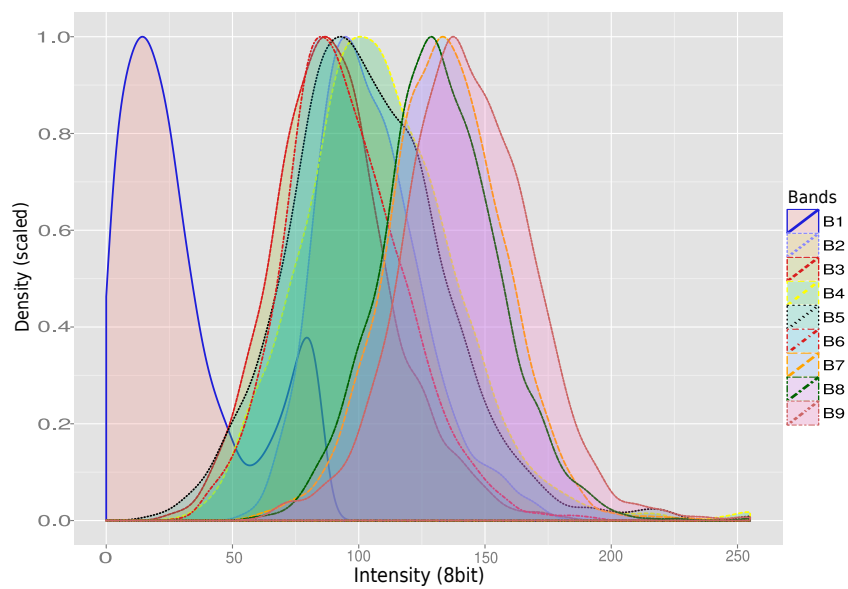


Figure D.5: Conditional density plot for all bands for the land cover *mangrove*.

(Conditional Density Plots for All Bands and the Different Land Covers)

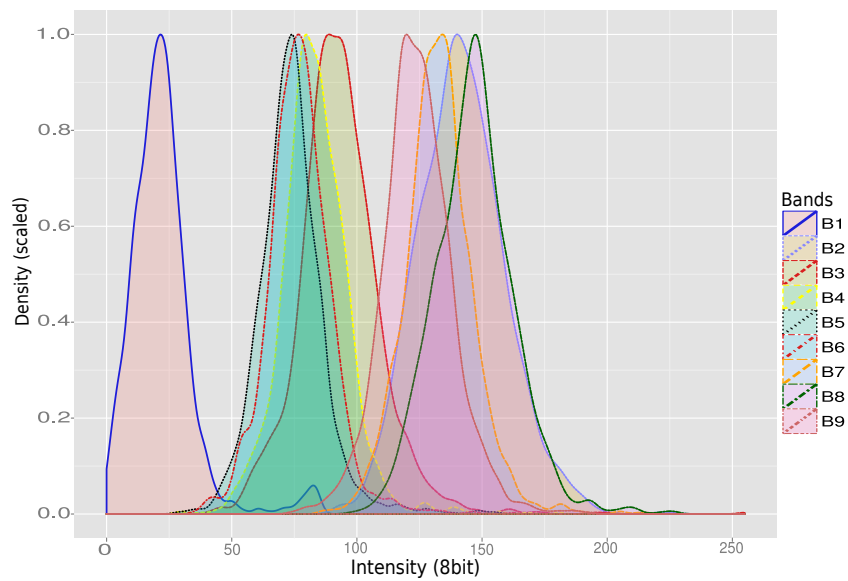


Figure D.6: Conditional density plot for all bands for the land cover *palm oil* (old plantation).

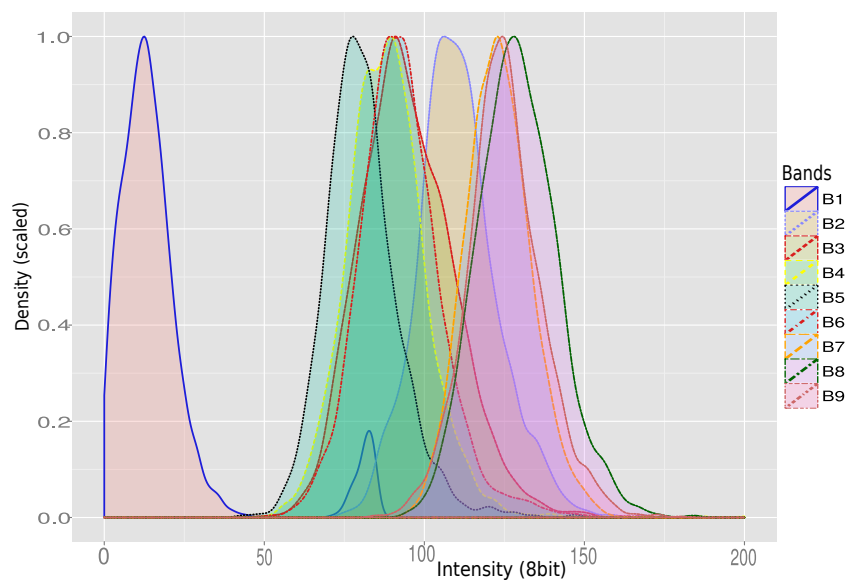


Figure D.7: Conditional density plot for all bands for the land cover *palm oil* (young plantation).

(Conditional Density Plots for All Bands and the Different Land Covers)

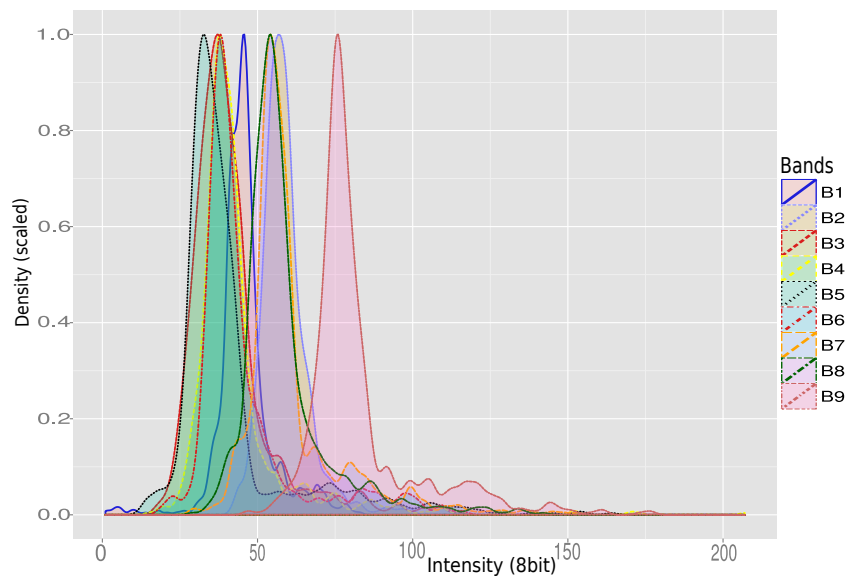


Figure D.8: Conditional density plot for all bands for the land cover *pasture*.

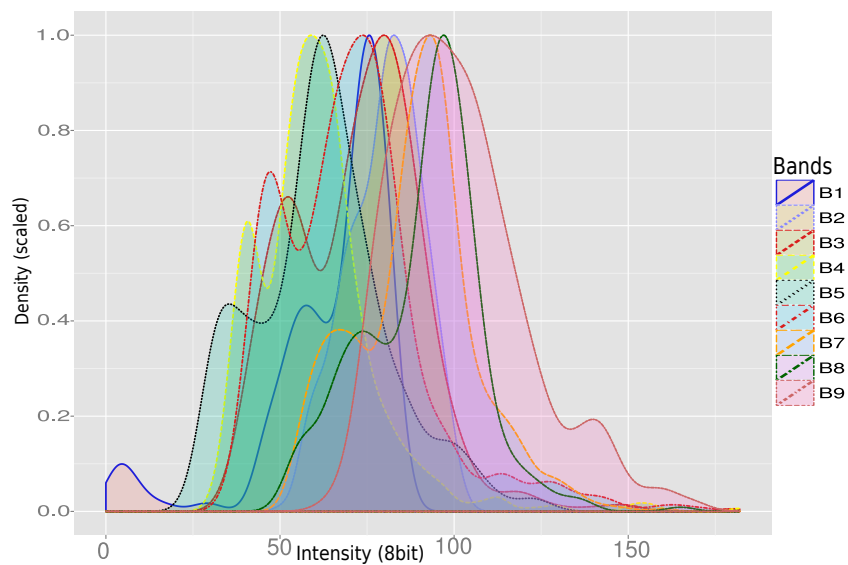


Figure D.9: Conditional density plot for all bands for the land cover *bare soil*.

(Conditional Density Plots for All Bands and the Different Land Covers)

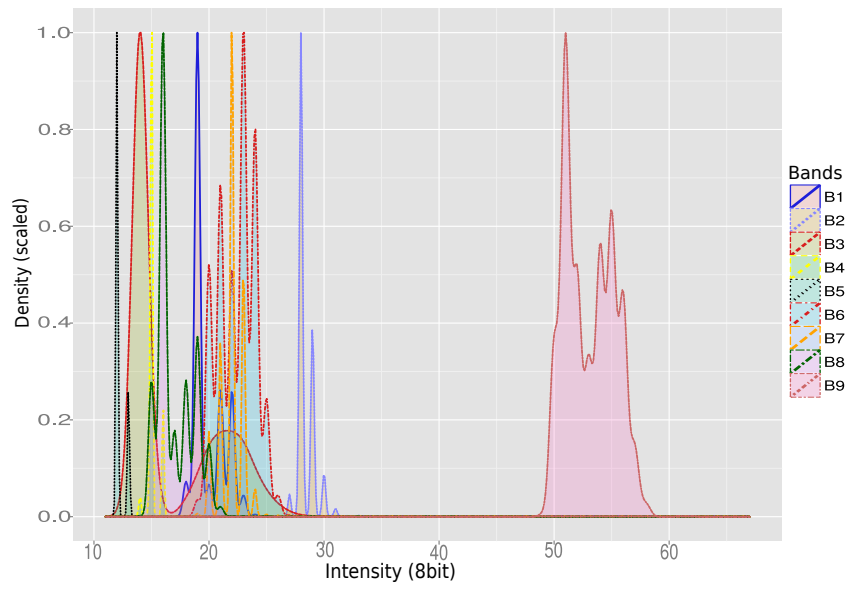


Figure D.10: Conditional density plot for all bands for the land cover *water*.

Appendix E

Table E.1: Comparison of the land covers by TerraSAR-X processed bands.

Cover (ROI)	B1		B2		B3		B4		B5		B6		B7		B8		B9	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
soil	65.01	17.89	79.97	10.56	72.16	18.72	60.18	17.35	62.18	19.21	70.56	20.98	89.18	17.31	90.91	16.62	101.56	19.14
forest_gen_undul	49.23	26.21	88.86	32.47	73.04	38.46	76.91	39.20	71.19	37.61	75.23	37.88	100.97	32.47	102.39	32.03	111.59	32.95
pastures	45.97	9.13	58.77	7.62	41.19	13.67	43.85	17.08	41.51	19.70	45.64	16.58	61.11	16.42	58.75	14.82	82.71	17.90
palm_old	22.26	13.26	141.69	17.28	93.18	16.69	83.25	13.52	73.83	13.09	77.13	12.83	132.80	13.72	146.96	15.99	124.34	14.05
palm_young	17.78	18.42	110.46	12.44	94.71	14.54	88.38	12.55	80.63	11.60	92.69	12.97	123.20	10.21	129.21	12.33	125.16	11.80
infrastructure	41.55	17.75	252.94	8.77	78.58	55.43	73.39	42.88	54.64	16.64	72.27	31.37	226.30	55.87	244.30	35.42	89.94	36.03
herbland	51.20	9.39	64.27	12.47	45.23	9.66	45.10	9.35	33.99	7.15	39.84	6.35	58.44	10.85	62.94	13.86	73.07	6.74
mangrove	29.80	24.44	105.07	20.79	90.05	23.78	108.51	30.50	103.36	31.11	94.97	24.05	134.35	23.29	132.24	23.49	143.16	25.46
water	19.91	1.37	28.48	1.46	16.84	3.92	15.13	0.45	12.23	0.48	22.44	1.58	22.00	1.06	17.04	1.73	53.08	2.13
forest_flat_raffia	18.42	21.61	123.79	15.53	81.63	11.55	81.75	11.08	71.29	10.49	75.40	10.35	119.95	10.25	130.20	13.16	117.90	10.68
forest_slopes	41.48	25.85	126.33	80.53	70.62	44.25	68.55	42.39	61.24	38.43	66.41	36.79	119.96	50.08	132.23	62.11	111.21	35.90

Appendix F

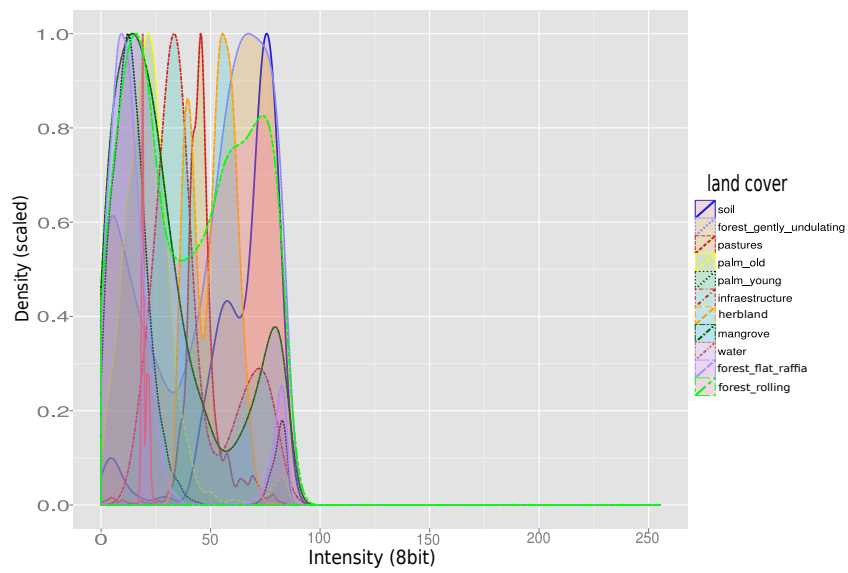


Figure F.1: Conditional density plot for the Band 1 and all land covers.

(Comparison of the Covers by TerraSAR-X Processed Bands)

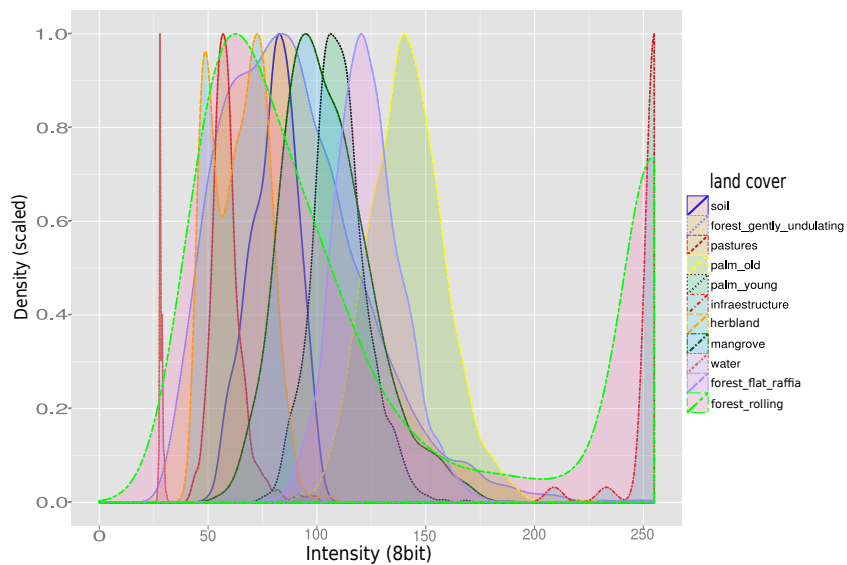


Figure F.2: Conditional density plot for the Band 2 and all land covers.

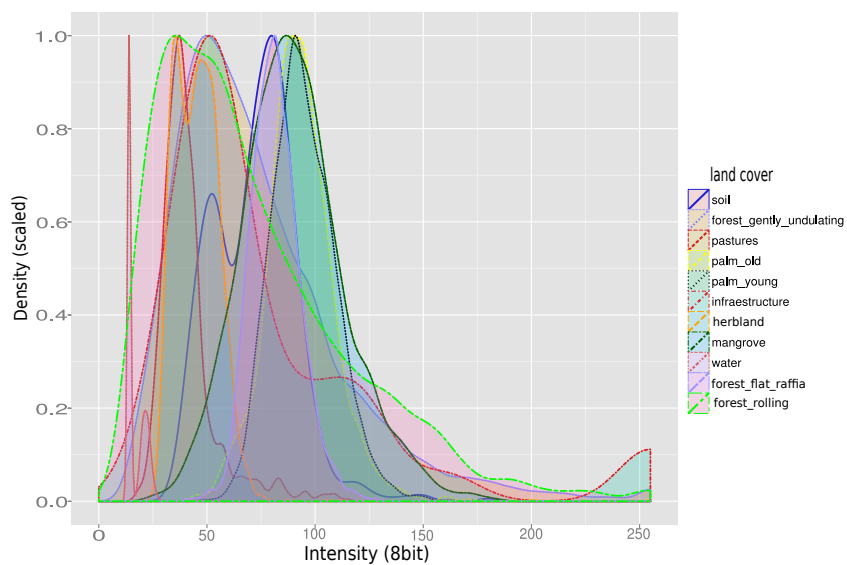


Figure F.3: Conditional density plot for the Band 3 and all land covers.

(Comparison of the Covers by TerraSAR-X Processed Bands)

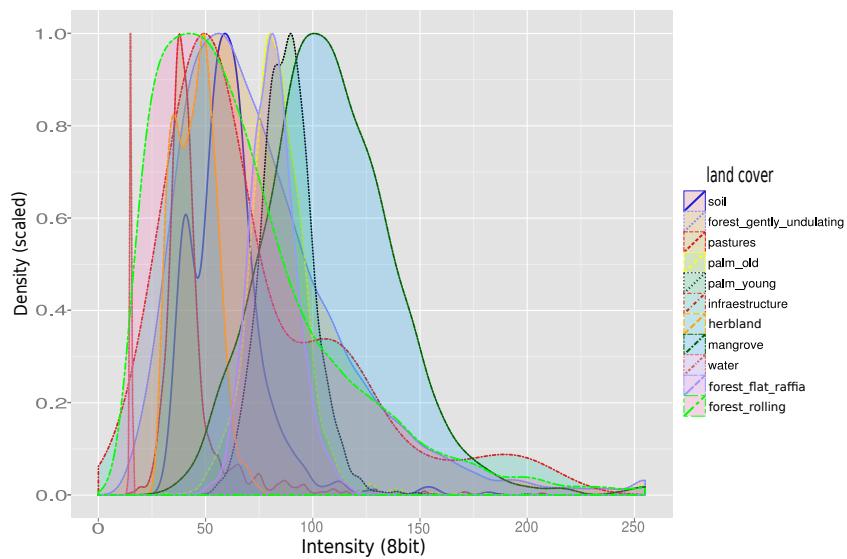


Figure F.4: Conditional density plot for the Band 4 and all land covers.

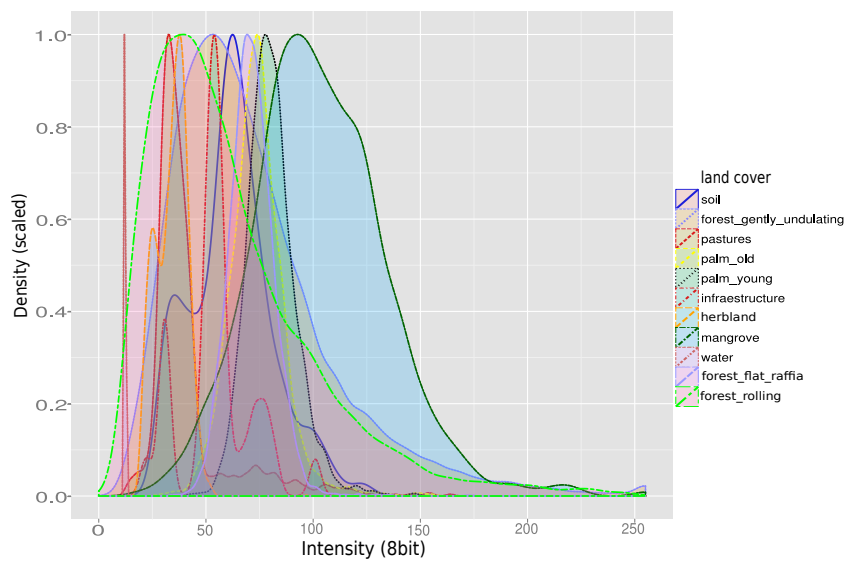


Figure F.5: Conditional density plot for the Band 5 and all land covers.

(Comparison of the Covers by TerraSAR-X Processed Bands)

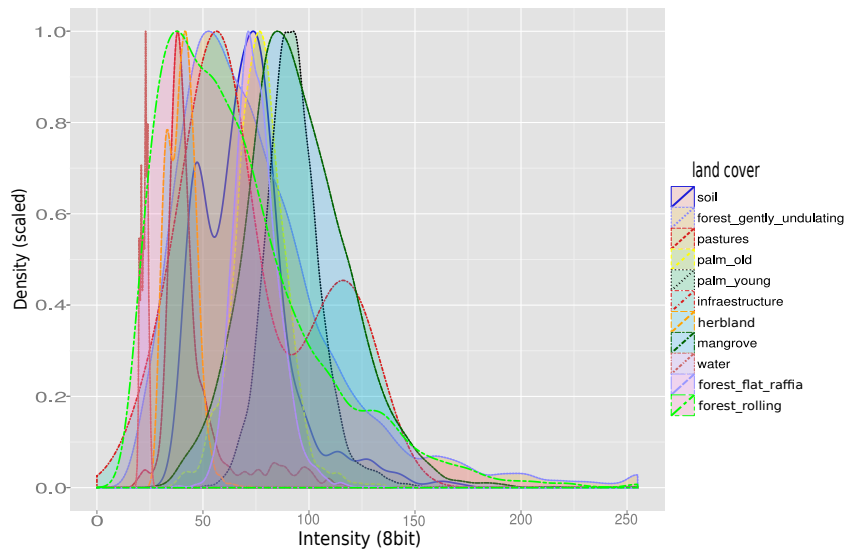


Figure F.6: Conditional density plot for the Band 6 and all land covers.

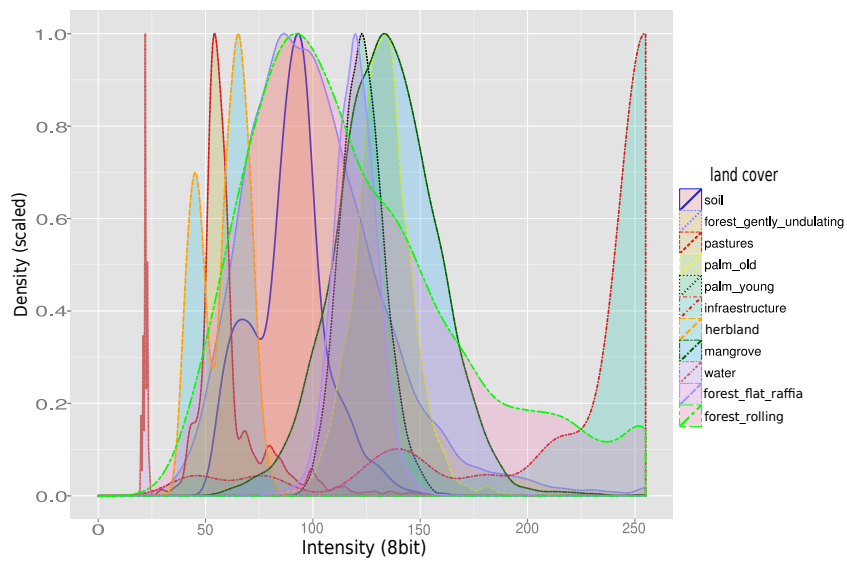


Figure F.7: Conditional density plot for the Band 7 and all land covers.

(Comparison of the Covers by TerraSAR-X Processed Bands)

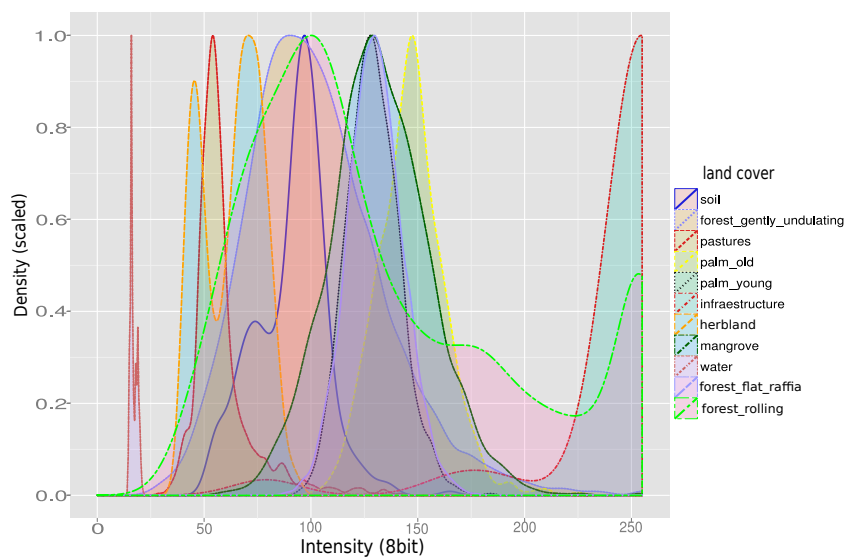


Figure F.8: Conditional density plot for the Band 8 and all land covers.

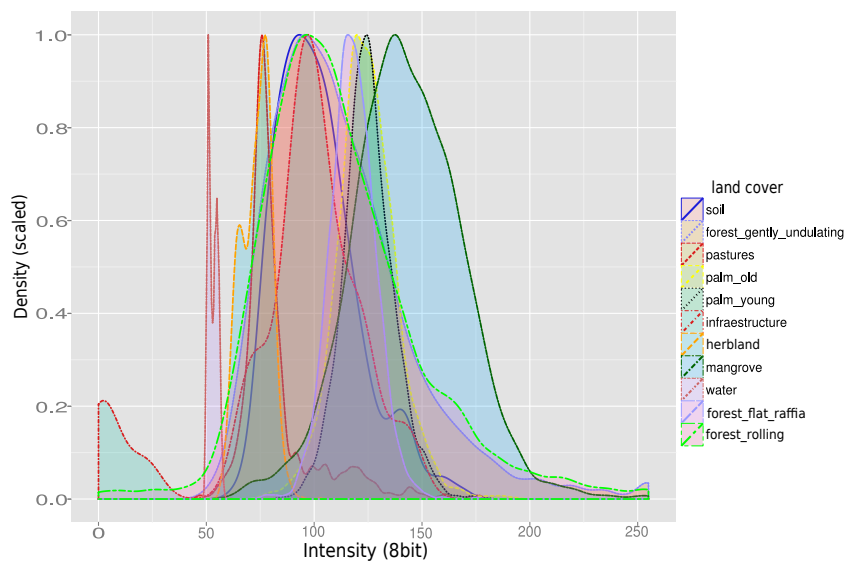


Figure F.9: Conditional density plot for the Band 9 and all land covers.