



UNIVERSITY OF THE PHILIPPINES

Master of Science in Geomatics Engineering

Jose Don T De Alban

Evaluation of multi-year ALOS/PALSAR mosaic data for mapping and monitoring of
forest cover types in the Philippines

Thesis Adviser:

Enrico C Paringit, Dr Eng
Department of Geodetic Engineering
University of the Philippines Diliman

Date of Submission:

January 2017

Thesis Classification:

F

This thesis is available to the public.

This thesis entitled, **EVALUATION OF MULTI-YEAR ALOS/PALSAR MOSAIC DATA FOR MAPPING AND MONITORING OF FOREST COVER TYPES IN THE PHILIPPINES**, prepared and submitted by **JOSE DON T DE ALBAN**, in partial fulfillment of the requirements for the degree of **GEOMATICS ENGINEERING**, major in **REMOTE SENSING** is hereby accepted.

ENRICO C PARINGIT, DR ENG
Thesis Adviser

Accepted as partial fulfillment of the requirements for the degree of **GEOMATICS ENGINEERING**, major in **REMOTE SENSING**.

DR RIZALINDA L DE LEON
Dean

Acknowledgement

This thesis is the product of perseverance and hard work from my long, meandering journey in pursuit of achieving a masters degree, and of wanting to do research that not only aligned with my own interests in applying remote sensing for forest conservation and management, but also of contributing an academic piece of work with a practical purpose. It is my hope that this thesis is found in the hands of individuals who by reading, studying, critiquing, and improving it find value in this work—all to further contribute to the greater body of knowledge—rather than just ending up unread in shelves, gathering dust.

Completing this masters thesis would not have been possible without the support and encouragement of family, friends, and colleagues, especially when my commitment to family and professional work took precedence over finishing my postgraduate studies. I am so grateful to all of you.

To Dr Enrico Paringit, my thesis adviser and professor, for sharing his knowledge and providing guidance in doing my thesis, and for his patience, understanding, persistence, and encouragement all throughout my masters studies.

To Dr Rijaldia Santos and Engr Ma Rosario Concepcion Ang, both of whom were my former professors during my undergraduate and masters studies, respectively.

Thank you for agreeing to be my thesis panel members and providing very helpful comments during my thesis viva and for improving the manuscript.

To this extraordinary group of friends, colleagues, research group mates wherein our professional, academic, and even personal lives have been intertwined: Angelica Kristina Monzon, Rizza Karen Veridiano, Margie Parinas, Sheryl Rose Reyes, Roven Tumaneng, Ray Neil Lorca, Raul Magabo, and Francisco Tavora Jr—this journey was all the more worthwhile because we were all in the thick of things together.

To my former professors, including Dr Ariel Blanco, Dr Rhodora Gonzalez, and Dr Ma Cecilia Rubio-Paringit, who kept encouraging me to finish my thesis at every opportunity that I meet them. This also goes for my boss at my new research work at the National University of Singapore, Dr Edward Webb—thanks for keeping me on my toes.

To close friends and colleagues: Edmund Leo Rico, Ronald Allan Altamirano, and Dr Neil Aldrin Mallari—thank you all for those long discussions, brainstorming sessions, and exchange of ideas of the things that make us look forward to waking up in the morning.

I am also grateful to Mari Trix Estomata, Nimol Vameourn, Christian Supsup, Dennis Tacmo, Dr Edwino Fernando, and Ms Rhudelene Serrano who have all supported me in small ways but nevertheless added up to help me put my thesis together.

This work has been undertaken within the framework of the Kyoto & Carbon Initiative led by Japan Aerospace Exploration Agency. The ALOS/PALSAR data used in this study have been provided by the JAXA Earth Observation Research Center.

I dedicate this academic work to Rina, my wife and best friend, and my twin daughters, Dahon Luntian and Ulap Amihan—you are all my inspiration.

And to God for whom all of this was meant for his greater honor and glory.

In memory of
Danilo S Balete
(1960-2017)
who in 2006 gave me the idea of
mapping the forest types of the Philippines
using geospatial technologies.
Thank you for sowing those seeds, Danny.

Abstract

Forests are important ecosystems that provide a broad range of goods and services, including social and economic benefits in the long term. To sustainably manage forest resources, remote sensing technologies are being employed as an accurate and cost-effective approach for mapping and monitoring forest conditions. In the Philippines, spaceborne remote sensing technology was similarly used in the past for forest resources assessment and management. The mapping efforts, however, were fraught with limitations in terms of temporal consistency and availability of optical data; inconsistent forest classification systems used; and replicability of approaches. This study evaluated the suitability of ALOS/PALSAR mosaic data for mapping and monitoring of forest cover types in the Philippines by assessing temporal consistency, forest classification hierarchy, and classification accuracy. Digital image processing approaches were used including object-based segmentation, extraction of texture features, hierarchical clustering, and decision tree classification. A combination of feature attributes including polarimetric, topographic, and texture information were assessed. Results showed that the ALOS/PALSAR mosaics were temporally consistent within a single acquisition year and across multiple acquisition years, making it a suitable data product for mapping and monitoring forests. Combining polarimetric,

topographic, and texture attributes yielded better classification accuracies compared to using polarimetric SAR data alone, although the addition of ancillary feature attributes marginally decreased temporal consistency of the input data. Employing a decision tree classification approach identified the most relevant image layers and feature attributes for classifying different classes. Hierarchical multi-level classification provided a consistent approach for quantifying the degree of accuracy in discriminating different forest cover types. However, even with the best combination of feature attributes used in study, low classification accuracies were obtained in discriminating forest cover types. Although compared to heuristic methods, using a hierarchical clustering analysis to find natural groupings of forest types in the data yielded better classification accuracies. ALOS/PALSAR was not suitable for discriminating forest types based on the FAO Global Forest Resource Assessment classification scheme, and exploring other classification schemes is recommended.

Contents

1	Introduction	1
1.1	Background	1
1.2	Remote sensing of Philippine forests	4
1.3	Overview of the study	8
1.4	Thesis structure	10
2	Literature Review	11
2.1	Synthetic aperture radar for forest applications	11
2.2	Land and forest cover type mapping using ALOS/PALSAR	15
2.3	Forest classification systems	17
2.3.1	FAO Global Forest Resources Assessment forest classification .	17
2.3.2	Multi-level hierarchical classification schemes	21
2.4	Cluster analysis	25
2.5	Decision tree classification	30
2.6	Texture: the grey-level co-occurrence matrices	34
2.7	Geographic object-based image analysis	36

2.8	Conceptual framework	39
2.9	Aims and objectives	40
2.10	Scope and limitations	41
3	Methodology	43
3.1	Study area	43
3.2	Data	44
3.3	Overall workflow	46
3.4	Image pre-processing	47
3.5	Design of a multi-level classification hierarchy	48
3.6	Identification of regions of interest	51
3.7	Image segmentation	52
3.8	Extraction of feature attributes and backscatter analysis	55
3.9	Hierarchical clustering	58
3.10	Decision tree classification and ruleset development	59
4	Results	62
4.1	Temporal consistency of PALSAR mosaic data	62
4.1.1	Mosaic data acquired in a single year but at different seasons .	62
4.1.2	Mosaic data acquired along one path of strip data but across different years	66
4.2	Consistency of forest classification hierarchy	71
4.3	Accuracy of multi-level hierarchical classification approach	75
4.4	Accuracy of heuristic classification approach	80
4.4.1	Closed and open canopy forests	81

4.4.2	Broadleaved, coniferous, and mixed forests	82
4.4.3	Area statistics of land cover types	83
5	Discussion	98
6	Conclusions and Recommendations	102
Appendices		131
A	Supplementary: Literature Review	132
B	Supplementary: Methods	134
C	Supplementary: Results	137
C.1	Classification trees and cross-validation graphs	137
C.2	Classification error rates for each case	170
C.3	Notes on implementation of the methodology	174
C.4	R scripts developed from this study	176

Acronyms

ALOS	Advanced Land Observing Satellite
AU	Approximately Unbiased
AVHRR	Advanced Very High Resolution Radiometer
CART	Classification and Regression Trees
DEM	Digital Elevation Model
ERS	European Resource Satellite
FAO	Food and Agriculture Organisation of the United Nations
FCCS	Forest Cover Classification System
FMB	Forest Management Bureau
GEOBIA	Geographic Object-Based Image Analysis
GFRA	Global Forest Resources Assessment
GIS	Geographic Information System
GLCM	Grey Level Co-occurrence Matrix
HH	Horizontal transmit - Horizontal receive
HV	Horizontal transmit - Vertical receive

IQR	Inter-Quartile Range
JAXA	Japan Aerospace Exploration Agency
JERS	Japan Earth Resources Satellite
LCCS	Land Cover Classification System
NAMRIA	National Mapping and Resource Information Authority
NFA	National Forest Assessment
OBIA	Object-Based Image Analysis
PALSAR	Phased Array type L-band Synthetic Aperture Radar
REDD+	Reducing Emissions from Deforestation and Forest Degradation
	- Plus
ROI	Region of Interest
SAR	Synthetic Aperture Radar
SIR	Shuttle Imaging Radar
SRTM	Shuttle Radar Topography Mission
SSC	Swedish Space Corporation
UN	United Nations
VH	Vertical transmit - Horizontal receive
VV	Vertical transmit - Vertical receive

List of Tables

1.1	National forest surveys in the Philippines using remotely sensed data.	5
1.2	Forest cover classification systems used in the 1987, 2003, 2010 maps.	6
2.1	Global Forest Resources Assessment categories and definitions.	19
2.1	Global Forest Resources Assessment categories and definitions.	20
2.1	Global Forest Resources Assessment categories and definitions.	21
3.1	Hierarchy and grouping of land cover from detailed to broad-level classes.	51
3.1	Hierarchy and grouping of land cover from detailed to broad-level classes.	52
3.2	Object-level feature attributes extracted from the images.	56
3.3	Four cases representing combinations of object-level feature attributes.	58
4.1	Baker's Gamma Indices (β coefficients) per case for each pair of years.	72
4.2	Summary of Baker's Gamma Indices (β coefficients) for pairwise comparison of dendrograms across all cases.	74
4.3	Summary of total classification error rates per case per year.	79
4.4	Summary of total classification error rates per case per year.	80
4.5	Classification error rates for closed and open canopy forests per case per year using a heuristic approach.	81

4.6	Classification error rates for closed and open canopy forests per case per year using a heuristic approach.	82
4.7	Area statistics of land cover types from classified annual PALSAR mosaics. .	83
4.7	Area statistics of land cover types from classified annual PALSAR mosaics. .	84
A.1	LCCS hierarchical structure adopted for the 2003/2010 NFA in the Philippines.	133
B.1	Land area statistics of forest cover types per province based on NAMRIA 2010 land cover map.	135
C.1	Summary table of classification error rates for Case 1: Polarisation. .	170
C.2	Summary table of classification error rates for Case 2: Polarisation + Topographic.	171
C.3	Summary table of classification error rates for Case 3: Polarisation + Texture.	172
C.4	Summary table of classification error rates for Case 4: Polarisation + Topographic + Texture.	173

List of Figures

1.1	Recent national land cover maps of the Philippines: (a) 1987 SSC; (b) 2003 NAMRIA; (c) 2010 NAMRIA.	7
2.1	Multi-level hierarchical classification system.	23
2.2	A dendrogram with arrows showing agglomerative and divisive hierar- chical clustering techniques.	26
2.3	Methods for computing the distance matrix: (a) single linkage; (b) complete linkage; (c) average linkage.	29
2.4	Texture variables derived from the grey-level co-occurrence matrix. .	35
2.5	The conceptual framework of the study shows forest as the central theme enclosed by elements concerning data, method, indicator, and application, and their interrelationships.	40
3.1	Study area in Northern Luzon, Philippines showing the 2010 land cover map based on NAMRIA data.	45
3.2	Overall image processing workflow implemented in the study.	47
3.3	Detailed sub-workflow for image pre-processing steps implemented on SAR and topographic datasets.	49

3.4	The classification hierarchy proposed for this study.	50
3.5	The distribution of the regions of interest selected over the study area in Northern Luzon, Philippines.	53
3.6	The annual PALSAR mosaics of Northern Luzon acquired in 2007, 2008, 2009, and 2010.	57
4.1	Backscatter values of forest types from PALSAR data acquired in a single year at different seasons.	69
4.2	Backscatter values of forest types from PALSAR data acquired in a single year at different seasons.	70
4.3	Box.1. The 2007 PALSAR mosaic image and the dates of acquisition of each strip/path that comprise the whole mosaic, and the distribution of mean HH backscatter for each forest cover type based on the 2007 mosaic image.	85
4.4	Box.2. A PALSAR mosaic image highlighting the strip/path where ROIs of forest cover types were taken across four years, and the distribution of mean HH backscatter for each forest cover type along the same strip at different years.	86
4.5	Backscatter values of forest types from PALSAR data acquired along the same strip (Cordillera) at different years.	87
4.6	Backscatter values of forest types from PALSAR data acquired along the same strip (Sierra Madre) at different years.	88

4.7	Box 3. Example dendrogram result from the hierarchical clustering of forest cover types using polarimetric feature attributes (Case 1) from the 2007 PALSAR mosaic image.	89
4.8	Dendograms from cluster analysis using Case 1.	90
4.9	Dendograms from cluster analysis using Case 2.	91
4.10	Dendograms from cluster analysis using Case 3.	92
4.11	Dendograms from cluster analysis using Case 4.	93
4.12	Box 4. Example classification trees corresponding to the first four classification levels at Case 1 of the 2007 PALSAR mosaic image. . .	94
4.13	Plots comparing error rates obtained for each year across classification levels per case.	95
4.14	Plots comparing error rates obtained for each case across classification levels per year.	96
4.15	Land cover and forest type maps generated from annual ALOS/PALSAR mosaics using decision tree classification.	97
B.1	Images of segmentation tests.	136
B.2	Images of selected segmentation parameters.	136
C.1	The Case 1 2007 classification trees generated by R tree package. . .	138
C.2	The Case 1 2007 cross-validation graphs generated by R tree package.	139
C.3	The Case 1 2008 classification trees generated by R tree package. . .	140
C.4	The Case 1 2008 cross-validation graphs generated by R tree package.	141
C.5	The Case 1 2009 classification trees generated by R tree package. . .	142
C.6	The Case 1 2009 cross-validation graphs generated by R tree package.	143

C.7	The Case 1 2010 classification trees generated by R tree package.	144
C.8	The Case 1 2010 cross-validation graphs generated by R tree package.	145
C.9	The Case 2 2007 classification trees generated by R tree package.	146
C.10	The Case 2 2007 cross-validation graphs generated by R tree package.	147
C.11	The Case 2 2008 classification trees generated by R tree package.	148
C.12	The Case 2 2008 cross-validation graphs generated by R tree package.	149
C.13	The Case 2 2009 classification trees generated by R tree package.	150
C.14	The Case 2 2009 cross-validation graphs generated by R tree package.	151
C.15	The Case 2 2010 classification trees generated by R tree package.	152
C.16	The Case 2 2010 cross-validation graphs generated by R tree package.	153
C.17	The Case 3 2007 classification trees generated by R tree package.	154
C.18	The Case 3 2007 cross-validation graphs generated by R tree package.	155
C.19	The Case 3 2008 classification trees generated by R tree package.	156
C.20	The Case 3 2008 cross-validation graphs generated by R tree package.	157
C.21	The Case 3 2009 classification trees generated by R tree package.	158
C.22	The Case 3 2009 cross-validation graphs generated by R tree package.	159
C.23	The Case 3 2010 classification trees generated by R tree package.	160
C.24	The Case 3 2010 cross-validation graphs generated by R tree package.	161
C.25	The Case 4 2007 classification trees generated by R tree package.	162
C.26	The Case 4 2007 cross-validation graphs generated by R tree package.	163
C.27	The Case 4 2008 classification trees generated by R tree package.	164
C.28	The Case 4 2008 cross-validation graphs generated by R tree package.	165
C.29	The Case 4 2009 classification trees generated by R tree package.	166

C.30 The Case 4 2009 cross-validation graphs generated by R tree package.	167
C.31 The Case 4 2010 classification trees generated by R tree package. . .	168
C.32 The Case 4 2010 cross-validation graphs generated by R tree package.	169

Measurement Units

B	billion
dB	decibel
ha	hectare
km	kilometer
m	meter
M	million
yr	year

Chapter 1

Introduction

1.1 Background

Forests are complex, functional systems of interacting and interdependent biological, physical, and chemical components, which creates unique combinations of climate, soil, plant and animal species that result in many different forest types around the world (Blanco and Lo, 2012). These ecosystems harbor the vast majority of biological species on Earth (Pan, Birdsey, Phillips, et al., 2013), account for 80% of the Earth's total plant biomass (Kindermann et al., 2008), and store more carbon in biomass and soils than is contained in the atmosphere (Pan, Birdsey, Fang, et al., 2011). Forests not only play an important role in supporting and maintaining ecological systems and cycles, but also provide a broad range of goods and services to humanity, including social and economic benefits in the long term (FAO, 2015b). Among the valuable ecosystem goods and services provided by forests are food, fibre, timber, medicine, clean water, aesthetic and spiritual values, climate regulation, a repository of carbon

storage, and mitigation of natural hazards such as floods (Pimentel et al., 1997; Pearce, 2001; Blanco and Lo, 2012; Pan, Birdsey, Phillips, et al., 2013; Chazdon et al., 2016).

Nearly 1.6 billion people comprising more than 25% of the worlds population depend on forest resources for their livelihoods (FAO, 2015a), and another estimated 410 million people are highly dependent on forests for their subsistence and income (Munang et al., 2011). Recent estimates of the global economic values of terrestrial forest ecosystems increased from US\$ 2,769/ha/yr in 1997 to US\$ 5,682/ha/yr in 2011 for tropical forests, and from US\$ 417/ha/yr in 1997 to US\$ 3,137/ha/yr in 2011 for temperate/boreal forests (standardised in terms of 2007 US\$ values; Costanza et al., 2014). The estimates of the economic value of mangrove forests were even higher at US\$ 13,786/ha/yr in 1997 to US\$ 193,843/ha/yr in 2011, which were mainly due to new studies of the value of mangroves in terms of storm protection, erosion prevention, and waste treatment (Costanza et al., 2014).

Despite their renowned importance, however, forests have declined at a staggering rate around the world. Achard et al. (2014) showed that 8.0 M ha/yr of tropical forests were lost between 1990 to 2000, which decreased to 7.6 M ha/yr between 2000 to 2010. Another study by Mayaux et al. (2005) further estimated the annual rates of forest loss in the 1990s for different forest types, particularly 5.8 M ha/yr of humid tropical rainforests, 2.2 M ha/yr of tropical moist deciduous forests, and 0.7 M ha/yr of tropical dry forests. For the period between 2000 to 2012, Hansen et al. (2013) mapped global forest cover change and estimated that a total of 230 M ha of forest were lost due to disturbance while 80 M ha of new forests were established, from which 20 M ha of land experienced both forest loss and forest gain.

Tropical deforestation was mainly driven by agricultural expansion, commercial wood extraction, and infrastructure extension, which were prominently underpinned by economic, political, and institutional factors (Geist and Lambin, 2002). Agricultural expansion was the main driver of forest cover loss across the tropics between the period of 1980 to 2000, wherein more than 55% of new agricultural land replaced intact forests and another 28% came from disturbed forests (Gibbs et al., 2010).

As of 2015, forests occupy 30.6% of the Earth's total land area (FAO, 2015b). But although the world's forests decreased from 4.128 B ha in 1990 to 3.999 B ha in 2015 (Keenan et al., 2015), the Food and Agriculture Organisation (FAO) reported that this 25-year period has seen positive results overall, such as in the rate of global net forest loss being cut by over 50% (i.e., decreased to 0.08% annually since 2010 being half the rate in the 1990s; Sloan and Sayer, 2015); the higher attention given to sustainable forest management worldwide (MacDicken et al., 2015); that forests are meeting increasing demand for a variety of forest products and services (Payn et al., 2015); and larger areas being designated for biodiversity conservation (Morales-Hidalgo, Oswalt, and Somanathan, 2015).

Sustainable forest management is key to ensuring that the ever-increasing demands of a growing human population are met while maintaining ecological functions of healthy forest ecosystems (MacDicken et al., 2015). Forest resources assessment forms an essential part of sustainable forest management, and determining the nature and distribution of land cover is the first step in any forest resources assessment (Sanden and Hoekman, 1999). Gathering better information to monitor and analyse the status and trends on forest resources can further improve forest management (Siry, Cubbage,

and Ahmed, 2005). To this end, accurate estimates of the extent of forests and forest types are indispensable for understanding forest ecosystems and their sustainable use (Foody, 2003), of which remote sensing techniques have been reliably employed over the past several decades as an accurate, systematic, and cost-effective approach for mapping and monitoring forest conditions and changes across a range of scales (Myers, 1988; Franklin, 2001; Dahdouh-Guebas, 2002; Boyd and Danson, 2005). Combining remote sensing with field observations provides a comprehensive monitoring approach in support of sustainable forest management at a range of scales (Keenan et al., 2015).

1.2 Remote sensing of Philippine forests

In the Philippines, a nation-archipelago located between the equator and the Tropic of Cancer in Southeast Asia, forests harbor extremely high floral and faunal diversity, and at the same time also serve as home to upland communities that are highly dependent on forest resources (Lasco and Pulhin, 2000). Studies have cited estimates of the extent of forest cover to be ~90% of the 30 M ha total land area of the Philippine islands before the Spanish colonisation in 1521, down to ~70% at the transition of colonial regimes between Spain and America by 1900s, and further reduced to ~60% during the Japanese occupation of the Philippines from 1942 to 1945 (Kummer, 1992a; Liu, Iverson, and Brown, 1993; Bankoff, 2007). During the 20th century, the average deforestation rate in the Philippines was estimated to be 148,000 ha/yr (Lasco and Pulhin, 2000). Recent official forestry statistics in 2015 reported a decline in forest cover from 7.17 M ha in 2003 to 6.84 M ha in 2010 (FMB, 2015). Timber harvesting and agricultural expansion during the Spanish colonisation,

followed by rapid and extensive commercial logging in the 20th century were recognised as the main drivers of historical forest loss in the Philippines (Eder, 1990; Kummer, 1992a; Chokkalingam et al., 2006; Bankoff, 2007).

Table 1.1: National forest surveys in the Philippines using remotely sensed data (adopted from Kummer (1992b)).

Year	Source	Forest Cover (%)	Data Source	Method of Interpretation
1973	Lachowski et al. (1979)	38.0	Landsat	Digital
1974	Bruce (1977)	29.8	Landsat	Visual
1976	Bonita & Revilla (1977)	30.0	Landsat	Visual
1980	Forestry Development Center (1985)	25.9	Landsat	Visual
1987	Swedish Space Corporation (1988)	23.7	SPOT	Visual
2003	National Mapping & Resource Information Authority (2004)	No data	Landsat	Visual
2010	National Mapping & Resource Information Authority (2014)	No data	Landsat, AVNIR	Visual

Note: At the time of its publication, Kummer's (1992b) study enumerated five national forest surveys using remotely sensed data conducted from the 1970s to the 1980s. Additional surveys conducted from 1990s to the present were included to update this list.

National forest surveys were conducted in the post-1950s period in support of managing the country's forest resources, which employed either ground-based forest surveys or spaceborne remote sensing technology to produce land and forest cover maps. To date, seven national forest surveys using remotely sensed data have been conducted in the Philippines (Table 1.1).

The forest cover maps generated from these past surveys were done mainly by the traditional method of visual or manual interpretation and analysis (except the 1973 maps) using optical remote sensing data, particularly Landsat. While visual interpretation of remotely sensed data is a simple and effective method that can result in excellent spatial information extraction, it is dependent on the extent of knowledge of the analyst regarding the area of study (Bhatta, 2013). Also, since the method depends solely on a human analyst it is more subjective, and tends to be tedious and slow compared to automated digital interpretation techniques (Horning et al., 2010).

Table 1.2: Forest cover classification systems used in the 1987 and 2003/2010 forest maps.

1987 map	2003/2010 map
Closed canopy, mature >50%	Closed forest, broadleaved
Open canopy, mature <50%	Closed forest, coniferous
Mossy forest	Closed forest, mixed
Pine forest	Open forest, broadleaved
Mangrove forest	Open forest, coniferous
Marshy area and swamp	Open forest, mixed
Submarginal forest	Forest plantation, broadleaved
	Forest plantation, coniferous
	Forest plantation, mangrove (2003 only)
	Mangrove forest

The exclusive use of optical satellite data in these mapping surveys meant that cloud cover was a prevalent concern. In particular, tropical regions such as the Philippines feature persistent cloud cover, which affects data availability and temporal consistency of optical data, thereby preventing regular observations (Myers, 1988). The lack of available optical data due to cloud cover contamination may be one of

the major limiting factors in the application of the resulting map products from these forest surveys for periodic, consistent, and wall-to-wall national land and forest cover mapping and monitoring (Fig. 1.1).

Different forest classification systems were also adopted in these forest surveys. For example, the 2003 and 2010 NAMRIA maps employed similar forest classification schemes based on the FAO Global Forest Resources Assessment (GFRA), but these maps differed from the 1987 Swedish Space Corporation (SSC) map that adopted a different system (Table 1.2). The forest definitions used to produce the 1987 map and the 2003/2010 maps were also different from each other (Swedish Space Corporation, 1988; FAO, 2010). These inconsistencies restrict the direct comparison of these maps and render the results incompatible for historical forest cover change analysis.

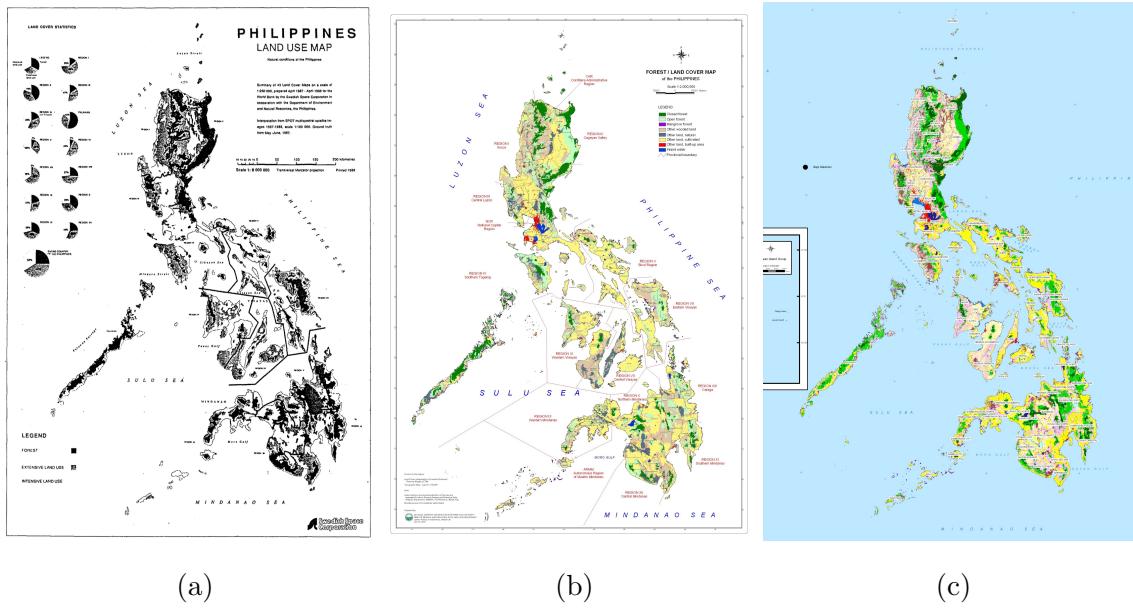


Fig. 1.1: Recent national land cover maps of the Philippines: (a) 1987 SSC; (b) 2003 NAMRIA; (c) 2010 NAMRIA.

Finally, there is little known documentation that exists to report on the methods implemented and levels of accuracy achieved in producing many of the forest cover maps from these surveys, except the 1973 and 1987 maps. Lachowski et al. (1979), for the 1973 Landsat-based maps, described the methods undertaken by the mapping work and mentioned an overall accuracy of 85% to 95% for the classification of forest types, albeit without sufficient detail. For the 1987 SPOT-based maps, SSC Swedish Space Corporation, 1988 described their methods and land cover classification system, but did not quantify the accuracy of the results of their interpretation. Kummer's study had cautioned on the reliability of the 1987 SSC map due to major differences with the results of the 1988 Philippine-German Forest Inventory (Kummer, 1992b). Reports for other mapping surveys were either not (or no longer) available or were not easily accessible. Since many of these forest mapping surveys employed visual interpretation techniques implemented by expert analysts, the absence of documentation also affects the replicability of the approaches for future work since the knowledge developed from these surveys were ultimately lost.

1.3 Overview of the study

There is an opportunity to explore alternative datasets and methods in support of finding solutions to improve the mapping and monitoring of Philippine forests using remote sensing technology. Synthetic aperture radar (SAR) is an active remote sensing technology that utilises the microwave region of the electromagnetic spectrum that may be used as an alternative (or complement) to optical data for forest mapping and monitoring (discussed in more detail in the next chapter). Innovations in computing

technology have leapfrogged over the last two decades and made it possible for computers to handle large remote sensing datasets and to implement sophisticated image processing algorithms for various applications.

This research intends to examine the potential of SAR data and digital image processing techniques for mapping forest cover extent and discriminating forest types in the Philippines based on the adopted FAO forest classification system. In contrast to previous national forest mapping surveys using remotely sensed data, which have utilised mainly optical data and visual interpretation for forest cover classification, this study employs image data from an active sensor that can at the onset overcome the limitations brought by persistent cloud cover, and aim to develop a replicable digital image processing and interpretation approach. This study documents the evaluation of the suitability of SAR data for mapping and discriminating forest types by providing quantitative accuracy assessments, and information of the relationships between image data predictor variables and forest types, which are important in developing robust and transparent forest monitoring systems.

This study also adopts the forest categories based on the FAO GFRA for consistency with the recent official national forest cover maps (i.e., 2003 and 2010). One of the key research thrusts in support of planning and decision-making of the National Mapping and Resource Authority (NAMRIA) is to utilise SAR imagery, either as an alternative to or in complementation with optical data, coupled with digital image processing approaches with the aim of developing improved wall-to-wall coverage baseline land and forest cover map products of the entire country on a periodic basis. This study can potentially support NAMRIA's aims by providing insights on the

suitability of radar data and a suite of digital image processing techniques for mapping and monitoring forest extent and forest types in the Philippines.

1.4 Thesis structure

This thesis is divided into six chapters. Chapter 1 gives a background of forest ecosystems and remote sensing of forests, and presents an overview of the study. Chapter 2 provides a review of relevant literature on forest applications of synthetic aperture radar, forest classification systems, and the approaches used in this study including cluster analysis, decision trees, texture measures, and object-based image analysis. Chapter 3 describes the study area, the datasets used, and the methods employed. Chapter 4 presents the results of the feature extraction, and clustering and classification procedures. Chapter 5 provides a thorough discussion of the results. Chapter 6 summarises the study and outlines recommendations for future research and application. The final sections contain the references cited in the text, and the appendices consisting of supplemental figures, rulesets, and scripts used in the study.

Chapter 2

Literature Review

2.1 Synthetic aperture radar for forest applications

Synthetic aperture radar (SAR) is a viable alternative to optical remote sensing for land cover characterisation, mapping, and monitoring to support environmental studies and resource management. SAR systems transmit short pulses of microwave energy toward the Earth's surface, which then interact with different surface features like vegetation, infrastructure, and water bodies, and the sensor provides information such as backscattering coefficients, which represent the strength of the microwave energy emitted from the antenna and returned after scattering by the target features (Thapa, Itoh, et al., 2014). Backscatter is characterised by polarised modes of the transmitted and received signals, particularly the co-polarisation channels: horizontal transmit and horizontal receive (HH) and vertical transmit and vertical receive (VV); and the cross-polarisation channels: horizontal transmit and vertical receive (HV) and vertical transmit and horizontal receive (VH). Scattering mechanisms that contribute

to backscattered energy include surface scattering from water bodies and slightly rough surfaces, double bounce scattering from dihedral surfaces and edges, and volume scattering from forests and other vegetation (Van Zyl, 1989; Freeman and Durden, 1998).

SAR sensors are unaffected by atmospheric conditions and can penetrate through clouds, haze, and smoke, thereby overcoming limitations of optical data in cloudy tropical regions (Achard, Stibig, et al., 2010; Shimada and Ohtaki, 2010). As an active remote sensing system, radar imagery can be calibrated to a very high degree of accuracy with its constant illumination geometry and well-established transmit/receive power spectra; hence allowing the comparison of backscatter values (σ^0) over time and space (i.e., structurally and electrically similar vegetation have the same backscattering behaviour anywhere on earth; Kellndorfer et al., 1998). SAR data can be acquired consistently on a repetitive basis regardless of weather conditions; hence, these data can potentially be used to complement or replace optical remote sensing systems in an operational programme for mapping and monitoring of forests (Almeida-Filho et al., 2007).

Notable studies, as described next, were aimed at understanding the relationships between SAR data and forest structure to explore the utility of SAR for forest applications. Ulaby et al. ((1986)) studied the suitability of radar data from two L-band SAR missions, Seasat and the Shuttle Imaging Radar (SIR-A), for mapping forest types in North and South America. They found that combining backscatter and texture measures improved discrimination of forest types using supervised classification techniques. Durden et al. ((1989)) described a theoretical scattering mechanism model

for forested areas and found good agreement between modelled polarisation signatures with observed polarisation measurements using ground data and AIRSAR L-band data. Le Toan ((1992)) analysed the relationships of multi-frequency (i.e., C, L, and P bands) quad-polarimetric (i.e., HH, HV, VH, VV) AIRSAR data with forest stand parameters measured from pine forests ranging from young to mature stages, which revealed that longer wavelength SAR data, particularly P- and L-bands, had strong correlations and high sensitivities to forest biomass. Moreover, cross-polarised signals showed high correlations with forest stand parameters. Imhoff ((1995)) investigated the effect of forest structural differences in equal-biomass broadleaf forest stands on radar backscatter similarly from multi-frequency quad-polarimetric AIRSAR data. The results indicated that structure exerted a substantial effect on backscatter, which tended to increase with forest structural consolidation into fewer and larger parts; hence providing a structural link to positive correlations of radar backscatter with aboveground biomass. Woodhouse et al. ((2012)) believes that “radar is typically the best satellite-based remote sensing tool for mapping forest extent, estimating forest structural variability, and detecting deforestation and degradation.”

The Advanced Land Observing Satellite (ALOS), launched on 24 January 2006, was developed for detailed observation of the Earths surface and frequent monitoring of global environmental changes, such as forest ecosystems, using both optical and SAR sensors (Shimada, Tadono, and Rosenqvist, 2010). Onboard ALOS is the Phased Array type L-band Synthetic Aperture Radar (PALSAR), a multi-polarisation L-band SAR system, which builds on the experience gained from its predecessor, the Japan Earth Resources Satellite (JERS-1), similarly a spaceborne L-band SAR sensor that

was successfully used for global rainforest and boreal forest mapping projects in the 1990s (Rosenqvist, Shimada, Chapman, Freeman, et al., 2000; Rosenqvist, Shimada, Chapman, McDonald, et al., 2004). L-band SAR (\sim 25 cm wavelength) is particularly useful in mapping forest areas due to its longer wavelengths and shorter frequencies (Thapa, Itoh, et al., 2014), which in turn allows greater penetration through forest vegetation by the transmitted signal and sensitivity for detecting deforestation (i.e., L-band backscatter is typically lower in non-forest areas compared to forests) and forest inundation (i.e., L-band backscatter is typically lower in inundated forest areas); thereby yielding important information for forest cover monitoring (Shimada, Tadono, and Rosenqvist, 2010; Shimada and Ohtaki, 2010). The co-polarised channel (i.e., HH) has been found to be sensitive to vertical structure, such as regrowth in clear-cut regions, while the cross-polarised channel (i.e., HV) is less sensitive to vertical structure in low vegetation or clear-cut areas (Shimada, Tadono, and Rosenqvist, 2010). ALOS also featured a unique systematic observation strategy to provide consistent, semi-annual, wall-to-wall observations at fine spatial resolution of the Earth's entire landmass during the satellite's lifetime (Rosenqvist, Shimada, Ito, et al., 2007).

Since its launch, ALOS/PALSAR data have been used for various forest applications such as detecting tropical deforestation (Almeida-Filho et al., 2009; Rahman and Sumantyo, 2010; Whittle et al., 2012; Motohka et al., 2014); time-series mapping of forest cover (Thapa, Shimada, et al., 2013; Shimada, Itoh, et al., 2014); assessing forest fragmentation (Dong et al., 2014); estimating mangrove biomass (Hamdan, Khali Aziz, and Mohd Hasmadi, 2014) and forest aboveground biomass (Mitchard et al., 2009; Morel et al., 2011; Sarker et al., 2012; Cartus, Santoro, and Kellndorfer,

2012; Rahman and Sumantyo, 2013; Mermoz, Réjou-Méchain, et al., 2015; Thapa, Watanabe, et al., 2015); mapping mangrove areas (Rocha de Souza Pereira, Kampel, and Cunha-Lignon, 2012) and changes in mangrove extent (Nascimento et al., 2013); and assessing processes of forest disturbances and successional dynamics (Joshi et al., 2015; Mermoz and Le Toan, 2016). PALSAR is also one of the main data sources for the Task on Forest Carbon Tracking, established by the Group on Earth Observation in 2009 as a global forest monitoring system in support to the United Nations Framework Convention on Climate Change post-Kyoto Protocol framework (Shimada, Tadono, and Rosenqvist, 2010).

2.2 Land and forest cover type mapping using ALOS/PALSAR

Studies have demonstrated the use of ALOS/PALSAR for mapping tropical forest types and broad land cover categories, which is the topic of interest of this study. For example, quad-polarimetric PALSAR data were used for mapping broad land cover categories. Rahman and Sumantyo (2010) mapped forest, forest re-growth, upland soil/shrub, lowland soil, settlement, and water/wetlands in Southern Chittagong, Bangladesh. They found that forests were discriminated well from other land cover categories, although forest re-growth was not detected, and only a 66% overall classification accuracy was achieved. Mishra et al. (2011) tested different classifiers for mapping broad land cover types (i.e., tall vegetation, short vegetation, urban, bare soil, and water) in Roorkee, India. Their results showed that decision tree classifier

obtained an 86% overall accuracy and performed better than other classification techniques. Avtar et al. (2012) assessed two images for characterising forest types (evergreen, deciduous, sparsely deciduous) and deforestation in Northern Cambodia. Their study showed that HV-polarised backscatter coefficient, cross-polarisation ratio (HH/HV), among other parameters gave better results for characterising forest types. Further, they observed that HV-polarised backscatter coefficient of evergreen forest was higher compared to deciduous and sparsely deciduous forest types, particularly due to the multiple canopies of evergreen forests that enhance volume scattering.

Previous studies have used dual-polarised PALSAR data combined with texture information derived from the radar images. Longepe et al. (2011) mapped land cover types (i.e., dry forest, wet forest, acacia, clear cut, oil palm, others) in Riau province, Sumatra, Indonesia using dual-polarised (HH, HV) PALSAR data and texture measures. The overall agreement with reference data was ~70% for land cover classification and ~87% for natural forest discrimination. Similarly, Rakwatin et al. (2012) mapped broad land cover types (i.e., forest, oil palm, acacia, clear cut, water) in Riau province, Indonesia, and reported a 73% overall classification accuracy. The overall accuracies for land cover classification increased by about 10% when combined with texture information. Li et al. (2012) used dual-polarised PALSAR data and texture measures for mapping broad land cover types (i.e., forest, succession, agropasture, water, wetland, urban) in Altamira, Para, Brazil. The best classification resulted in a 74% overall accuracy, of which textural information was valuable for improving vegetation classification. They also reported difficulties in discriminating between forest classes and between forest succession classes.

Discriminating tropical forest types based on successional stages have been studied due to its importance in understanding the role of forest regeneration in global carbon cycles. Li et al. (2012) found that neither L-band PALSAR nor C-band RADARSAT-2 dual-polarisation data can accurately separate between detailed forest types (upland, flooding, liana) and between forest succession stages (initial, intermediate, advanced). Upon testing different classifiers, they also found that no algorithm could successfully separate forest succession stages. Liesenberg & Gloaguen (2013) evaluated different PALSAR data modes (e.g., single-, dual-, quad-polarimetric, and interferometric) to characterize and classify primary, successional secondary, and riparian forest classes in Eastern Amazon, Brazil. They found that while quad-polarised data provided better results compared to single or dual-polarised data, the overall accuracies remained low. Radar backscatter of primary and riparian forests were higher than those observed for successional secondary forest types (initial, intermediate, advanced). Backscatter values tended to increase from advanced to initial secondary stage forest, although no substantial differences were observed between successional secondary forest types.

2.3 Forest classification systems

2.3.1 FAO Global Forest Resources Assessment forest classification

The above examples on mapping forest cover types using ALOS/PALSAR data have adopted classification schemes based on forest age or successional stages (e.g., primary, secondary), disturbance gradients (e.g., pristine, degraded), or on bespoke

classes suited to the area of interest. However, using different classification schemes makes compatibility between datasets difficult to ascertain and hinders the comparative analysis of land cover types across space and time in a consistent manner (Ahlqvist, 2008). Although, at present, no definitive universally accepted land cover classification system exists (Townshend et al., 2010), such a classification system is crucial for the evaluation, comparison, and change analysis of land cover products (Giri, 2012).

The Global Forest Resources Assessments (GFRA), coordinated by FAO at the request of its member countries, provides periodic information on the situation of the world's forests to support decision-making in investment and policymaking in forestry and sustainable development (FAO, 2015b). One of the primary sources of data for these assessments is the Country Report prepared by national correspondents, which are based on National Forest Assessments (NFA) and are submitted to FAO for consolidation in a global report produced every 5 years. Forest inventories, which are based on systematic field sampling and complemented by remote sensing, are the primary means for data collection within these NFAs. The FAO formulates guidelines for country reporting, including forest classification terms and definitions, to ensure compatibility and interoperability of information across countries and time for global forest resources monitoring, as well as facilitating comparative analysis of land cover change within national-level assessments.

In the Philippines, national remote sensing surveys were conducted by NAMRIA to complement national forest inventories for the reporting periods in 2005 and 2010 (i.e., 2003 and 2010 land cover map products), which adopted the FAO GFRA classification and definitions (Table 2.1). Forest inventories were only implemented

for the 2005 reporting, but were not carried out for the 2010 reporting (FAO, 2010).

For consistency, this study similarly adopted the same forest classification system for evaluating L-band PALSAR data for distinguishing forest cover types based on this system, which would allow comparison with the official land cover maps.

Table 2.1: Categories and definitions based on the Global Forest Resources Assessment (Source: FAO, 2005; FAO, 2010).

Category	Definition
Forest	Land with tree crown cover (or equivalent stocking level) of more than 10% and area of more than 0.5 ha.
Broadleaved forest	Forest with predominance (more than 75% of tree crown cover) of trees of broadleaved species.
Coniferous forest	Forest with predominance (more than 75% of tree crown cover) of trees of coniferous species.
Bamboo/palm formations	Forest on which more than 75% of the crown cover consists of tree species other than coniferous or broadleaved species (e.g. tree-form species of the bamboo, palm and fern families).
Mixed forest	Forest in which neither coniferous, nor broadleaved, nor palms, bamboos, account for more than 75% of the tree crown cover.
Closed forest ($\geq 40\%$)	Natural forest where trees in the various storeys and undergrowth cover 40% of the ground. These formations do not have a continuous dense grass layer. They are either managed or unmanaged forests primary or in an advanced state of reconstitution and may have been logged-over one or more times, having kept their characteristics of forest stands, possibly with modified structure and composition. Typical examples of tropical closed forest formations include tropical rainforest and mangrove forest.

Table 2.1: Categories and definitions based on the Global Forest Resources Assessment (Source: FAO, 2005; FAO, 2010).

Category	Definition
Open forest (10 to <40%)	Formations where trees form a discontinuous layer covering 10 to 40% of the ground. This forest usually includes a continuous grass layer allowing grazing activities and the spreading of fires.
Forest plantation	Forest stands established by planting or/and seeding in the process of afforestation or reforestation. They are either of introduced species (all planted stands), or intensively managed stands of indigenous species, which meet all the following criteria: one or two species at plantation, even age class, regular spacing.
Open broadleaved forest plantation (10 to <40%)	Forest plantation where the crown cover is between 10 and 40% of the area.
Closed broadleaved forest plantation (\geq 40%)	Forest plantation where the crown cover is above or 40% of the area.
Other wooded land	Land either with a crown cover (or equivalent stocking level) of 5 to 10% of trees able to reach a height of 5 m at maturity in situ; or a crown cover (or equivalent stocking level) of more than 10% of trees not able to reach a height of 5 m at maturity in situ (e.g., dwarf or stunted trees); or with shrub or bush cover of more than 10%.
Shrubs	Refers to vegetation types where the dominant woody elements are shrubs i.e. woody perennial plants, generally of more than 0.5 m and less than 5 m in height on maturity and without a definite crown. The height limits for trees and shrubs should be interpreted with flexibility, particularly the minimum tree and maximum shrub height, which may vary between 5 and 7 m approximately.

Table 2.1: Categories and definitions based on the Global Forest Resources Assessment (Source: FAO, 2005; FAO, 2010).

Category	Definition
Fallow	It encompasses forest fallow where the woody vegetation is under 5 m in height. It refers to woody vegetation deriving from the clearing of natural forest for shifting agriculture. It is part of a forest fallow consisting of a mosaic of various reconstitution phases. The vegetation does not reach a height of 5 m.
Wooded grasslands (5 to <10%)	Land where the trees cover between 5 to 10% of the area and their height may reach 5 m at maturity.
Other land	Land not classified as forest or other wooded land, as described above. Including cultivated land, grasslands and pastures, built-up areas, barren land etc.
Inland water	Area occupied by major rivers, lakes, and reservoirs.

2.3.2 Multi-level hierarchical classification schemes

According to Di Gregorio (2005), a hierarchically structured land cover classification offers more consistency due to its ability to accommodate different levels of information (i.e., beginning with broad-level classes that are systematically subdivided into detailed sub-classes). The following examples describe some of the early development and full implementations of multi-level hierarchical land cover classification schemes applied to SAR data.

Pierce et al. (1994) and Dobson et al. (1996) were among the initial studies that implemented a “knowledge-based” approach for classifying land cover types using polarimetric AIRSAR data over Northern Michigan. The design of the approach uses

knowledge of the nature of radar backscatter from target features to formulate decision rules for classification in a sequential format. The land cover classes were defined on the basis of simple structural attributes of widespread applicability, relying heavily on their expert knowledge, and decision rules were determined empirically using training data to define simple thresholds (Dobson, Pierce, and Ulaby, 1996). Their knowledge-based classifier operated sequentially at each level of the class hierarchy. Overall accuracies above 90% were reported from the results of the classification by both studies.

In another study, Walker et al. (2010) devised a bespoke hierarchical, five-level land cover classification scheme (see Fig. 2.1) for large-area land and forest cover classification and mapping of the Xingu watershed in the Brazilian Amazon using dual-polarised PALSAR data and feature extraction (such as texture measures). Starting from detailed land cover types at the first level, classes were aggregated in each subsequent level until the last level, which consisted of forest and non-forest classes. The forest sub- categories were forest and *cerrado*, each of which were further sub-categorised into intact, degraded, or regenerating forests. Overall accuracies were low at the detailed classification level (58% at 15 classes), which increased as classes were aggregated at subsequent levels, eventually obtaining a high accuracy at the broad classification level (92% at 2 classes).

Currently, one of the most comprehensive, flexible, and internationally applied framework for land cover characterisation is the United Nations (UN) Land Cover Classification System (LCCS) developed by FAO (Giri, 2012). The UN LCCS was developed in response to a need for reliable, harmonised, and standardised information

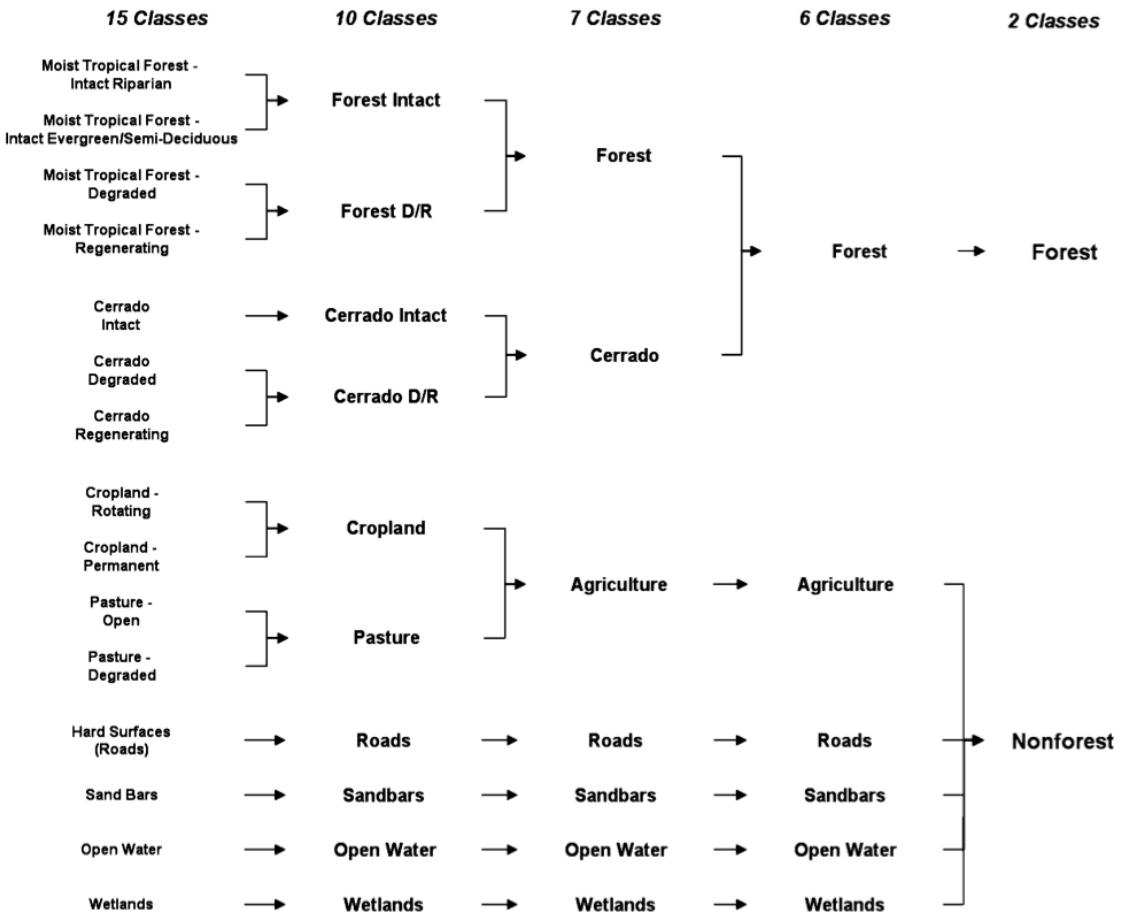


Fig. 2.1: Multi-level hierarchical classification system adopted by Walker et al. (2010).

on land cover and land cover change (Di Gregorio, 2005). It provides descriptors of land cover classes through a dichotomous, modular, hierarchical system for translation between typologies (Defourny and Bontemps, 2012) such that the comparison of land cover products, including qualitative and quantitative evaluations of product differences, are enabled (Kuenzer et al., 2014). The first fully operational version of the UN LCCS was developed initially for the implementation of the AFRICOVER project (Kalensky, 1998; Herold and Di Gregorio, 2012). Some global land cover products that are compliant with the UN LCCS standards include the GLC2000 (Bartholomé and Belward, 2005) and GlobCover (Arino et al., 2007).

Two studies described as follows used PALSAR data for FAO LCCS-compliant land and forest cover mapping. Hoekman et al. (2010) mapped forest and land cover types in sub-continental Borneo using single- and dual-polarised PALSAR data in 2007. The resulting demonstration product was 85.5% in full agreement with an independent reference dataset. The map legend consisted of 22 classes, including eight forest types, namely: tropical lowland forest, broadleaved evergreen; tropical mountain forest, broadleaved evergreen; riverine forest; swamp forest; mangrove forest; nipah mangrove forest; peat swamp (pole) forest; peat swamp (*padang*) forest; and forest mosaics/fragmented or degraded forest. Except for nipah mangrove forest, the degree of full agreement was very high for lowland forest and fair/good for other natural forest types.

Another study by Otupei et al. (2014) mapped land cover types in Bwindi Impenetrable National Park, Uganda based on the FAO Africover LCCS using quad-polarimetric ALOS/PALSAR data. The legend consisted of eight land cover types

including dense evergreen forest, dry bare farmland, wet bare farmland, mixed farmland, mixed rangeland, degraded wetland, aquatic vegetation, and open water. An overall classification accuracy of 86% was achieved, of which mixed rangeland and farmland were found to be the least discriminated.

While these studies and implementations, including the FAO GFRA-based land cover/forest type classification (see Appendix A), applied hierarchically structured classification approaches, the determination of the sequence and order of the classes in these hierarchies were heuristically determined based on expert knowledge or on what made logical sense (i.e., similar to the “classification logic” proposed by Running et al. (1995) and implemented by DeFries et al. (1998) and Hansen et al. (2000) for generating AVHRR-based global land cover products). For this study, a hierarchical clustering analysis was implemented (described in more detail in the next section) to determine a natural hierarchy of forest cover classes based on the multi-year PALSAR data, which would then define the multi-level class hierarchy to guide the execution of the classification.

2.4 Cluster analysis

Cluster analysis, or clustering, is the art of finding groups or structure in the data (Kaufman and Rousseeuw, 1990). The purpose of cluster analysis is to group observations into clusters in such a way that observations in the same cluster are more similar to one another than they are to observations in other clusters (Bajorski, 2011). Cluster analysis, also often called unsupervised learning, does not require prior information about the groupings of observations and is commonly employed for

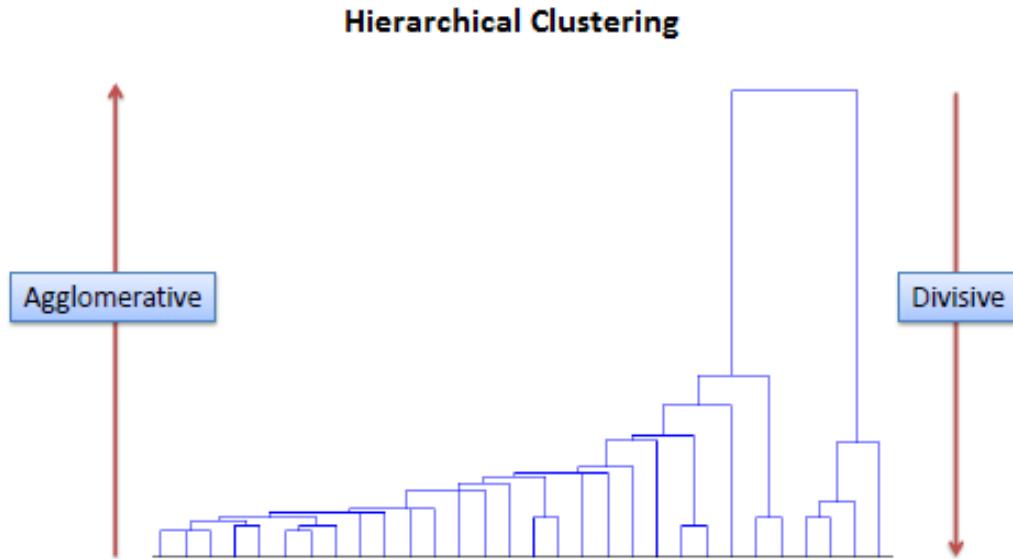


Fig. 2.2: A dendrogram with arrows showing agglomerative and divisive hierarchical clustering techniques (Source: Sayed, 2016).

descriptive or exploratory purposes to better understand the structure of the data (Bajorski, 2011).

To measure the degree of similarity or dissimilarity between observations, distances are computed on the basis of their proximity in two-dimensional space in such that the smaller the distance between observations represents greater similarity and larger distances as less similarity (Hair et al., 2013). Euclidean distance is the most commonly used measure of similarity between two observations (Kaufman and Rousseeuw, 1990), which is essentially a measure of the length of a straight line drawn between two observations when shown graphically.

There are two types of clustering procedures: hierarchical and non-hierarchical (or partitioning). Non-hierarchical clustering procedures divide the observations into a

pre-defined number of clusters, which can be defined by the user or calculated within the algorithm (Bajorski, 2011). The K-means is one of the most commonly used algorithms for non-hierarchical methods, which works by partitioning observations into a specified number of clusters and iteratively reassigns observations until a criteria for cluster homogeneity is met (Hair et al., 2013).

Hierarchical clustering procedures, on the other hand, do not require observations into a pre-defined number of clusters. The result is a tree structure, more commonly known as a dendrogram (see Fig. 2.2), which is a graphical representation of the results of the hierarchical procedure that shows how the clusters are combined at each step of the procedure until all are contained in a single cluster (Hair et al., 2013). The dendrogram is a one-dimensional display of similarity with the height of the join indicating dissimilarity (Ripley, 1996). According to Bajorski (2011), hierarchical clustering methods often provide more clustering information than other methods.

Johnson (1967) provided a succinct outline of the hierarchical clustering process. Given a dataset of N observations to be clustered and $N \times N$ distance (or similarity) matrix:

1. Assign each observation to its own cluster, i.e., N observations will be equal to N clusters. Let the distances (or similarities) between clusters equal the distances similarities between observations they contain;
2. Search for the closest (most similar) pair of clusters and combine them into a single cluster, such that there is one less cluster;
3. Compute the distances (similarities) between the new cluster and each of the old clusters;

4. Repeat steps #2 and #3 until all items are grouped into a single cluster of size N.

In hierarchical clustering procedures, there are two types of methods: agglomerative and divisive (Kaufman and Rousseeuw, 1990). These two methods construct their hierarchy in opposite directions (Fig. 2.2). Divisive methods begin with all observations under one cluster and subsequently divide the data into the two meaningful clusters, then the process continues until the number of clusters is achieved. Agglomerative methods start with each observation forming a separate cluster, and are subsequently merged into the most similar clusters until all observations are in a single cluster.

The algorithm used in a hierarchical clustering procedure defines how similarity is measured in the clustering process (Hair et al., 2013). There are numerous algorithms for agglomerative methods, which only vary in their definition of between-cluster dissimilarity (Kaufman and Rousseeuw, 1990). Some of the most common methods include (2.3): single linkage, complete linkage, and average linkage, which are described as follows (Podani, 2000; Bajorski, 2011; Hair et al., 2013):

- Single linkage algorithm (also called minimum distance or nearest neighbor algorithm) defines similarity as the distance between the two closest elements between two clusters;
- Complete linkage algorithm (also called maximum distance or farthest neighbor algorithm) defines similarity as the distance between the two farthest elements from the two clusters; and

- Average linkage algorithm (also called average distance or unweighted pair-group average method) defines similarity as the average distance over all possible pairs of elements from the two clusters. The average linkage method gives a more robust solution compared to the two other techniques (Sokal and Sneath, 1963; Johnson, 1967).

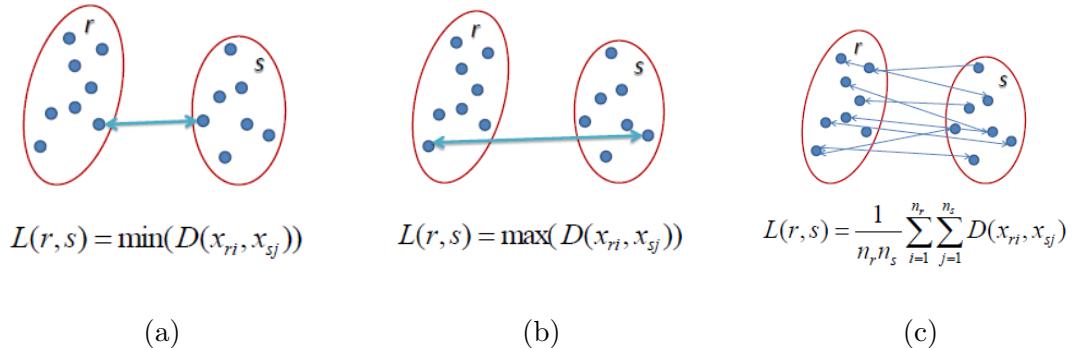


Fig. 2.3: Methods for computing the distance matrix: (a) single linkage; (b) complete linkage; (c) average linkage (Source: Sayed, 2016).

In Fig. 2.3, x pertains to a set of objects within clusters r and s , of which L is the distance between two clusters that may be defined as the shortest (single) or longest (complete) distance between two points in each cluster, or the average distance between each point in n number of points in one cluster to every point in the other cluster. A distance matrix D is calculated where the i th row and j th column is the distance between the i th and j th elements.

Hierarchical clustering approaches to data classification are well-established in the biological sciences (i.e., numerical taxonomy) and the humanities (e.g., sociology, psychology), of which the common application is classification according to the relationships among the data observations being clustered (Kaufman and Rousseeuw,

1990; Podani, 2000).

In this study, a hierarchical agglomerative clustering procedure was implemented to determine and visualise the “natural” hierarchical structure of forest types based on the data with the aim of defining the forest classification hierarchy in contrast to the heuristic approaches such as those implemented by Pierce et al. (1994) and Dobson et al. (1996) to define classification hierarchies. The dendograms from the cluster analyses were also compared to assess temporal consistency of the multi-year PALSAR mosaics.

2.5 Decision tree classification

Classification of SAR data using traditional statistical techniques makes use of the unique local statistics of each image, which in turn necessitates local training and validation of each image; hence limiting its application at regional or global scales (Dobson, Pierce, and Ulaby, 1996). Knowledge-based or decision tree approaches were found to be more suitable for classifying SAR data (Pierce et al., 1994; Dobson, Pierce, and Ulaby, 1996). This is because decision tree classifiers apply hierarchical decision rules to differentiate land cover classes, and these decision rules or classifiers, once trained, become applicable to other images irrespective of scale and time (Dobson, Pierce, and Ulaby, 1996).

The decision tree algorithm, or classification and regression trees (CART), is based on the seminal work of Breiman et al. (1984). The decision tree classifier is a simple, non-parametric, binary recursive partitioning procedure in the form of hierarchical rules that are successively applied to the input data or feature space

(Strobl, Malley, and Tutz, 2009). The goal of the recursive partitioning is to use a set of predictor variables to estimate the means of one or more response variables; thereby constructing a nested binary tree from the repeated splitting of the data into homogenous subsets (Michaelsen et al., 1994). The full dataset (parent node) is partitioned into two groups based on a single predictor variable. Each resulting subset (child node) is further partitioned such that the variability in the response variables are minimised and the nodes contain more homogenous observations (Berk, 2008). Once partitioning is complete, the means of the response variables in each terminal node serve as rules for predicting future observations (Michaelsen et al., 1994; Berk, 2008; Strobl, Malley, and Tutz, 2009).

Decision trees can grow to their maximum size since the splitting of each node continues recursively without a stopping criterion, and eventually the process stops until no further splits are possible due to lack of data (Wu and Kumar, 2009). Larger decision trees have fewer classification errors, hence less bias, but have terminal nodes with few observations, hence greater variance and instability (Berk, 2008). A pruning sequence is then implemented to find a sensible trade-off. Pruning, a process aimed at constraining the size of the tree to avoid overfitting the data, involves removing inefficient branches of the decision tree by consolidating nodes that do not reduce heterogeneity despite the added extra tree complexity (Berk, 2008; Strobl, Malley, and Tutz, 2009). An optimal tree is identified by evaluating the predictive performance of every tree in the pruning sequence either on independent test data or through cross-validation (Wu and Kumar, 2009). Cross-validation, as implemented in this study, involves partitioning the full dataset into subsets, excluding one subset, growing

the tree on the remaining data, and testing it on the excluded subset (Michaelsen et al., 1994). The chances of overfitting is minimised since essentially different data are used to train and test the decision trees.

Several examples have successfully demonstrated the use of decision tree classifiers for the classification of remotely sensed data. With optical data, decision trees have been used for generating global land cover classification map products from the AVHRR sensor (DeFries et al., 1998; Hansen, Defries, et al., 2000); for mapping land and forest cover change in Palawan and Eastern Mindanao in the Philippines from Landsat data (Pereira et al., 2006); and for mapping decadal mangrove cover change in the Philippines using Landsat (Long et al., 2013), among others.

With SAR data, Simard et al. (2000) used decision trees and multi-scale texture measures on the Japan Earth Resources Satellite (JERS-1) SAR data for thematic mapping of tropical forest in Gabon in Central Africa, of which decision trees resulted in improving classification results. Both JERS-1 and the European Resources Satellite (ERS-1) SAR composites were used with a decision tree classifier for mapping land cover in Michigan in the United States (Dobson, Pierce, and Ulaby, 1996), and for mapping coastal vegetation in Gabon in Central Africa (Simard, De Grandi, et al., 2002). Dong et al. (2014) mapped broad land cover categories (i.e., forest, cropland, urban, water) in Southeast Asia using dual-polarised PALSAR data and a decision tree classifier, and achieved high overall accuracy. The producer's and user's accuracies of forest were 86% and 93%, which was partly attributed the simple classification scheme.

Decision trees have a number of advantages over traditional classification methods.

Hansen et al. (1996) showed that decision trees slightly outperformed maximum likelihood algorithm for classifying monthly composites of AVHRR data. Moreover, decision trees provided an added advantage of reducing data dimensionality by revealing the hierarchical nature of predictor variables, specifically those metrics that contributed to discriminating among cover types. Friedl & Brodley (1997) found that decision tree algorithms consistently outperformed the maximum likelihood and linear discriminant classifiers in terms of classification accuracies applied to global AVHRR data. Pal & Mather (2003) showed that decision trees when applied to Landsat and hyperspectral data performed acceptably well compared to the maximum likelihood and artificial neural network classifiers, except with high-dimensional data. Otukei & Blaschke (2010) found that decision trees performed better than maximum likelihood and support vector machines for land cover change assessment using Landsat data. Mishra et al. (2011) evaluated different classification algorithms for classifying land cover using fully polarimetric ALOS/PALSAR data, and found that decision trees produced more accurate results than parallelepiped, minimum distance, maximum likelihood, and Wishart classifiers.

In this study, a decision tree classifier was used similarly due to its better performance over other classification algorithms, its suitability for classifying SAR data, and its potential applicability of decision rules to images over other areas once the classifier has been trained.

2.6 Texture: the grey-level co-occurrence matrices

Image texture refers to the spatial variability or the patterns of spatial relationships of grey levels (tone) within a neighbourhood of pixels, which is commonly described in terms of roughness or smoothness (Mather and Koch, 2011). Texture is one of the important elements of image interpretation that are used to describe the characteristics of features in remotely sensed images (Campbell and Wynne, 2011).

Previous research have utilised different texture information with SAR data for various aspects of land and forest cover type mapping such as improving the discrimination of forest types (Ulaby et al., 1986); detecting clearings in primary forest (Oliver, 2000); discriminating different subcategories of forest vegetation types (Saatchi et al., 2000); mapping of tropical forest (Simard, Saatchi, and De Grandi, 2000); discriminating inundated vegetation (Podest and Saatchi, 2002); increasing the accuracy of biomass estimation in tropical forests (Kuplich, Curran, and Atkinson, 2005); and improving the classification accuracies of broad land cover types (Walker et al., 2010; Longépé et al., 2011; Li et al., 2012; Rakwatin et al., 2012).

First- and second-order texture measures are among the descriptors that can be measured in an image. First-order texture measures calculate the spectral variability or statistical moments (e.g., mean, variance, kurtosis, skewness) within a certain window (group of pixels or kernel), of which the moments do not contain any direction (Kurvonon and Hallikainen, 1999). Second-order texture measures consider the spatial relationships between groups of two neighboring pixels to a certain direction within the window (Gallardo-Cruz et al., 2012). The calculation of second-order texture measures involves the construction of Grey-Level Co-occurrence Matrices (GLCM),

Texture variable	Formula	Description
Mean	$\text{MEAN} = \sum_{i,j=0}^{N-1} i P_{i,j}$	Mean of the probability values from the GLCM. It is directly related to the image spectral heterogeneity.
Variance	$\text{VAR} = \sum_{i,j=0}^{N-1} P_{i,j}(i - \text{MEAN})^2$	Measure of the global variation in the image. Large values denote high levels of spectral heterogeneity.
Correlation	$\text{COR} = \sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(i - \text{MEAN})(j - \text{MEAN})}{\text{VAR}} \right]$	Measure of the linear dependency between neighbouring pixels.
Contrast	$\text{CONT} = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$	Quadratic measure of the local variation in the image. High values indicate large differences between neighbouring pixels.
Dissimilarity	$\text{DISS} = \sum_{i,j=0}^{N-1} P_{i,j} i-j $	Linear measure of the local variation in the image.
Homogeneity	$\text{HOM} = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$	Measure of the uniformity of tones in the image. A concentration of high values along the GLCM diagonal denotes to a high homogeneity.
Angular second moment	$\text{ASM} = \sum_{i,j=0}^{N-1} P_{i,j}^2$	Measure of the order in the image. It is related to the energy required for arranging the elements in the system.
Entropy	$\text{ENT} = - \sum_{i,j=0}^{N-1} P_{i,j} \ln P_{i,j}$	Measure of the disorder in the image. It is inversely related to ASM.

The abbreviations, formulas and descriptions of the eight texture variables used to model successional vegetation attributes are presented. $P_{i,j}$ is the (i, j) element of the GLCM, and represents the probability of finding the reference pixel value i in combination with a neighbor pixel value j . Note that $\sum_{i,j} P_{i,j}=1$.
doi:10.1371/journal.pone.0030506.t003

Fig. 2.4: Texture variables derived from the grey-level co-occurrence matrix (Source: Gallardo-Cruz et al., 2012).

which are matrices containing the probabilities of co-occurrence of pixel values for pairs of pixels in a given direction and distance (Gallardo-Cruz et al., 2012).

The GLCM, a concept first described by Haralick et al. (1973), outlines the distance and angular relationships among a neighbourhood of pixels. It computes the joint probability of occurrence $p(i,j)$ of the pairs of grey levels i and j separated by a given distance dij and direction, and the image being initially quantised into N_g grey levels (Haralick, Shanmugam, and Dinstein, 1973). In this study, eight texture variables in three groups were used to describe the degree of contrast between pixels (contrast, dissimilarity, homogeneity), the regularity in the pixels within a neighbourhood of pixels (angular second moment, entropy), and the statistics derived from the GLCM (mean, correlation, variance) (Fig. 2.4; Gallardo-Cruz et al., 2012).

In this study, the texture measures listed in Fig. 2.4, in addition to radar backscatter and topographic variables, were evaluated whether they contributed to improving the classification and discrimination of forest cover types based on the FAO forest classification.

2.7 Geographic object-based image analysis

Geographic Object-Based Image Analysis (GEOBIA) is a recent sub-discipline of geographic information science (including its own theory, methods, and tools) that intends to develop automated approaches to partition remote sensing imagery into meaningful image-objects, and assess their characteristics through spatial, spectral, and temporal scales (Hay and Castilla, 2008). Its main objective is to generate new geographic information (in GIS-ready format) from which spatial knowledge can be obtained (Blaschke et al., 2014). It serves as a critical bridge to the raster (remote sensing) and vector (GIS) domains through the generation of image-objects, which represent meaningful geographic entities that can be distinguished in an image. Fundamentally, GEOBIA consists of image segmentation, classification of objects, and modeling based on characteristics of objects (Johansen, Bartolo, and Phinn, 2010).

Object-based approaches were developed to address the limitations of traditional pixel-based approaches for image analysis, classification, and interpretation. Burnett & Blaschke (2003) argued that pixel-based methods do not maximise the use of the spatial concepts of neighbourhood, proximity, or homogeneity for characterising land use/land cover using remotely sensed data. The pixel-centered approach, which adheres to the concept of the pixel as a spatial entity (Fisher, 1997), assumes the

pixel to be equivalent to objects in the landscape (Burnett and Blaschke, 2003). Wu (1999) suggested that the hierarchical patch dynamics framework can be used model landscapes as a hierarchical mosaic of patches such that ecological systems can be recognised as spatially nested patch hierarchies, wherein larger patches are composed of smaller patches. In simple terms, an image composed of pixels is partitioned into segments (or patches), and groups of related segments form larger segments (e.g., individual tree segments combine to form a forest segment).

In its broadest sense, image segmentation refers to the process of partitioning an image, which is basically an array of measurements, into multiple spatially discrete regions on the basis of spectral homogeneity (Blaschke et al., 2014; Canty, 2014). It is the first step in a processing workflow to derive meaningful image-objects. Multi-scale segmentation or object relationship modelling, the method developed by Burnett & Blaschke (2003) to derive meaningful objects from images, relies heavily on the hierarchical patch dynamics theory, image segmentation, and in defining the semantic rules to establish relationships of the lower landscape units to higher levels of organisation. Semantics based on descriptive assessment and knowledge is used in GEOBIA to translate spectral characteristics of image-objects to real-world features, such that groups of pixels are now put into context (Blaschke et al., 2014).

Object-based approaches have been used for land and forest cover mapping and characterisation, specifically using SAR data. Some examples of these applications include delineation of forest cover maps and detection of deforestation in Europe (Thiel et al., 2008); mapping changes in the largest continuous Amazonian mangrove belt (Nascimento et al., 2013); mapping mangroves in Guinea, West Africa (Flores

De Santiago, Kovacs, and Lafrance, 2013); and classifying wetland vegetation types (Hong et al., 2015). In some of these studies, implementing an object-based approach allowed the inclusion of other data sources such as digital elevation models (DEM) or texture measures, in addition to the SAR images themselves, to provide more information in aid of image analysis, classification, and interpretation. Others studies also combined of multi-sensor and multi-temporal SAR data such as for mapping land cover and seasonal inundation in the Brazilian Pantanal Wetland by using L-band ALOS/PALSAR and C-band RADARSAT-2 data (Evans, Costa, Telmer, et al., 2010; Evans and Costa, 2013; Evans, Costa, Tomas, et al., 2014).

A notable example that is useful for this present study is the paper by Walker et al. (2010), which compared the performance of Landsat and ALOS/PALSAR for large-area land and forest cover classification and mapping in the Brazilian Amazon. They designed an object-based image analysis approach combining imagery with information from elevation layers and feature extraction, and applied a decision tree classifier within the context of a hierarchical, multi-level classification scheme. Their study confirmed the potential of ALOS/PALSAR as an accurate source for spatially explicit estimates of forest cover, and pointed to the important role of spaceborne radar sensors in complementing optical remote sensing for forest cover mapping and monitoring.

For this study, an object-based approach was implemented, in contrast to traditional pixel-based approaches, since previous research have applied this on SAR data for forest cover mapping with better results and classification accuracies. One advantage of an object-based approach over a pixel-based approach is the inclusion

of ancillary datasets to improve the analysis and classification of remotely sensed data. For example, topographic data, such as elevation and slope, can be included to complement dual-polarised ALOS/PALSAR mosaic data to provide additional information that can aid in land and forest cover classification and mapping since topography influences forest typology.

2.8 Conceptual framework

This study is about the evaluation of multi-year ALOS/PALSAR mosaic data for mapping and monitoring of forest cover types in the Philippines. The conceptual framework of this study presents Forest as the central theme and the utilisation of remote sensing data and techniques for forest applications (Fig. 2.5). The four main elements of the study consist of Data, Method, Indicator, and Application. Data elements consist of SAR, DEM, and ancillary land cover maps. Method elements include FCCS, GEOBIA, GLCM, clustering, and CART. Indicator elements include consistency, hierarchy, and accuracy. Application elements consist of mapping, classification, and monitoring.

The Data-Method relationship describes the dataset as inputs to the methodology and the suite of methods implemented on the data. The Method-Indicator relationship describes the methodology for exploring these indicators and the indicators provide a measure of the efficacy of the methods in producing the intended result. The Indicator-Application relationship describes the quality of these indicators for the intended applications and the applications provide feedback to these indicators over time. The Data-Application describes the use of the data for the intended applications,

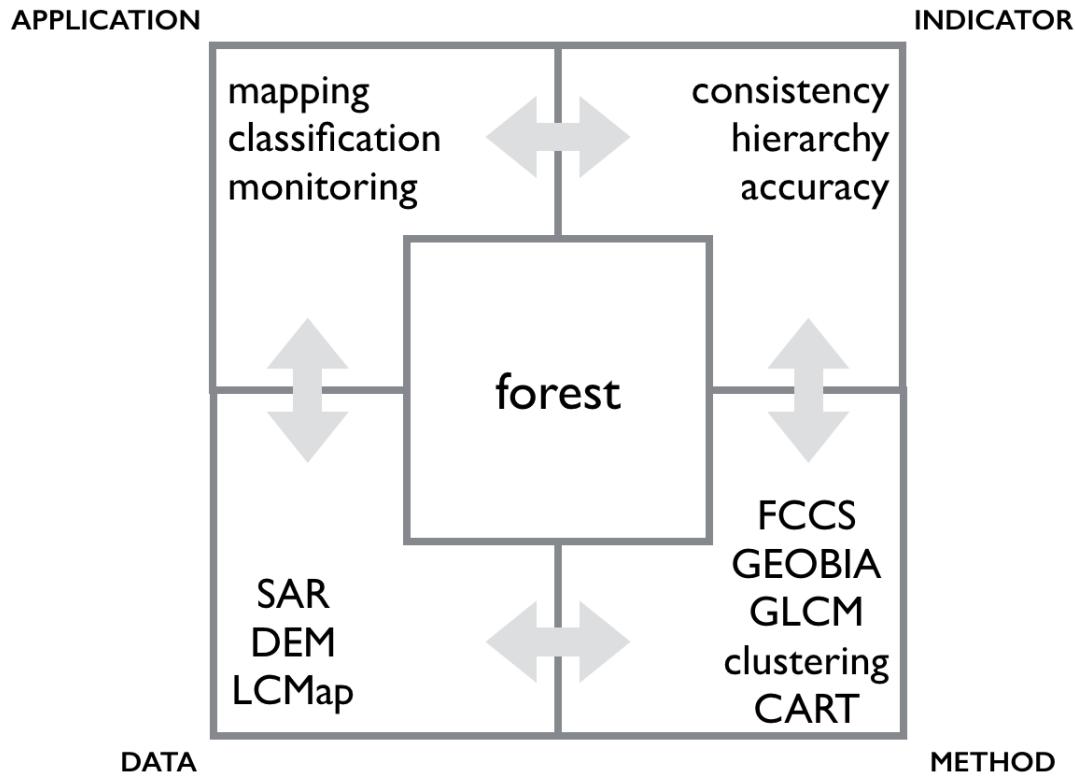


Fig. 2.5: The conceptual framework of the study shows forest as the central theme enclosed by elements concerning data, method, indicator, and application, and their interrelationships.

of which in turn also provide feedback on the applicability and utility of the data for the intended purpose.

2.9 Aims and objectives

Given the capabilities of ALOS/PALSAR data and above-mentioned image processing techniques, it is worth exploring their potentials for land and forest cover mapping in the Philippines. The general aim of this masters research is to evaluate the suitability of ALOS/PALSAR mosaic data for mapping and monitoring of forest

cover types in the Philippines based on the FAO forest cover classification system. To achieve this, the specific objectives of this thesis are as follows:

1. To ascertain whether ALOS/PALSAR mosaic data are temporally consistent within a single acquisition year and across multiple acquisition years for the purpose of periodic monitoring;
2. To establish a forest classification hierarchy based on the FAO forest classification system, and assess its consistency temporally and across feature attributes;
3. To assess discrimination of forest cover types using ALOS/PALSAR mosaic data based on polarisation, topographic, and texture characteristics;
4. To determine whether ancillary feature attributes (i.e., topographic, texture), in addition to polarisation, contribute to improving the classification accuracy of forest cover types; and
5. To develop a hierarchical, stepwise classification approach for assessing the degree at which ALOS/PALSAR mosaic data can discriminate between forest cover types.

2.10 Scope and limitations

Reference data used for this study, particularly as training and validation for the image classification, does not make use of ground-truth data. Rather, the study made use of the 2010 NAMRIA land cover map as the reference data. The official land cover map product was validated in the field by NAMRIA and fulfills acceptable classification

accuracies (Raul Magabo, 2015, *personal communication*), although documentation of the methods and accuracy assessment to accompany the map product has not yet been made publicly available. The multi-year ALOS/PALSAR mosaic data used in this study, which are different from the standard strip data product, were accessed through JAXA's Kyoto & Carbon (K&C) Initiative and were provided as is. The findings from this study may apply to the standard products, but may potentially differ unless the same pre-processing steps carried out on the mosaic data (Shimada and Ohtaki, 2010) are similarly applied on the standard products.

Chapter 3

Methodology

3.1 Study area

The region of Northern Luzon, Philippines (approx. geographic coordinates: upper-left: 120° E, 18° N; lower-right: 123° E, 16° N) was chosen as an appropriate study area because all nine forest cover types based on the FAO FCCS were present in the region. The study area covers 16 provinces, either fully or partially, including: Abra, Apayao, Aurora, Benguet, Cagayan, Ifugao, Ilocos Norte, Ilocos Sur, Isabela, Kalinga, La Union, Mountain Province, Nueva Ecija, Nueva Vizcaya, Pangasinan, and Quirino (Fig. 3.1). Northern Luzon experiences tropical cyclones very frequently, and spans three climate regimes based on the modified Coronas classification, particularly: Type I over Ilocos region (pronounced wet and dry season); Type III over Cagayan River Basin (short dry season with no pronounced maximum rain period); and Type IV over Northern Sierra Madre (no dry season and no pronounced maximum rain period).

The study area covers a total land area of 4.9 M ha, of which approximately 3.1 M ha (63%) is comprised of non-forest land cover and 1.8M ha (37%) is comprised of forest cover types based on the 2010 NAMRIA land cover map. Expansive agricultural areas are situated in the Cagayan Valley between the Cordilleras and Northern Sierra Madre ranges. Of the total land area of forest cover types, the majority is comprised of closed forest broadleaved (FCFB) at 38%; open forest broadleaved (FOFB) at 46%; open forest coniferous (FOFC) at 9%; and the remaining 7% by other forest cover types. A detailed tabulation of total land area of forest cover types per province is presented in Appendix B.

3.2 Data

For this study, the following datasets were used:

1. Dual-polarised 25 m spatial resolution ALOS/PALSAR mosaic data (HH, HV) from JAXA through the K&C Initiative (Rosenqvist, Ogawa, et al., 2001; Rosenqvist, Shimada, Lucas, et al., 2010). The PALSAR mosaics, provided as 1x1 degree tiles, are radiometrically and geometrically calibrated large-scale datasets generated using JAXA's mosaicking algorithm, which consists of the pre-processing the long-strip PALSAR data; orthorectification and slope correction using a DEM; suppression of differences in intensity between neighbouring strips; and metadata preparation (e.g., dates of launch, local incidence angle, radar shadow, layover, and valid/invalid data) to support data interpretation (Shimada and Ohtaki, 2010). Multi-year ALOS/PALSAR mosaic data (2007, 2008, 2009, 2010) were used to assess temporal consistency and suitability for periodic

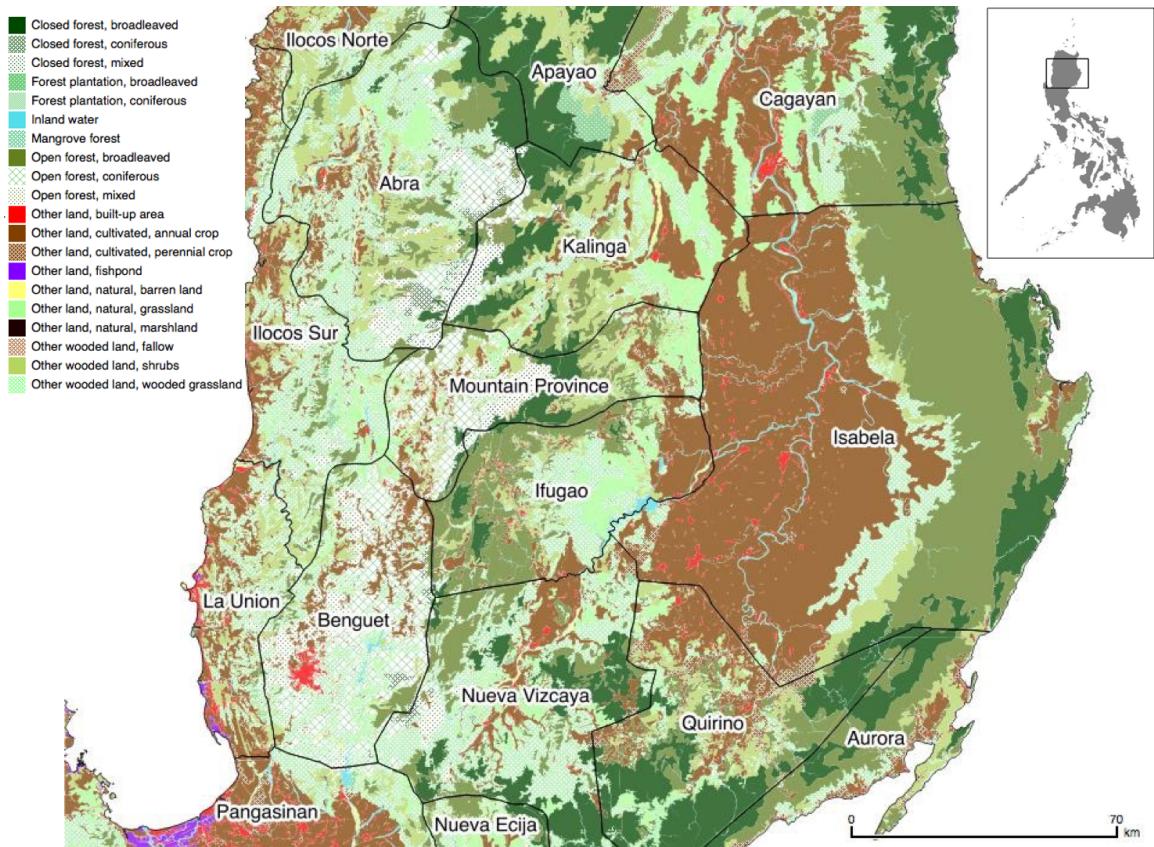


Fig. 3.1: Study area in Northern Luzon, Philippines showing the 2010 land cover map based on NAMRIA data.

monitoring of forest cover. A total of 24 ALOS/PALSAR mosaic tiles were used with six tiles per year (i.e., 17N120E, 17N121E, 17N122E, 18N120E, 18N121E, 18N122E). All PALSAR tiles consist two channels, the HH and HV polarisation. The metadata layers, specifically the date of acquisition layer and the mask layer (for excluding radar shadow and layover pixels), corresponding to each mosaic tile per year were also used.

2. Shuttle Radar Topography Mission (SRTM) elevation data product available from the United States Geological Survey. The SRTM DEM 1 arc-second (~ 30 m spatial resolution) global void-filled data was downloaded through the EarthExplorer portal. Nine DEM tiles were used (i.e., 16N120E, 16N121E, 16N122E, 17N120E, 17N121E, 17N122E, 18N120E, 18N121E, 18N122E).
3. National 2010 NAMRIA land cover map. The data in vector format was accessed through the Forest Management Bureau. The land cover and forest type classes were based on the FAO GFRA categories (Table 2.1) composed of a total of 20 land cover classes consisting of nine forest types.

3.3 Overall workflow

The primary steps in the overall image processing workflow, which are discussed in the proceeding sub-sections, consists of image pre-processing (Sec. 3.4), identification of regions of interest (Sec. 3.6), image segmentation (Sec. 3.7), extraction of feature attributes and backscatter analysis (Sec. 3.8), hierarchical clustering (Sec. 3.9), and decision tree classification and ruleset development (Sec. 3.10). The overall workflow

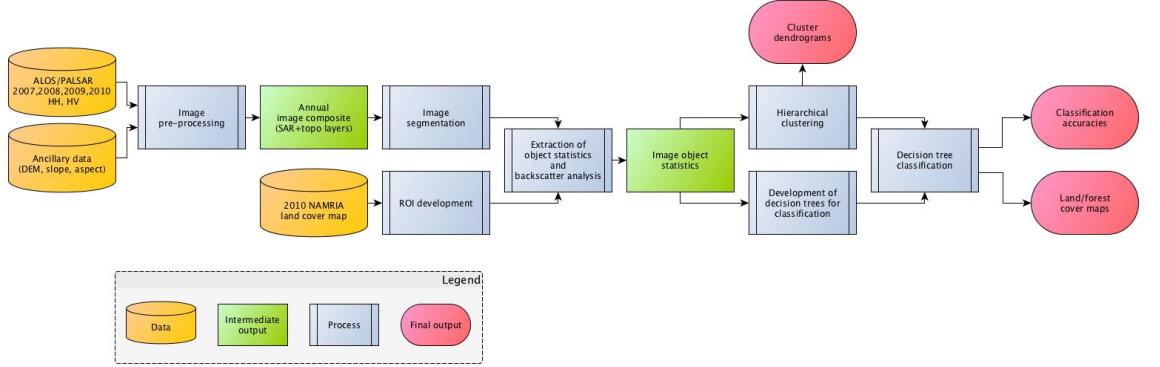


Fig. 3.2: Overall image processing workflow implemented in the study.

diagram shows the data used, intermediate outputs, processes implemented, and the final outputs (Fig. 3.2).

3.4 Image pre-processing

Image pre-processing steps, which are presented as a sub-workflow diagram in Fig. 3.3, were carried out in ENVI v.4.8 software (Exelis VIS, Inc. USA). For SAR data, each PALSAR data layer was mosaicked into a single regional image prior to geocoding and re-projection to UTM Zone 51 North coordinate system and WGS84 datum. The pixel dimensions of each image were 12844 columns x 8953 rows. To reduce signal noise from the radar data, a Lee filter was applied to preserve image sharpness and detail while suppressing speckle noise (Lee, 1980; Lee, 1981), and with a small 5x5 kernel window to preserve texture information (Lee et al., 1994). After minimising the speckles, ratio images (HH/HV) were generated using the HH and HV regional mosaics for each year. Then, HH and HV polarisation images were converted from amplitude to normalised radar cross-section, or sigma-nought (σ^0 ; units in dB)

using Eq. 3.1:

$$\sigma^0 = 10 \cdot \log_{10}[DN^2] + CF \quad (3.1)$$

Where DN is the digital number and CF is the calibration factor with a value of 83.0 dB for ALOS/PALSAR data (Rosenqvist, Shimada, Ito, et al., 2007). A masking process was implemented on the PALSAR regional mosaics using the corresponding mask layers for each year to exclude unwanted pixels affected by shadow and layover in the radar data.

For the topographic data, the SRTM DEM tiles were mosaicked together prior to re-projection to UTM Zone 51 North WGS84, and subsequently resampled to 25 m spatial resolution and cropped to match the extents and dimensions of the SAR mosaic images. The slope and aspect layers were subsequently computed from the mosaicked DEM. Lower kernel sizes (i.e., 3x3, 5x5, and 7x7) were initially tested but the resulting models contained artifacts; hence, a 9x9 kernel size was used to generate the final slope (in percent) and aspect models.

3.5 Design of a multi-level classification hierarchy

A multi-level land cover classification hierarchy was designed to guide the selection of regions of interest (ROI) and the decision tree classification (Fig. 3.4). The first three levels were designed to classify the image into broad-level classes and extract the Forest class. Within the Forest class, after setting aside all Non-Forest class, the class hierarchy of subsequent levels was defined based on the result of the hierarchical

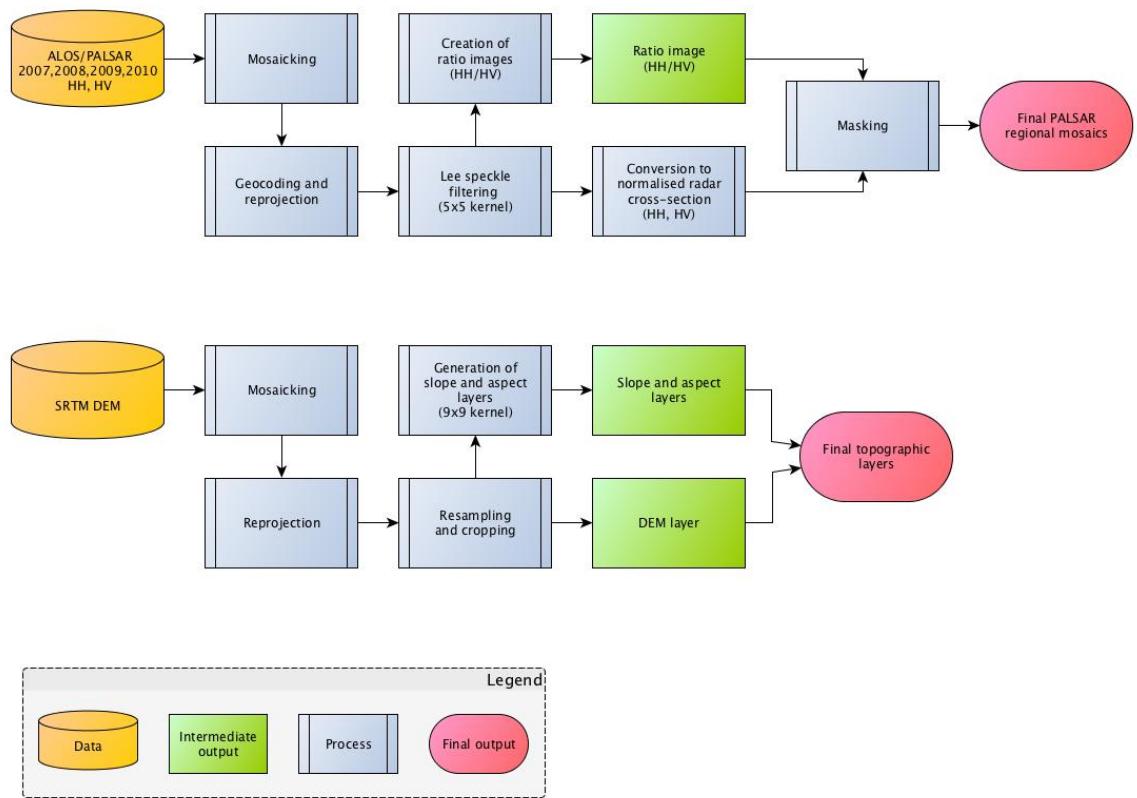


Fig. 3.3: Detailed sub-workflow for image pre-processing steps implemented on SAR and topographic datasets.

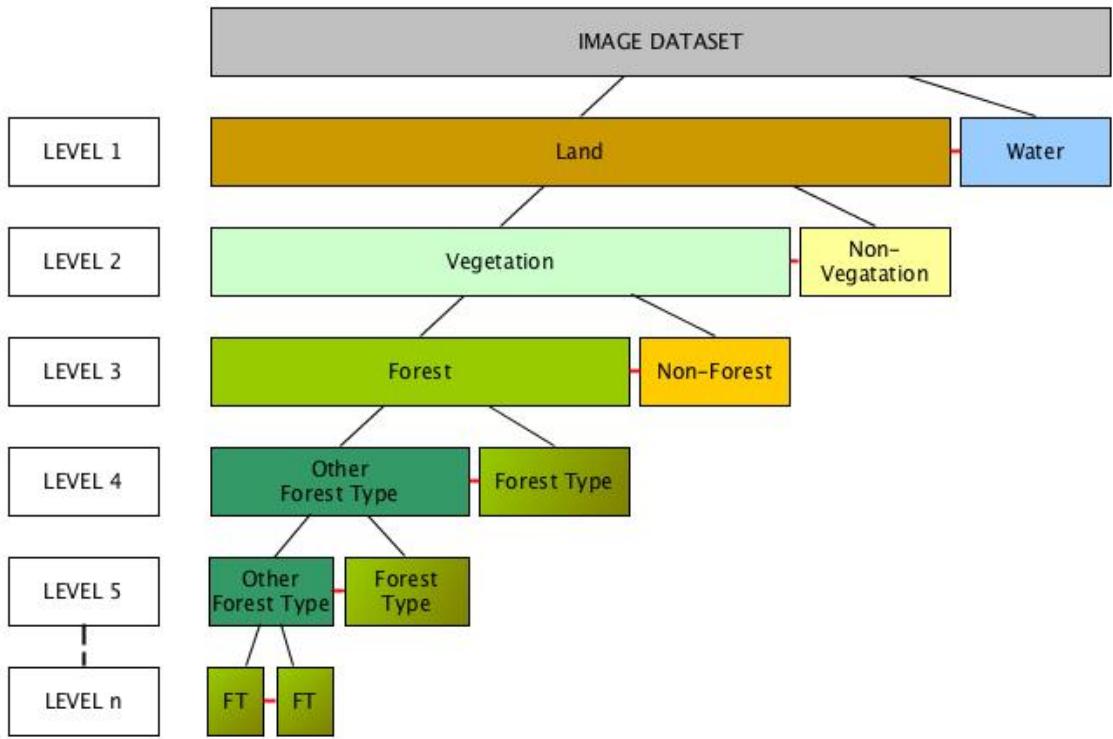


Fig. 3.4: The classification hierarchy proposed for this study (note: FT is Forest Type).

clustering.

Starting with the entire image dataset, each level bifurcated into two classes and carried out sequentially one level after another. At Level 1, the data was split into either Land or Water classes. Then at Level 2, Water classes were set aside, and the Land class was assigned into either Vegetation or Non-Vegetation classes. At Level 3, the Vegetation class was split into either Forest or Non-Forest classes while the Non-Vegetation class was set aside. At Level 4, the splitting into forest types was carried out following the dendrograms resulting from the hierarchical cluster analysis.

3.6 Identification of regions of interest

The identification of the ROIs was carried out in Quantum GIS v.2.12 software (QGIS Project Team). Prior to delineating ROIs, the detailed land cover classes were grouped into broad-level classes based on the multi-level classification hierarchy (Fig. 3.4; Table 3.1). To aid in identifying and delineating the ROIs, the 2010 NAMRIA land cover map was used as a base map. The PALSAR mosaics were also visually inspected during the delineation of each ROI before assigning it to a specific class and to ensure a 5x5 kernel size polygon. Other ancillary vector data were also used to aid in identifying the ROIs, including OpenStreetMap data such as roads, buildings, and rivers (Geofabrik GmbH and OpenStreetMap Contributors, 2015); and the Landsat-based 2010 mangrove map of the Philippines developed by Long et al. (2013) to aid in identifying regions of interest in mangrove forests. A total of 1,451 ROIs were identified representing the land cover classes, and each ROI record reflected the multi-level classification hierarchy in the attribute table (Table 3.1). The distribution of the ROIs selected over the study area is shown in Fig. 3.5.

Table 3.1: Hierarchy and grouping of detailed land cover classes into broad-level classes.

Level 1	Level 2	Level 3	2010 NAMRIA Land Cover Classes	Number of ROIs
Land	Vegetation	Forest	Closed forest, broadleaved (FCFB)	100
			Closed forest, coniferous (FCFC)	100
			Closed forest, mixed (FCFM)	100
			Open forest, broadleaved (FOFB)	186
			Open forest, coniferous (FOFC)	101
			Open forest, mixed (FOFM)	100
			Forest plantation, broadleaved (FFPB)	90

Table 3.1: Hierarchy and grouping of detailed land cover classes into broad-level classes.

Level 1	Level 2	Level 3	2010 NAMRIA Land Cover Classes	Number of ROIs
			Forest plantation, coniferous (FFPC)	30
			Mangrove forest (LVMG)	48
	Non-Forest	Other wooded land, shrubs		317
		Other wooded land, wooded grassland		
		Other land, grassland		
		Other land, annual crop		
		Other land, perennial crop		
Non-Vegetation		Other wooded land, fallow		67
		Other land, built-up area		
		Other land, barren land		
Water		Inland water		212
		Fishpond		

3.7 Image segmentation

Image segmentation was implemented in eCognition Developer v.8.9 software (Trimble Germany GmbH). The segmentation method that is implemented in eCognition is an iterative process that minimises the average heterogeneity of generated segments (Baatz and Schäpe, 2000). The measure of heterogeneity consists of a spectral and spatial component (Benz et al., 2004; Happ et al., 2010). Spectral heterogeneity is defined on the values of the spectral responses of the pixels within the segments. Spatial heterogeneity is based on two shape attributes: smoothness and compactness of the segments. Another parameter, scale, is the stopping criterion for the optimisation process. Prior to fusion of two adjacent objects, the resulting

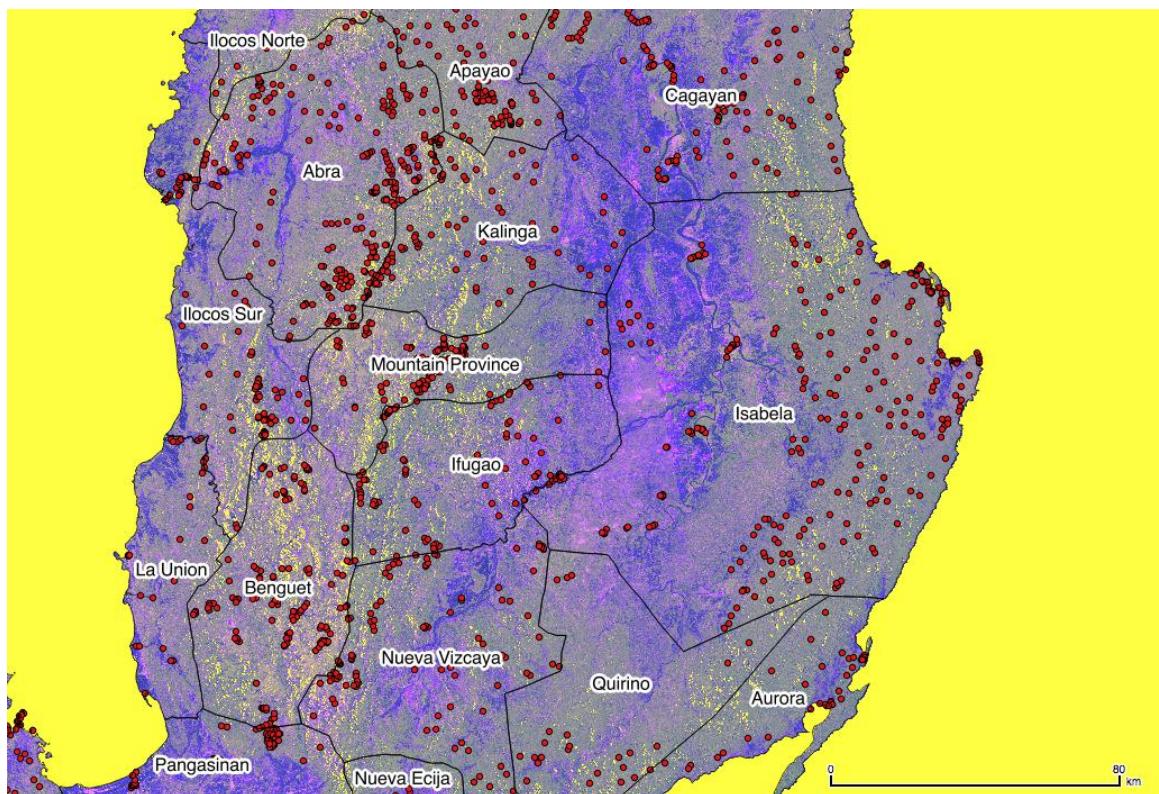


Fig. 3.5: The distribution of the regions of interest (red points) selected over the study area in Northern Luzon, Philippines. The 2010 PALSAR mosaic image is shown in the background.

increase in heterogeneity is computed such that if the resulting increase exceeds a threshold determined by the scale parameter then no further fusion takes place and the segmentation stops (Benz et al., 2004). A merge with a better degree of fitting (i.e., smaller value) than the scale parameter fulfils the homogeneity criterion (Baatz and Schäpe, 2000). Hence, the larger the scale parameter, the more objects can be fused and the larger the segments grow (Baatz and Schäpe, 2000).

The input data for segmentation consisted of the set of images, including the 12 ALOS/PALSAR data layers (i.e., HH and HV polarisation and HH/HV ratio images for each year) and 3 topographic data layers (i.e, elevation, slope, and aspect), totalling 15 input image layers. To determine the values of the parameters to be used for image segmentation procedure, different values were tested on each parameter to see whether the resulting segments made an acceptable delineation of image features (Appendix B). For multi-resolution segmentation, the user sets the following parameters: scale, shape, compactness, and image weights. The values of 5, 10, and 20 were tested for the scale parameter. For shape and compactness, the values of 0.1, 0.5, and 0.9 were used for each parameter, thereby testing nine combinations of values between shape and compactness (i.e., 0.1-0.1; 0.1- 0.5; 0.1-0.9; 0.5-0.1; 0.5-0.5; 0.5-0.9; 0.9-0.1; 0.9-0.5; and 0.9-0.9). For image weights, a value of 1 was used uniformly on all layers, both for the SAR and topographic layers.

After a visual assessment of the segmentation tests, a scale parameter value of 10 was chosen since a value of 5 was too detailed and a value of 20 had generalised small features. Next, a shape value of 0.1 and a compactness value of 0.9 were selected since the segments closely followed the form of image features. For image weights,

a value of 1 was retained for SAR data layers only. Topographic layers were given image weight values of 0 since they introduced regular-shaped rectangular segments that were not capturing the form of image features. The final values chosen for the multi-resolution segmentation parameters were: scale = 10; shape = 0.1; compactness = 0.9; image weights = 1 for SAR layers and 0 for topographic layers (Appendix B). The total number of segments delineated for each PALSAR regional mosaic were as follows: 2007 = 182,876 segments; 2008 = 184,917 segments; 2009 = 186,369 segments; 2010 = 191,588 segments.

3.8 Extraction of feature attributes and backscatter analysis

Object-level feature attributes were computed for all the delineated land cover ROIs in eCognition software. Feature attributes were calculated, particularly the mean and standard deviation, both from the yearly PALSAR regional mosaics and topographic layers; and eight GLCM measures from each of the PALSAR data layers. The GLCM texture measures include angular second moment, contrast, correlation, dissimilarity, entropy, homogeneity, mean, and standard deviation (variance). Thirty feature attributes were extracted from the SAR data layers per year (including polarisation and texture attributes) and 6 attributes from the topographic layers (Table 3.2). Tabular databases were generated containing these feature attributes such that four tables were produced, each corresponding to the attributes per annual PALSAR regional mosaics plus the topographic layer attributes. These tabular databases

were then used as inputs for subsequent backscatter analysis, the hierarchical clustering, and the decision tree classification in R statistical software (R Core Team, 2016).

Table 3.2: Object-level feature attributes extracted from the images.

Polarisation	Topographic	Texture		
		GLCM HH	GLCM HV	GLCM HH/HV
Mean HH	Mean DEM	Ang. 2 nd Mom	Ang. 2 nd Mom	Ang. 2 nd Mom
Mean HV	Mean Slope	Contrast	Contrast	Contrast
Mean HH/HV	Mean Aspect	Correlation	Correlation	Correlation
SD HH	SD DEM	Dissimilarity	Dissimilarity	Dissimilarity
SD HV	SD Slope	Entropy	Entropy	Entropy
SD HH/HV	SD Aspect	Homogeneity	Homogeneity	Homogeneity
		Mean	Mean	Mean
		SD	SD	SD

Note: HH, horizontal transmit - horizontal receive; HV, horizontal transmit - vertical receive; GLCM, grey level co-occurrence matrix; DEM, digital elevation model; SD, standard deviation.

PALSAR mosaic data are comprised of strip (or path) data acquired on different dates that are pre-processed and mosaicked together as end-user products (Fig. 3.6; Shimada and Ohtaki, 2010). The strip data comprising one annual regional PALSAR mosaic were acquired over different seasons/dates. Also, the strip data along the same path acquired over the same locality were acquired over different years for the annual mosaics. To assess the temporal consistency of the PALSAR mosaic data, a radar backscatter analysis of forest cover types was done for PALSAR mosaics, first, across adjacent strips within a single acquisition year; and second, within the same strip across multiple acquisition years. Box-whisker plots were generated using the *ggplot2*

package in R software (Wickham and Chang, 2015) to analyse the backscatter values of different forest types over the different dates of acquisition of the images.

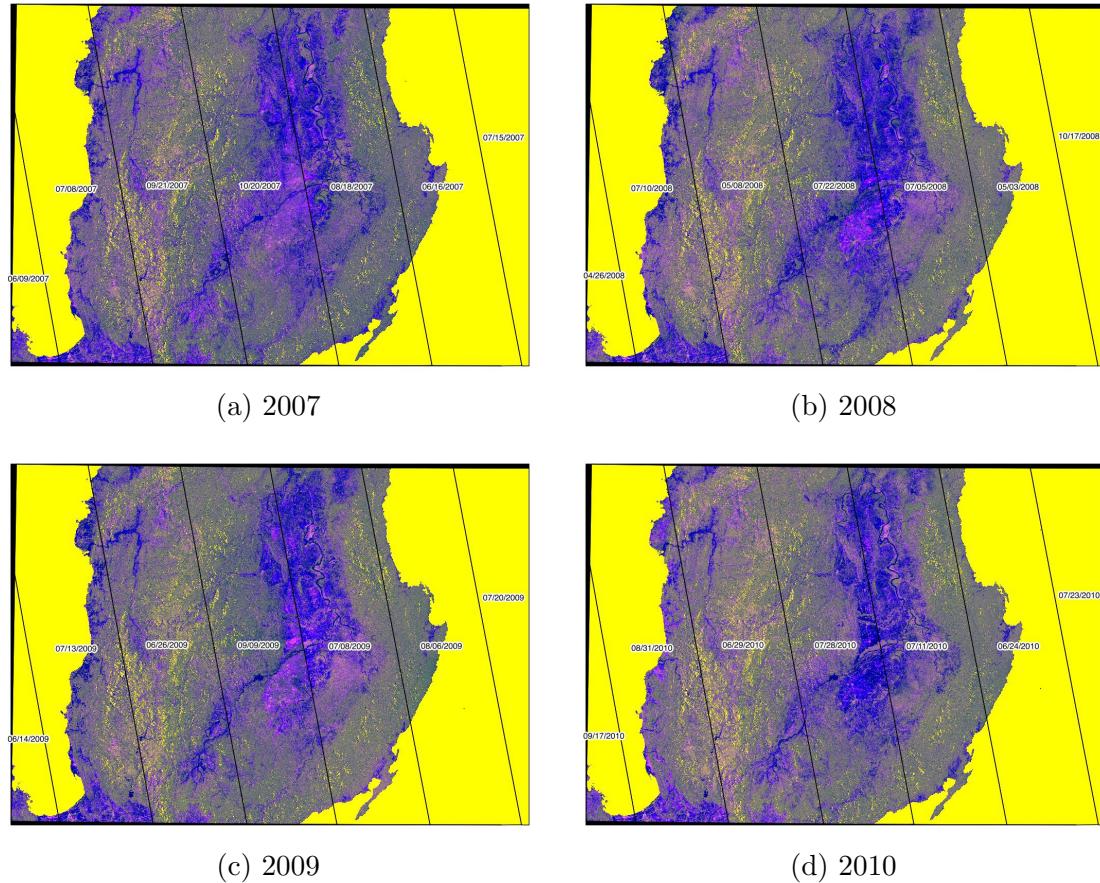


Fig. 3.6: The annual PALSAR mosaics of Northern Luzon acquired in (a) 2007, (b) 2008, (c) 2009, and (d) 2010. The extent of each strip (or path) data and the dates of acquisition of each strip are shown for each annual mosaic.

3.9 Hierarchical clustering

Agglomerative hierarchical clustering was implemented using the *pvclust* package in R software (Suzuki and Shimodaira, 2006; Suzuki and Shimodaira, 2014) with the tabular databases for each year as inputs. Both the temporal consistency of the multi-year SAR data and the cluster hierarchy of forest types were evaluated using the dendrograms. The *pvclust* package was useful to assess the uncertainty or stability in hierarchical cluster analysis by calculating probability values (Suzuki and Shimodaira, 2014). An approximately unbiased (AU) p-value ≥ 0.95 indicates consistent or stable clustering/grouping and is strongly supported by the data.

Table 3.3: The four cases representing combinations of object-level feature attributes.

Case	Number of variables	Combinations of object-level feature attributes
1	6	Polarisation
2	12	Polarisation + Topographic
3	30	Polarisation + Texture
4	36	Polarisation + Topographic + Texture

Four cases representing combinations of object-level feature attributes were used in evaluating the temporal consistency of the cluster hierarchy of forest types (Table 3.3). Hierarchical clustering was applied across the four cases with each case involving the four annual PALSAR mosaics. For Case 1, only the Polarisation feature attributes were evaluated. For Case 2, the Polarisation and Topographic feature attributes were evaluated. For Case 3, the Polarisation and Texture feature attributes were evaluated. Finally, for Case 4, the all feature attributes were evaluated.

In any case, consistency of the cluster hierarchy means that the hierarchical structure of the dendrogram does not change and the order and location of the nodes represented by forest types do not vary. The consistency of dendrograms was assessed through pairwise comparisons of dendrograms, totalling six pairs in each case. Baker's Gamma Indices were computed using the *dendextend* package in R software (Galili et al., 2015; Galili, 2015) for comparing clustering dendrograms. The Bakers Gamma Index (referred to as β coefficient from here on) is a measure of association (similarity) between two dendrograms of hierarchical clustering (Baker, 1974). It is defined as a rank correlation between the stages at which pairs of objects combine in each of the two dendrograms (Fowlkes and Mallows, 1983). A Spearman's correlation is computed and the values can range between -1 to 1 with near 0 values meaning that the two dendrograms are not statistically similar (Galili et al., 2015). The Baker's Gamma Indices for pairwise comparison of dendrograms for all cases were summarised and ranked to determine which case (or combination of feature attributes) gave the higher similarity between dendrograms.

3.10 Decision tree classification and ruleset development

A decision tree classification approach using the *tree* package in R software (Ripley, 2015), which is based on the CART algorithm by Breiman et al. (1984), was used to construct decision tree models. The tabular databases were used as the inputs for decision tree classification, designating the feature attributes as predictor variables

and assigning classes depending on the classification level. At Level 1 classification, all objects were assigned into either Land or Water classes. Then at Level 2, Water classes were set aside, and only objects classified as Land from Level 1 were subsequently assigned into either Vegetation or Non-Vegetation classes. At Level 3, only objects classified as Vegetation from Level 2 are assigned into either Forest or Non-Forest classes while setting aside Non-Vegetation classes. In subsequent classification levels, forest cover types were classified one by one based the results of the hierarchical clustering until all forest cover types were classified.

Decision tree models were developed based on four cases or combinations of feature attributes to determine whether ancillary feature attributes (i.e., topographic, texture), in addition to polarisation, contribute to improving the classification accuracy of forest cover types. All decision trees were pruned using k-fold cross-validation to prevent overfitting. The case yielding the lowest misclassification error rates was selected and used as the basis for generating classification trees for each level for each year. A total of 44 classification trees were generated consisting of 11 trees in each year. The decision tree models produced in R software were translated as the bases for constructing the multi-level classification rulesets to produce forest cover maps in eCognition software.

An additional assessment of classification accuracies was done to discriminate closed and open canopy forests without following the hierarchical clustering result. The first three classification levels were maintained. At Level 4, mangrove forest (LVMG) was separated from all other forest types. For the final classification at Level 5, closed and open canopy forests were separated. Closed canopy forests included:

closed broadleaved (FCFB), closed coniferous (FCFC), and closed mixed (FCFM). Open canopy forests included: open broadleaved (FOFB), open coniferous (FOFC), open mixed (FOFM), broadleaved forest plantation (FFPB), and coniferous forest plantation (FFPC).

Chapter 4

Results

4.1 Temporal consistency of PALSAR mosaic data

Two sets of figures were used for analysing the distribution of radar backscatter vis-a-vis data acquisition dates. The first set consists of PALSAR mosaic data that were acquired in a single year but the data strips used to put the mosaic together were taken at different seasons; and the second consists of PALSAR mosaic data acquired along the same data strip or path but at different years.

4.1.1 Mosaic data acquired in a single year but at different seasons

Figs. 4.1 and 4.2 shows the distribution of backscatter values of forest cover types from PALSAR mosaic data acquired in a single year but at different seasons. The backscatter values of different forest types were plotted with respect to polarisation.

Each mosaic is comprised of six strip data paths, particularly where there is land.

As an example, the figures in Fig. 4.3 show the 2007 PALSAR mosaic image and the dates of acquisition of each strip/path that comprise the whole mosaic (top), and the distribution of mean HH polarisation backscatter of each forest type based on the ROIs on the 2007 mosaic image (bottom). The images were taken from Fig. 3.6 and Figs. 4.1 and 4.2, respectively. In the mosaic image, the dates of acquisition of the strips from left to right were taken at different months of the year, specifically: Jun 09, Jul 08, Sep 21, Oct 20, and Jun 16. Then, for the boxplots, the mean HH backscatter (y-axis) was plotted from the ROIs of each of the different forest cover types (x-axis), and under each forest cover type there can be a number of boxplots where each colour denotes the date of acquisition of the corresponding PALSAR strip data. For each box-whisker plot, the upper and lower hinges of the box correspond to the 1st and 3rd quartiles (25th and 75th percentiles). The band inside the box corresponds to the median (or the 2nd quartile). The upper whisker extends from the hinge to the highest value that is within $1.5 \times \text{IQR}$ of the hinge, where IQR is the inter-quartile range, or distance between the 1st and 3rd quartiles. The lower whisker extends from the hinge to the lowest value within $1.5 \times \text{IQR}$ of the hinge. Data beyond the end of the whiskers are outliers and are plotted as points.

The range of mean HH backscatter values fell between -2.50 dB and -11.25 dB across all forest cover types. It can be seen that for broadleaved forests, including closed and open canopy and forest plantation (FCFB, FOFB, FFPB), the mean HH backscatter values were recorded over four data strips/paths. Forest cover types with ROI samples recorded in three data strips/paths include open canopy coniferous

forest (FOFC) and open canopy mixed forest (FOFM). The other forest types were distributed along only two data strips/paths, particularly: closed canopy coniferous forest (FCFC), closed canopy mixed forest (FCFM), and mangrove forest (LVMG), of which mangroves were only found along strip path data over coastlines. It should be noted that ROI samples for mangrove forest along the far left data strip/path (i.e., Jun 09) were not sufficient for plotting, hence the box-whisker plot was truncated. And finally, the coniferous forest plantation (FCFP) was recorded only within one data strip/path (i.e., Sep 21).

For FCFB, the ranges of mean backscatter values as shown by the boxplots generally overlapped (i.e., Jun 16, Aug 18, Sep 21), and hence can be considered temporally consistent, although the Oct 20 boxplot was slightly greater than the other plots. For FCFC and FCFM, boxplots denoting two adjacent data strips (i.e., Sep 21, Oct 20) also overlapped; hence their mean backscatter values can be treated as temporally consistent. For FFPB, three boxplots overlapped (i.e., Jul 08, Sep 21, Oct 20), which denote that their mean backscatter values can be considered temporally consistent. The Aug 18 boxplot shows lower mean backscatter, which can be attributed to low backscatter returning to the antenna due to karst substrate (i.e., Penablanca Protected Landscape in Tuguegarao, Cagayan province) upon checking the locations of the ROIs found along the data strip. For FFPC, mean backscatter values were recorded on only one data strip (i.e., Sep 21). For FOFB, FOFC, and FOFM, all boxplots overlapped, indicating that mean backscatter values acquired over different months were within the same range, and hence were temporally consistent. For LVMG, the two boxplots overlapped (i.e., Jun 16, Aug 18), and hence their mean backscatter

values can be considered as temporally consistent. Overall, the general overlaps shown by boxplots of each forest cover type suggest that the range of mean HH backscatter values over PALSAR mosaic data strips acquired in different months can be treated as temporally consistent.

Now looking at Figs. 4.1 and 4.2, three plots are shown for each of the four years (hence, a total of 12 plots) that correspond to boxplots of mean HH backscatter, mean HV backscatter, and HH/HV ratio across the different forest cover types.

The mean HH backscatter of most forest types across four years (2007-2010) was approximately between -2.50 dB to -11.25 dB in terms of their minimum and maximum values, indicating values were generally consistent across different seasons. For mangrove forest (LVMG), the backscatter values were approximately below -5.0 dB. Broadleaved forests, both closed and open canopy (FCFB, FOBF), exhibited narrow ranges denoting lower variability, and their median values were also generally consistent at the same level and did not vary across different seasons or different years.

The mean HV backscatter of most forest types across four years was approximately between -7.50 dB to -17.50 dB in terms of their minimum and maximum values, also indicating that values were generally consistent across different seasons within mosaic image acquired within a single year. Mangrove forest (LVMG) backscatter values were approximately between -11.0 dB and -17.5 dB, and generally had lower backscatter values compared to other forest cover types (notably the 25th to 75th percentile range of boxplots). Similar to mean HH backscatter, broadleaved forests, both closed and open canopy (FCFB, FOBF), also exhibited narrow ranges denoting lower variability, and their median values were also generally consistent at the same level and did not

vary across different seasons or different years.

For the mean HH/HV ratio plots, values were approximately from 1.5 to 2.5 for all forest types (excluding outliers). Mangrove forest (LVMG) exhibited higher median values, generally observed at 2.0 and above, compared to other forest types that had values from 1.5 to 2.0.

4.1.2 Mosaic data acquired along one path of strip data but across different years

Figure 13 shows the distribution of backscatter values of forest cover types from PALSAR mosaic data from only one path of strip data but across different years. Similar to Figure 12, the backscatter values of different forest types were also plotted with respect to polarisation, and each mosaic is comprised of six strip data paths, particularly where there is land.

As an example, the figures in Fig. 4.4 show the one of the PALSAR regional mosaics with a highlight along the specific strip/path where ROIs of forest cover types were taken across different years, specifically from 2007 to 2010 (top), and the distribution of mean HH polarisation backscatter of each forest type along the same strip/path at different years (bottom). The images were taken from Fig. 3.6 and Figure 13, respectively. The dates of acquisition along the same strip/path taken at different years were: 21 Sep 2007, 08 May 2008, 26 Jun 2009, and 29 Jun 2010. Then, for the boxplots, the mean HH backscatter (y-axis) was plotted from the ROIs of each of the different forest cover types (x-axis), except for mangrove forest (LVMG) since there were no mangroves present within the data strip. Under each forest cover type

there can be a number of boxplots where each colour denotes the date of acquisition along the same strip at different years.

The range of mean HH backscatter values fell between -2.50 dB and -11.25 dB across all forest cover types across different years. Median values of mean HH backscatter at different years consistently overlapped for each forest cover type, except for some very slight fluctuations. Overall, the overlaps shown by boxplots of each forest cover type suggest that the range of mean HH backscatter values over the same PALSAR mosaic data strip acquired in different years can be considered as temporally consistent.

Now looking at Figure 13, three plots are shown for each of the two highlighted data strips/paths that correspond to boxplots of mean HH backscatter, mean HV backscatter, and HH/HV ratio across the different forest cover types. Backscatter values were assessed for two different data strips/paths: one in Cordillera (Strip 1) and the other in Sierra Madre (Strip 2). Eight forest types were present along the Cordillera strip (except mangrove forest) while four forest types were present in the Sierra Madre strip (except coniferous and mixed forest types).

For the Cordillera data strip (Strip 1), the mean HH backscatter values were approximately from -2.5 dB to -10.0 dB, while the mean HV backscatter values were approximately from -7.5 dB to -15.0 dB. Median values were at consistent levels across different years for each forest type. The mean HH/HV ratio values were generally from 1.0 to 2.0 with consistent median values across different years for each forest type. For the Sierra Madre data strip (Strip 2), the mean HH backscatter values were approximately from -5.0 dB to -10.0 dB, while the mean HV backscatter values were

approximately from -10.0 dB to -17.5 dB, of which mangrove forest (LVMG) had lower values compared to broadleaved forest types. Median values were generally at consistent levels across different years for each forest type. The mean HH/HV ratio values were also generally from 1.0 to 2.0, except for mangrove forest with values of 2.0 and above.

Overall, the distribution of mean backscatter of forest cover types for both HH and HV polarisation overlapped. The values of forest cover types in HH/HV ratio also overlapped, except for mangrove forest (LVMG), which had slightly higher values compared to other forest types. This emphasises that these forest types cannot be discriminated using solely polarimetric data.

The backscatter analysis demonstrated that the ALOS/PALSAR mosaics were temporally consistent despite the difference in seasons of acquisition dates of the strip data (usually within a range of 3-5 months) or the difference in annual dates of acquisition.

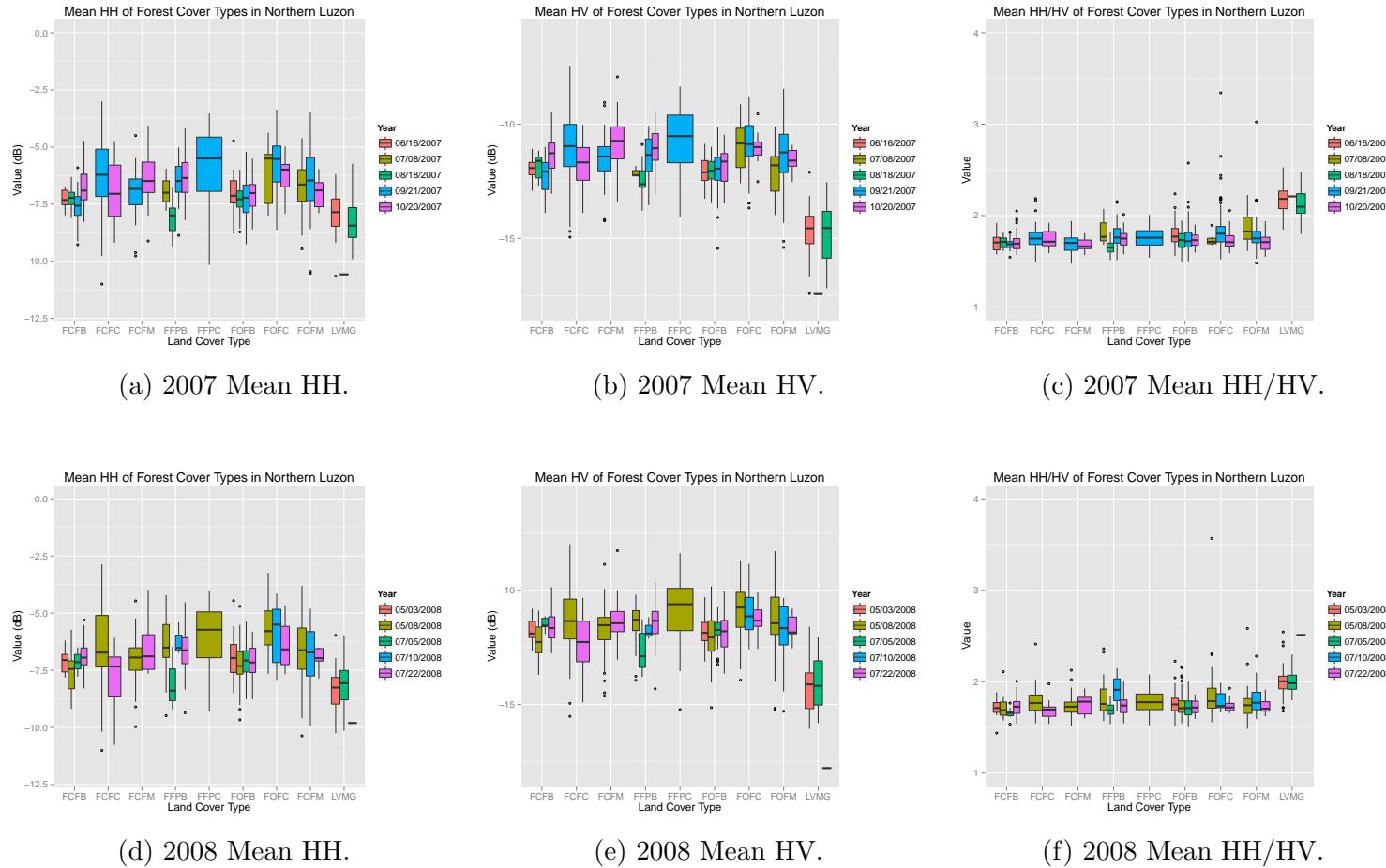


Fig. 4.1: Backscatter values of forest types from PALSAR data acquired in a single year at different seasons. Shown are box-whisker plots of mean backscatter values in 2007 (a-c) and 2008 (d-f).

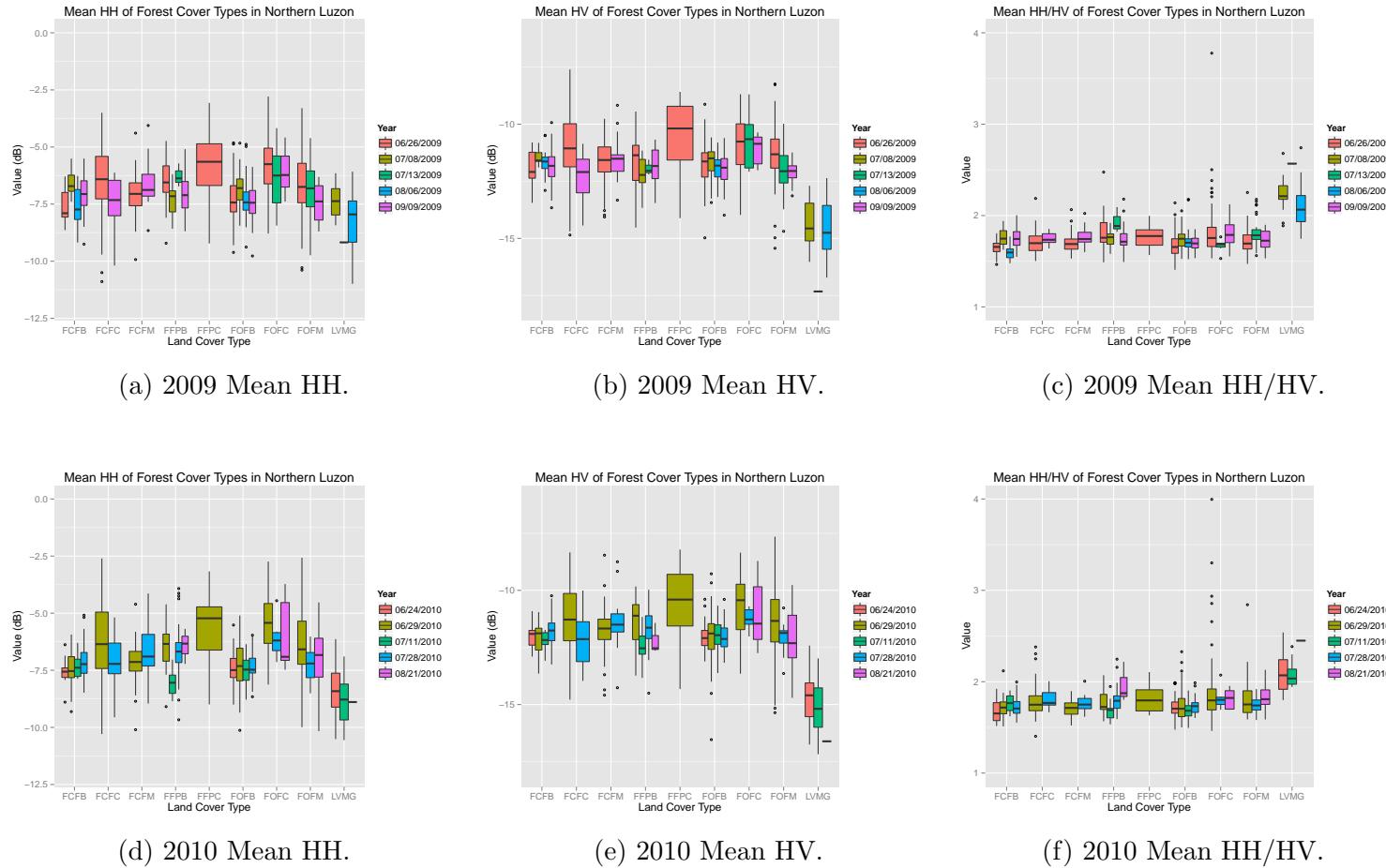


Fig. 4.2: Backscatter values of forest types from PALSAR data acquired in a single year at different seasons. Shown are box-whisker plots of mean backscatter values in 2009 (a-c) and 2010 (d-f).

4.2 Consistency of forest classification hierarchy

A total of 16 dendograms were produced from the hierarchical cluster analysis consisting of four dendograms for each case (covering four years in each case). Each case represents a combination of object-level feature attributes (Table 3.3), particularly: Case 1 using Polarisation feature attributes; Case 2 using Polarisation and Topographic; Case 3 using Polarisation and Texture; and Case 4 using all feature attributes (Polarisation, Topographic, Texture).

As an example, the figure in Fig. 4.7, which was taken from Fig. 4.8, shows the dendrogram resulting from the hierarchical clustering of forest cover types using polarimetric feature attributes from the 2007 PALSAR mosaic image. The dendrogram shows which forest types are similar or dissimilar from other types. The y-axis indicates the computed distance between clusters using an average linkage algorithm in determining which clusters are similar. Within the tree structure, the edge numbers denoted in grey colour indicate the sequence of the grouping of nodes into similar clusters given that an agglomerative hierarchical clustering technique was used, and the red value indicates the AU p-value, of which values ≥ 0.95 represents consistent grouping. The red box indicates which clusters or groups of clusters have ≥ 0.95 AU p-values, and hence have consistent groupings.

In this dendrogram, Cluster #1 showed that closed canopy mixed forest (FCFM) and broadleaved forest plantation (FFPB) were similar with a 97 AU p-value. Next, Cluster #2 showed closed canopy coniferous forest (FCFC) and coniferous forest plantation (FFPC) were similar with 91 AU p-value. Cluster #3 showed closed canopy broadleaved forest (FCFB) and open canopy broadleaved forest (FOFB) were similar

with 100 AU p-value, suggesting that closed and open canopy broadleaved forests were similar in this case using polarimetric attributes. Cluster #4 showed Cluster #2 being grouped together with open canopy mixed forest (FOFM) at 89 AU p-value, and subsequently grouped with Cluster #1 as Cluster #5 at 97 AU p-value. All the forest types in Cluster #5 were subsequently grouped together with Cluster #3 at only 37 AU p-value. Then, all the forest types in Cluster #6 were grouped with open canopy coniferous forest (FOFC) as Cluster #7 with 97 AU p-value. Finally, Cluster #8 showed mangrove forest (LVMG) grouped with all other forest types in Cluster #7, suggesting that mangroves had greatest dissimilarity from the rest of the forest types.

Table 4.1: Baker's Gamma Indices (β coefficients) per case for each pair of years.

Years	Baker's Gamma (β Coefficient)			
	Case 1	Case 2	Case 3	Case 4
2007-2008	0.8966454	0.9896552	0.8786650	0.8797485
2007-2009	0.8527756	0.5834679	0.8078116	0.8393977
2007-2010	0.8849630	0.9634359	0.8301529	0.8250326
2008-2009	0.8853727	0.6208636	0.8750541	0.8914641
2008-2010	0.9899121	0.9706925	0.8943540	0.8950236
2009-2010	0.8858105	0.6269121	0.9539378	0.8526550

For Case 1, the dendograms showed the clustering of forest types using only polarimetric feature attributes (Fig. 4.8). Except for 2009, AU p-values for the other years were all greater than 0.95 indicating reliable clustering results, specifically for 8 of 9 forest types. Mangrove forest (LVMG) and open coniferous forest (FOFC) appeared as distinct clusters across all years. The β coefficients were all high (>0.85),

which indicate relatively similar clusters or dendrogram structure across all four years (Table 4.1).

For Case 2, the dendrograms showed the clustering of forest types using polarimetric and topographic feature attributes (Fig. 4.9). Both 2007 and 2008 exhibited AU p-values above 0.95. For 2009, only two clusters exhibited above 0.95 AU p-values, while 2010 did not at all. Mangrove forest (LVMG) also appeared as distinct clusters across all years. The β coefficients were high (>0.96) for pairwise combinations that do not include 2009, of which values were average (range from 0.58 to 0.63) (Table 4.1). Interestingly, except 2009, broadleaved forest types (i.e., closed, open, and plantation) and both coniferous and mixed forest types (i.e., closed, open, and plantation) branched into separate clusters, suggesting the distinction of these general forest types.

For Case 3, the dendrograms showed the clustering of forest types using polarimetric and texture feature attributes (Fig. 4.10). All years exhibited stable clustering results for 8 of 9 forest types, except 2009 with only 7 of 9 forest types. Mangrove forest (LVMG) and coniferous forest plantations (FFPC) appeared as distinct clusters across all years. The β coefficients were all high (>0.80), which indicate relatively similar clusters or dendrogram structures across all four years (Table 4.1).

For Case 4, the dendrograms showed the clustering of forest types using all feature attributes, including polarimetric, topographic, and texture (Fig. 4.11). The AU p-values consisting of majority of forest types were above 0.95 for 2007, 2008, and 2009. For 2010, only one terminal cluster exhibited above 0.95 AU p-value. Similar to Case 3, mangrove forest (LVMG) and coniferous forest plantations (FFPC) appeared as distinct clusters across all years. The β coefficients were all high (>0.82), which indi-

cate relatively similar clusters or dendrogram structure across all four years (Table 4.1).

Table 4.2: Summary of Baker’s Gamma Indices (β coefficients) for pairwise comparison of dendograms across all cases.

Case	07-08	07-09	07-10	08-09	08-10	09-10	Total	Rank
1	0.89664	0.85277	0.88496	0.88537	0.98991	0.88581	5.39547	1
2	0.98965	0.58346	0.96343	0.62086	0.97069	0.62691	4.75502	4
3	0.87866	0.80781	0.83015	0.87505	0.89435	0.95393	5.23997	2
4	0.87974	0.83939	0.82503	0.89146	0.89502	0.85265	5.18332	3
Total	3.64471	3.08345	3.50358	3.27275	3.74998	3.31931	3.42896	

The Baker’s Gamma Indices for pairwise comparison of dendograms were tabulated across all cases to assess the effect of topographic and texture feature attributes, in addition to polarisation, on the consistency of the hierarchies between dendograms (Table 4.2). Pairwise dendrogram comparisons involving the 2009 ALOS/PALSAR image showed lower consistencies as shown by vertical sum of coefficients for each pairwise comparison.

High Baker’s Gamma Index coefficients (i.e., values close to 1) resulting from pairwise comparisons of dendograms between acquisition years indicated that the hierarchical clustering of forest types were similar and consistent across years, and reinforced the temporal consistency of the ALOS/PALSAR mosaic data acquired in different years. Ranking showed that polarimetric variables (Case 1) had the highest consistencies between pairwise dendrogram comparisons. This suggests that polarimetric data by itself is temporally consistent across different acquisition years, and the forest classification hierarchy is also similarly consistent across different acquisition years. The inclusion of other feature attributes slightly lowers the consistency in

both aspects. Nevertheless, nominal values of Bakers Gamma Index coefficients were still high, suggesting that the inclusion of topographic or texture variables or both only marginally decreases the temporal consistency of ALOS/PALSAR mosaics and hierarchical consistency of the forest type classification.

4.3 Accuracy of multi-level hierarchical classification approach

Classification trees were generated for all cases or combinations of feature attributes for all years (Appendix C.1). For each case, 11 trees were generated from running the decision tree classification in each year, totalling 44 classification trees for one case across four years. Each tree corresponds to one of 11 classification levels based on the multi- level classification hierarchy, of which is defined by a chosen cluster dendrogram (discussed later).

An example of the generated classification trees is shown in Fig. 4.12 (from Appendix C.1), which shows four classification trees corresponding to the first four classification levels at Case 1 of the 2007 PALSAR mosaic image. Classification Level 1 splits into either Land or Water classes. Then at Level 2, with the Water class set aside, the Land class was split into either Vegetation or Non-Vegetation classes. At Level 3, the Vegetation class was split into either Forest or Non-Forest classes while the Non- Vegetation class was set aside. For Level 4, with the Non-Forest class set aside, the Forest class was split into either Mangrove forest or the other forest types group (here denoted as Forest Group 7). To interpret the classification tree, the data is split

starting at top level of the tree, in this case looking at Level 1 classification tree as an example. The tree is split into two branches using Mean HH as the predictor variable, of which values less than -10.6321 in HH backscatter are assigned into the left branch, and those greater than this value are assigned into the right branch. Considering the left branch first, it is then subsequently split again into two branches again using Mean HV as the predictor variable. In this case, however, regardless whether the values are less than or greater than -20.2073 in HV backscatter, these are assigned into the Water class, which are the terminal nodes. Next, considering the right branch from the topmost node, it is split using Mean HV as the predictor variable, of which values less than -13.8405 in HV backscatter are assigned into the left branch, and those greater than this value are assigned into the right branch. Again, treating the left branch first, it is split using SD of HV layer with values less than 1.78303 assigned into the Land class and those greater than this value are assigned into the Water class, which are also both terminal nodes. Lastly, treating the right branch, it is split using Mean HV as the predictor variable. Again in this case, regardless of whether the values are less than or greater than -12.7619 in HV backscatter, these are assigned into the Land class. Once the data in Level 1 are all assigned into either the Land or Water classes, the Water class are set aside, and the data assigned to the Land class are then classified using the Level 2 classification tree, splitting into either Vegetation or Non-Vegetation classes. The classification process continues until all the data are assigned to classes ending at the Level 11 classification tree.

The error rates obtained at each classification level per case per year were tabulated (Appendix C.2). Two sets of plots were generated: the first set of plots presented per

case compared the error rates obtained for each year across classification levels (Fig. 4.13); the second set of plots presented per year compared the error rates for each case across classification levels (Fig. 4.14).

In Fig. 4.13, the misclassification error rates across all years for all cases followed a consistent trend from classification levels 1 to 5, and turned erratic with no clear trend from classification level 6 onwards. Error rates at the first two classification levels were consistently very low across all years, indicating the suitability of PALSAR mosaics for discriminating between Land and Water (Level 1) and Vegetation and Non-Vegetation (Level 2), and this consistency was observed for multi-year PALSAR mosaics.

At Level 1, error rates decreased from Case 1 to Case 4, of which Case 4 had the lowest error rates, suggesting that the inclusion of topographic, texture, or both in addition to polarimetric variables improved classification accuracy. At Level 2, Case 2 and Case 4 showed lower error rates compared to the other two cases, suggesting that inclusion of topographic variables (Case 2) or all variables (Case 4) resulted in better separation of Vegetation and Non-Vegetation classes. Between Case 2 and Case 4, however, no significant differences in error rates were observed, suggesting that the addition of topographic variables (Case 2) or topographic and texture (Case 4) to polarimetric variables can produce the same result. At Level 3, Case 2 and Case 4 also showed lower error rates compared to the other two cases, suggesting the inclusion again of topographic variables or all variables can provide better discrimination of Forest and Non-Forest classes. At Level 4, mangrove forest (LVMG) was consistently separated from the other forest types across all cases and across all years. Notably,

error rates at Level 4 resulted into zero for cases that involved topographic attributes, particularly at Case 2 and Case 4, which was primarily due to elevation and nothing else as the most parsimonious variable that discriminated mangroves from other forest types (Appendix C.2). This indicates that the decision tree classifier selected elevation as a determining variable for classifying mangrove forests, which is true since mangroves are situated on low elevations. However, elevation as the sole variable determining the classification of mangroves is not feasible as not all areas on low elevation are mangroves. Case 1 and Case 3 both showed low error rates in discriminating between mangrove forest and other forest types. Between the two cases, Case 3 showed slightly lower error rates compared to Case 1 suggesting that adding texture variables to polarimetric variables can marginally improve the classification accuracy. At Level 5, open canopy coniferous forest (FOFC) was consistently discriminated from other forest types in Case 1 across all years. In Case 2, closed canopy broadleaved forest (FCFB) was consistently discriminated from other forest types across all years. For Case 3 and Case 4, coniferous forest plantation (FFPC) was consistently discriminated from other forest types across all years. This particular discrimination of one forest type from the others may suggest the influence of specific predictor variables (or combinations thereof) for discriminating these forest types at this classification level. Finally, for Level 6 onwards, no apparent trend was observed except the erratic behaviour of the graphs of error rates, suggesting the varied influence of predictor variables at specific cases in the discrimination of forest types.

In Fig. 4.14, the misclassification error rates were presented this time per year to compare each case across classification levels. Similar to Fig. 4.13, error rates across

all cases for all years followed a consistent trend from classification levels 1 to 5, and turned erratic with no clear trend from classification level 6 onwards. The error rates at the first two classification levels, specifically Level 1 (Land and Water) and Level 2 (Vegetation and Non-Vegetation) were consistently very low across cases for all years. The error rates spiked at Level 3 (Forest and Non-Forest) for all cases, of which Case 2 and Case 4 exhibited the lower error rates compared to the other cases (Appendix C.2), and subsequently Case 4 had the lowest error rates compared to Case 2. At Level 4, error rates were low for Case 1 and Case 3, and resulted to 0 for Case 2 and 4. At Level 5, Case 1 and 2 showed higher error rates compared to the other two cases across all years.

Table 4.3: Summary of total classification error rates per case per year.

Case	2007	2008	2009	2010
1	163.745	158.784	155.709	160.277
2	101.349	103.588	91.795	102.603
3	95.333	101.463	98.812	96.342
4	74.556	78.679	69.118	70.525

A subsequent summary of the total classification error rates per case per year was computed (Table 4.3). The table showed that the inclusion of topographic and texture attributes (Case 2 and Case 3, respectively) tend to reduce the misclassification error rates compared to using solely polarimetric attributes (Case 1). The addition of topographic to polarimetric attributes (Case 2) also showed higher error rates compared to a polarimetric and texture combination (Case 3). The inclusion of all feature attributes (Case 4), including polarimetric, topographic, and texture, produced the lowest misclassification error rates across years; and consequently, it formed the

bases of the classification hierarchies that were constructed through rulesets to produce the annual land cover maps (Fig. 4.15).

The mean error rates of land cover classes for each case and each year were tabulated (Table 4.4), which indicates that the mean error rates were still the lowest for Case 4 compared to other cases across all years. The inclusion of either topographic or texture attributes to polarimetric data tended to lower the classification error rates, of which error rates with texture included was lower compared to error rates with topographic attributes included. This emphasises that the inclusion of both topographic and texture feature attributes improve the classification accuracy of forest cover types.

Table 4.4: Summary of total classification error rates per case per year.

Case	2007	2008	2009	2010
1	14.886 ± 10.04	14.435 ± 10.11	14.155 ± 12.68	14.571 ± 10.870
2	9.214 ± 6.137	9.417 ± 6.777	8.345 ± 5.221	9.328 ± 6.486
3	8.667 ± 8.138	9.224 ± 7.618	8.983 ± 9.273	8.758 ± 8.743
4	6.778 ± 6.458	7.153 ± 6.155	6.283 ± 5.793	6.411 ± 6.416

4.4 Accuracy of heuristic classification approach

Additional assessments of classification accuracies were done to discriminate: (a) closed and open canopy forests; and (b) broadleaved, coniferous, and mixed forests through a heuristic approach without following the hierarchical clustering result. It should be noted here that the error rates obtained for the first four classification levels

for both groups were the same values obtained in the previous sub-section.

4.4.1 Closed and open canopy forests

The final classification level at Level 5 dealt with separating closed canopy and open canopy forests, of which each group consisted of the following specific forest types: for closed canopy forests, the forest types included closed broadleaved (FCFB), closed coniferous (FCFC), and closed mixed (FCFM); for open canopy forests, the forest types included open broadleaved (FOFB), open coniferous (FOFC), open mixed (FOFM), broadleaved forest plantation (FFPB), and coniferous forest plantation (FFPC).

Table 4.5: Classification error rates for closed and open canopy forests per case per year using a heuristic approach.

Case	2007	2008	2009	2010
1	36.35	37.22	37.22	35.24
2	30.89	32.13	29.90	29.40
3	36.35	37.22	37.22	35.24
4	30.89	32.13	32.13	29.40

Results showed that error rates at Level 5 were exactly the same in Case 1 and Case 3, and almost the same for Case 2 and Case 4 with the exception of values in 2009 (Table 4.5). Error rates in Case 1 and Case 3 were similar, suggesting that texture variables do not contribute to improving classification accuracy in discriminating between closed and open canopy forests. Case 2 and Case 4 showed lower error rates compared to the other two cases, suggesting that topographic variables contributed to improving classification accuracy in discriminating between closed and open canopy

forests by 5-7%. Overall, classification error rates at Level 5 were high at approximately 29-38%.

4.4.2 Broadleaved, coniferous, and mixed forests

The final classification level at Level 5 dealt with separating broadleaved, coniferous, and mixed forests, of which each group consisted of the following specific forest types: broadleaved forests included closed broadleaved (FCFB), open broadleaved (FOFB), and broadleaved forest plantation (FFPB); coniferous forests included closed coniferous (FCFC), open coniferous (FOFC), and coniferous forest plantation (FFPC); and mixed forests included closed mixed (FCFM) and open mixed (FOFM).

Table 4.6: Classification error rates for closed and open canopy forests per case per year using a heuristic approach.

Case	2007	2008	2009	2010
1	43.42	44.04	41.44	41.19
2	32.26	33.37	33.25	31.89
3	43.42	44.04	41.44	41.19
4	32.26	33.37	33.25	33.75

Results showed that classification error rates at Level 5 were exactly the same in Case 1 and Case 3, and almost the same for Case 2 and Case 4 with the exception of values in 2010 (Table 4.6). Error rates in Case 1 and Case 3 were similar, suggesting that texture variables do not contribute to improving classification accuracy in discriminating between broadleaved, coniferous, and mixed forests. Case 2 and Case 4 showed lower error rates compared to the other two cases, suggesting that topographic

variables contributed to improving classification accuracy in discriminating between broadleaved, coniferous, and mixed forests by 8-11%. Overall, classification error rates at Level 5 were high at approximately 31-44%, of which were also higher compared to error rates obtained in classifying between closed and open canopy forests (Table 4.5).

4.4.3 Area statistics of land cover types

The area statistics of each land cover type was computed from the resulting land cover maps from the decision tree classification of multi-year PALSAR mosaics, specifically based on Case 4 (utilising all predictor variables, namely polarimetric, topographic, and texture) from the hierarchical clustering approach (Table 4.7).

Significant differences can be seen from the area figures across years for all land cover types. This may be attributed to the high error rates, specifically in the classification of different forest types. Differences in area statistics can also be observed between figures obtained from the PALSAR mosaics and the NAMRIA 2010 land cover map. (Note that differences in total area figures at the bottom row may be attributed to masked areas in the PALSAR mosaics due to shadow and layover.)

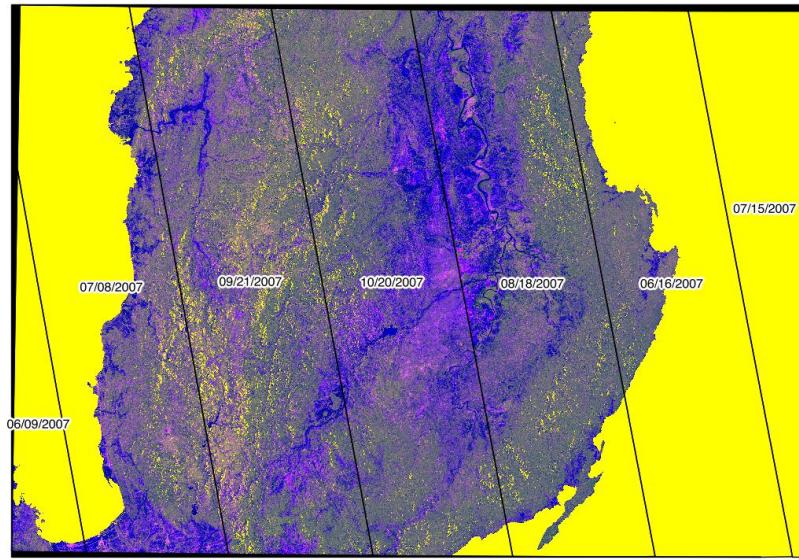
Table 4.7: Area statistics of land cover types from classified annual PALSAR mosaics.

Class	PALSAR				NAMRIA
	2007 (ha)	2008 (ha)	2009 (ha)	2010 (ha)	2010 (ha)
Water	634,067	461,769	345,434	349,074	
Non-vegetation	376,640	421,192	302,471	114,440	
Non-forest	805,576	1,670,234	1,526,462	1,851,864	3,085,588
Mangrove forest	183,238	90,844	151,386	88,734	2,122
CFB	4,045	42,453	127,200	104,958	685,677
CFC	9,556		106,556	144,578	15,905

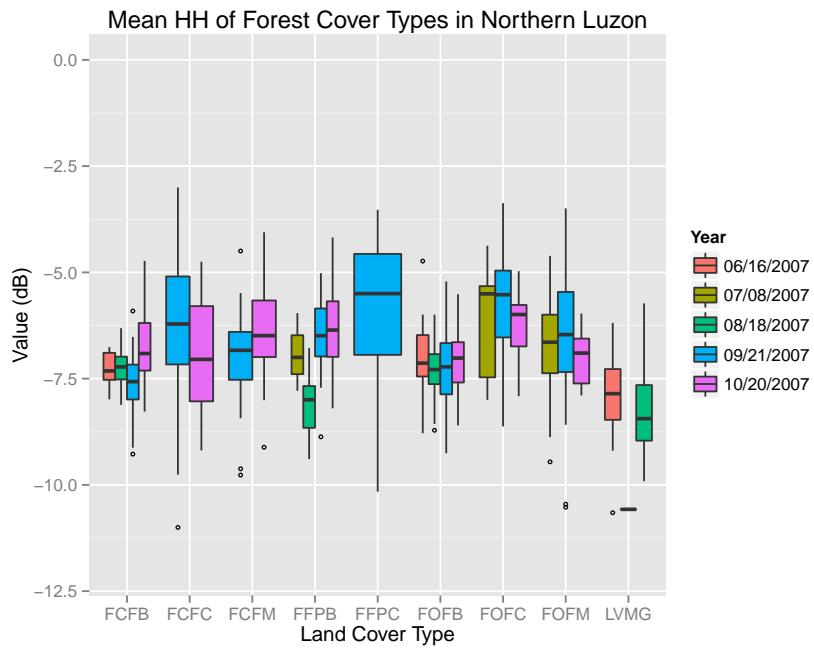
Table 4.7: Area statistics of land cover types from classified annual PALSAR mosaics.

Class	PALSAR				NAMRIA
	2007 (ha)	2008 (ha)	2009 (ha)	2010 (ha)	2010 (ha)
CFM	124,515	258,253	101,324	209,390	26,072
OFB	1,608,354	616,830	1,531,202	1,065,904	841,185
OFC		190,995	6,670	55,054	163,806
OFM	1,444	277,572	56,492	106,434	65,604
FPB	528,260	246,608	3,627	185,417	22,067
FPC	1,519	215	18,183	1,068	802
Total	4,276,915	4,276,915	4,276,915	4,276,915	4,908,828

Note: CFB: Closed forest, broadleaved; CFC: Closed forest, coniferous; CFM: Closed forest, mixed; OFB: Open forest, broadleaved; OFC: Open forest, coniferous; OFM: Open forest, mixed; FPB: Forest plantation, broadleaved; FPC: Forest plantation, coniferous.

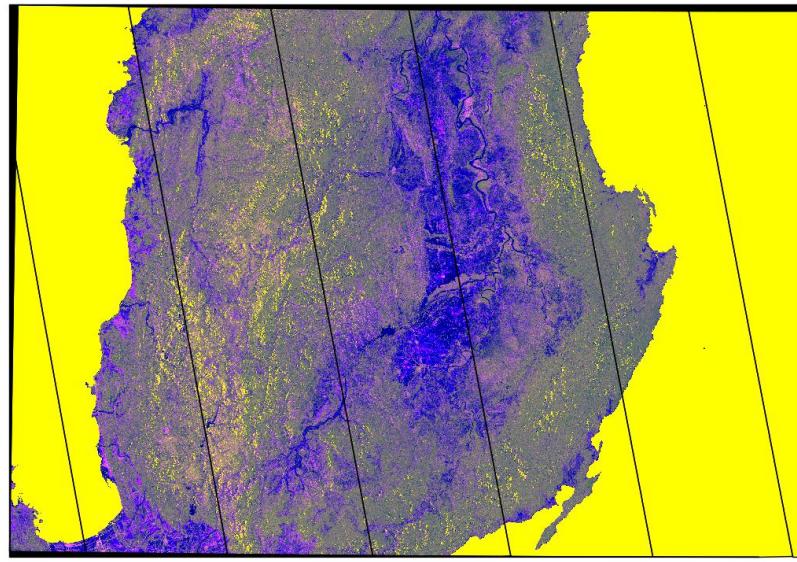


(a)

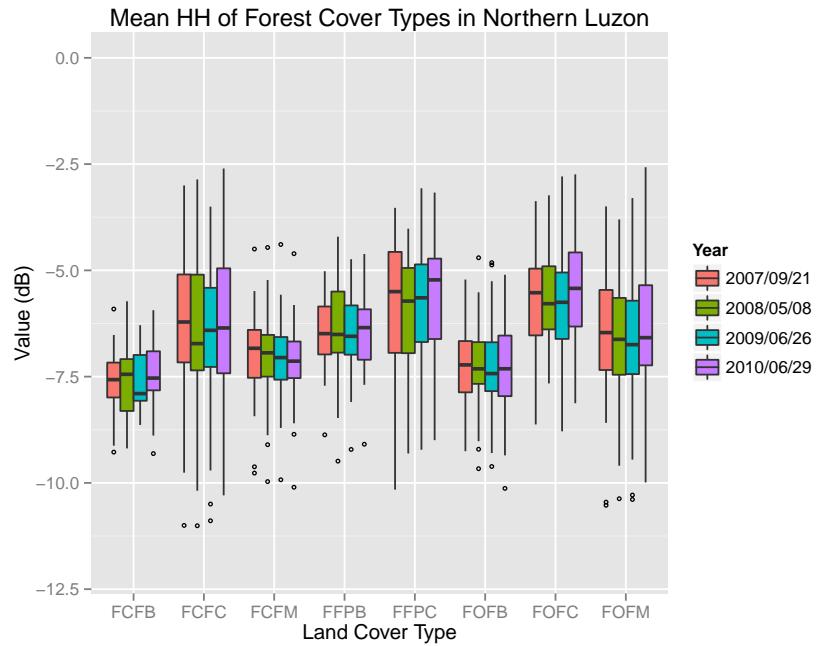


(b)

Fig. 4.3: The 2007 PALSAR mosaic image and the dates of acquisition of each strip/path that comprise the whole mosaic (a), and the distribution of mean HH backscatter for each forest cover type (b) based on the 2007 mosaic image.



(a)



(b)

Fig. 4.4: A PALSAR mosaic image highlighting the strip/path where ROIs of forest cover types were taken across four years (a), and the distribution of mean HH backscatter for each forest cover type (b) along the same strip at different years.

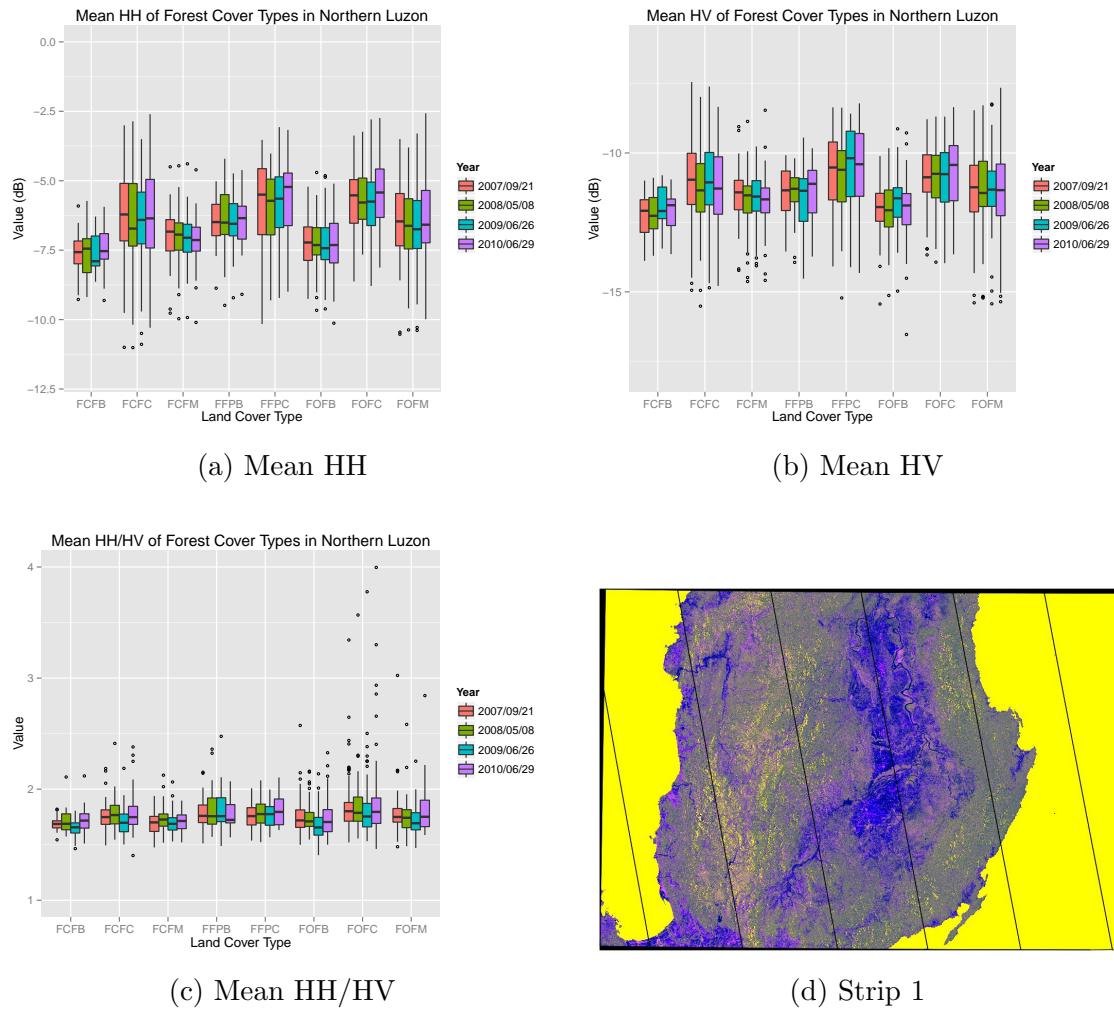


Fig. 4.5: Backscatter values of forest types from PALSAR data acquired along the same strip (Cordillera) at different years.

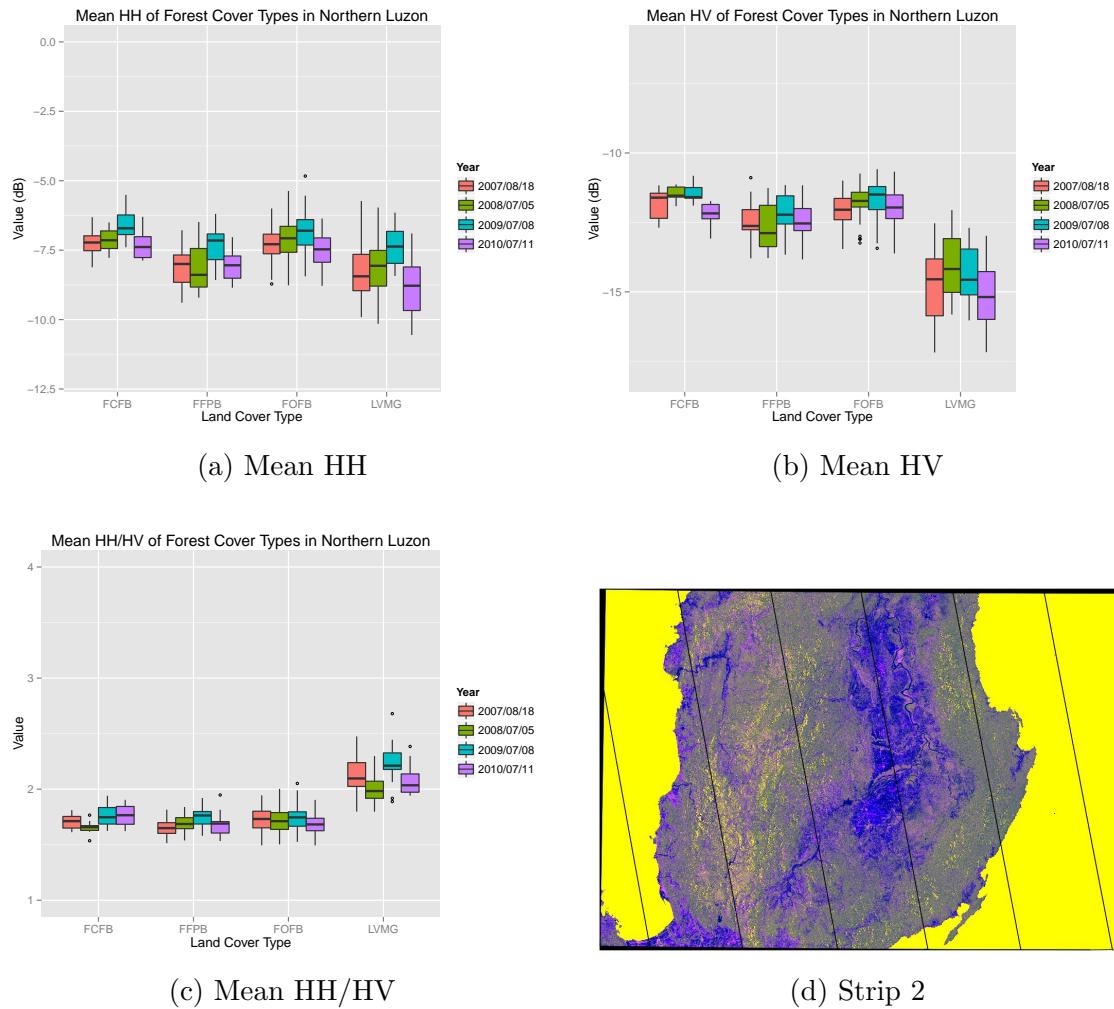


Fig. 4.6: Backscatter values of forest types from PALSAR data acquired along the same strip (Sierra Madre) at different years.

Dendrogram: 2007 Case 1

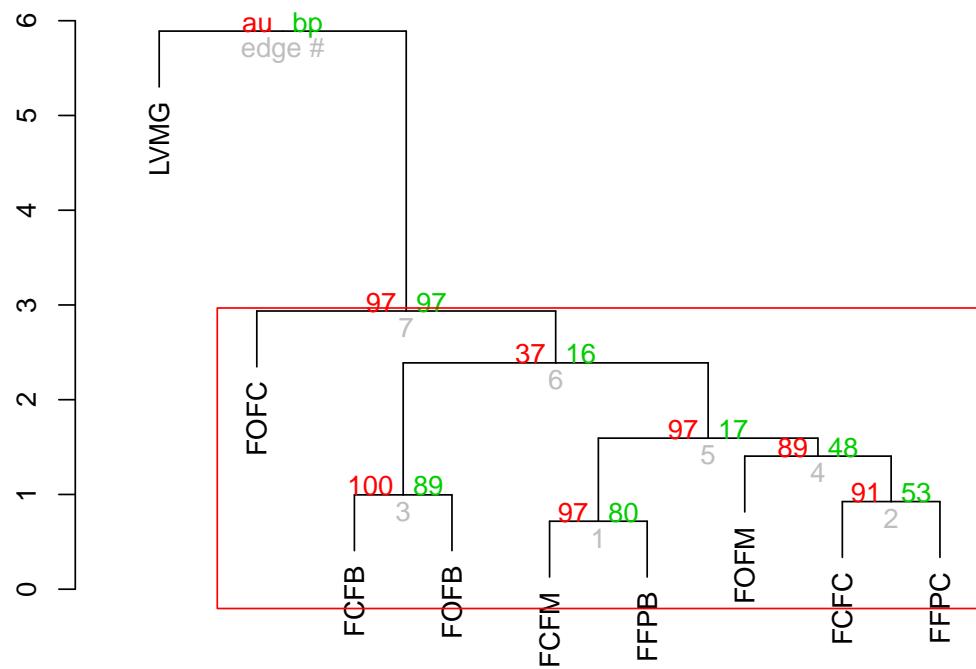


Fig. 4.7: Example dendrogram result from the hierarchical clustering of forest cover types using polarimetric feature attributes (Case 1) from the 2007 PALSAR mosaic image.

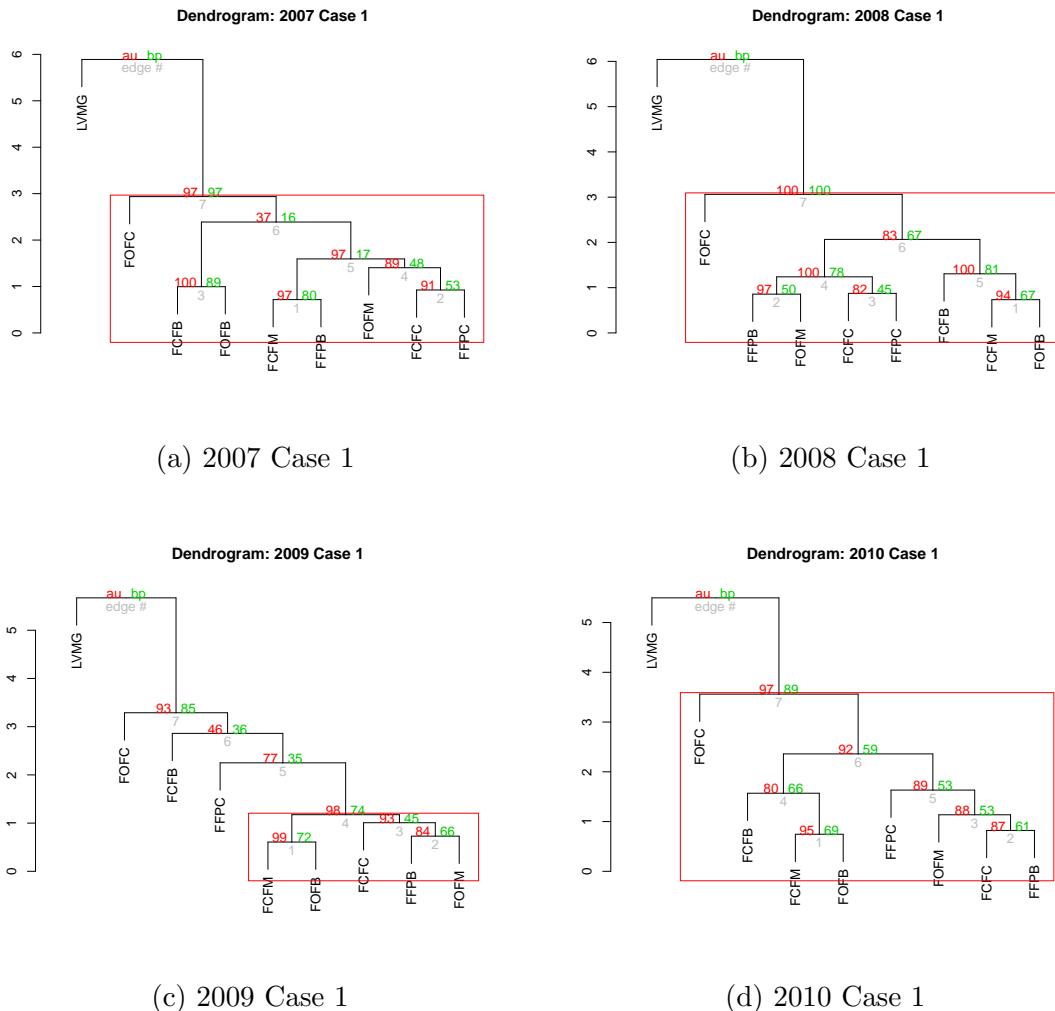


Fig. 4.8: Dendograms from cluster analysis using Case 1. Red boxes AU p-value ≥ 0.95 .

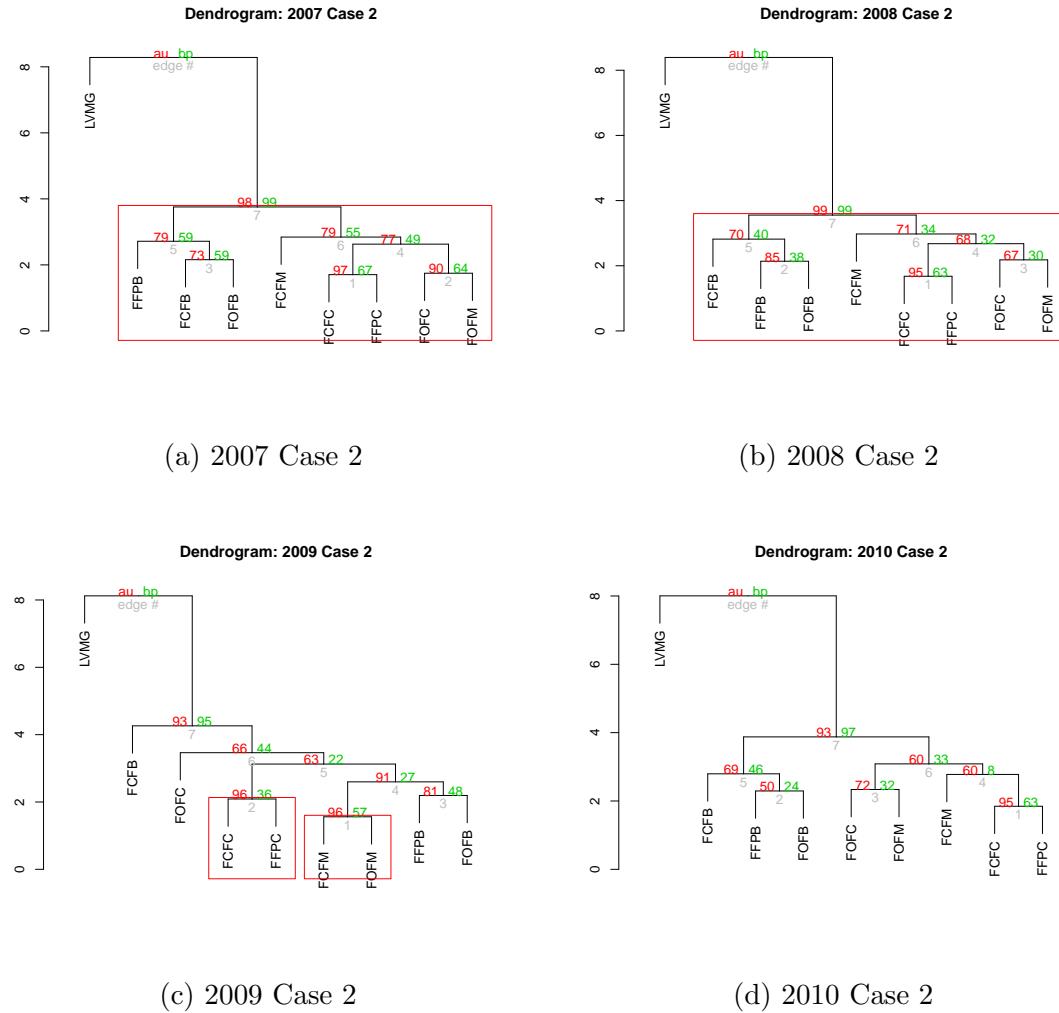


Fig. 4.9: Dendograms from cluster analysis using Case 2. Red boxes AU p-value ≥ 0.95 .

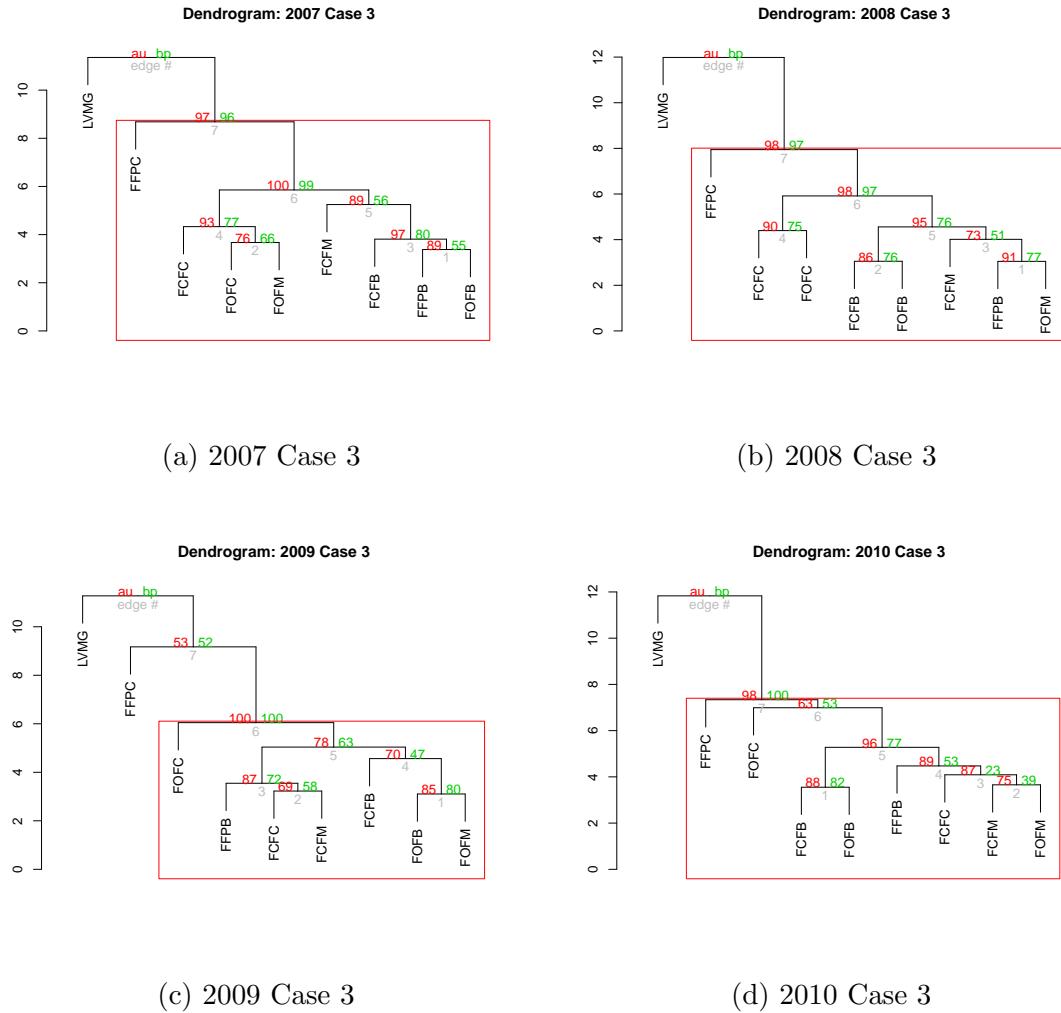


Fig. 4.10: Dendograms from cluster analysis using Case 3. Red boxes AU p-value ≥ 0.95 .

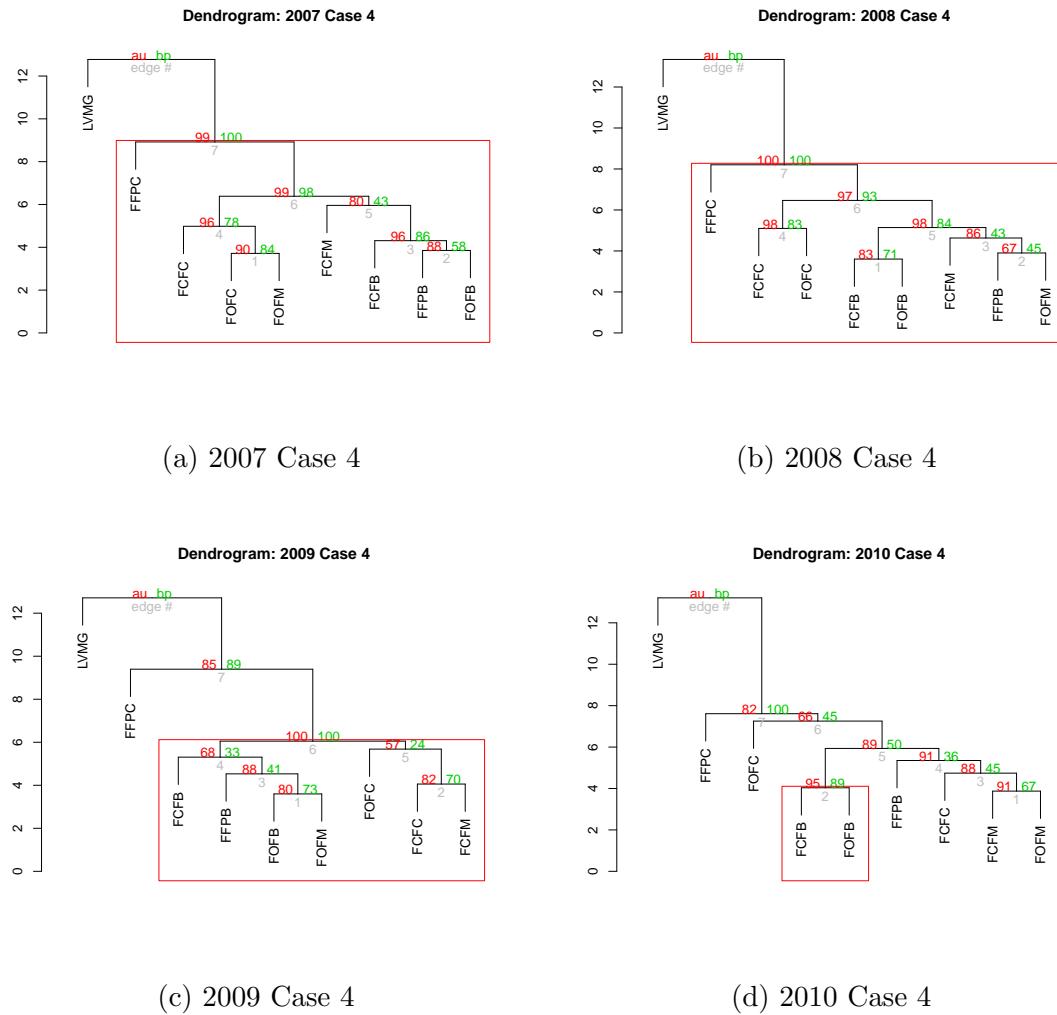


Fig. 4.11: Dendrograms from cluster analysis using Case 4. Red boxes AU p-value ≥ 0.95 .

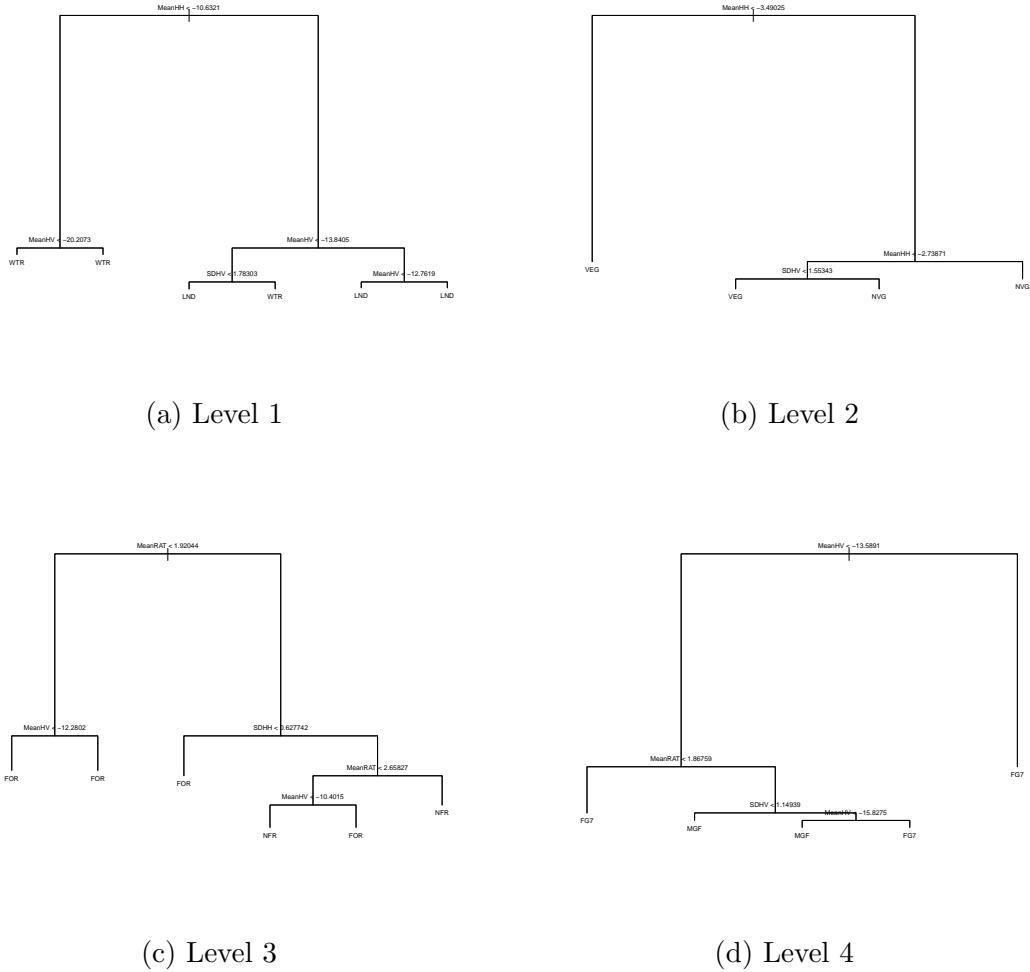


Fig. 4.12: Example classification trees corresponding to the first four classification levels at Case 1 of the 2007 PALSAR mosaic image. Shown are Level 1 (land-water), Level 2 (vegetation - non-vegetation), Level 3 (forest - non-forest), and Level 4 (mangrove - non-mangrove).

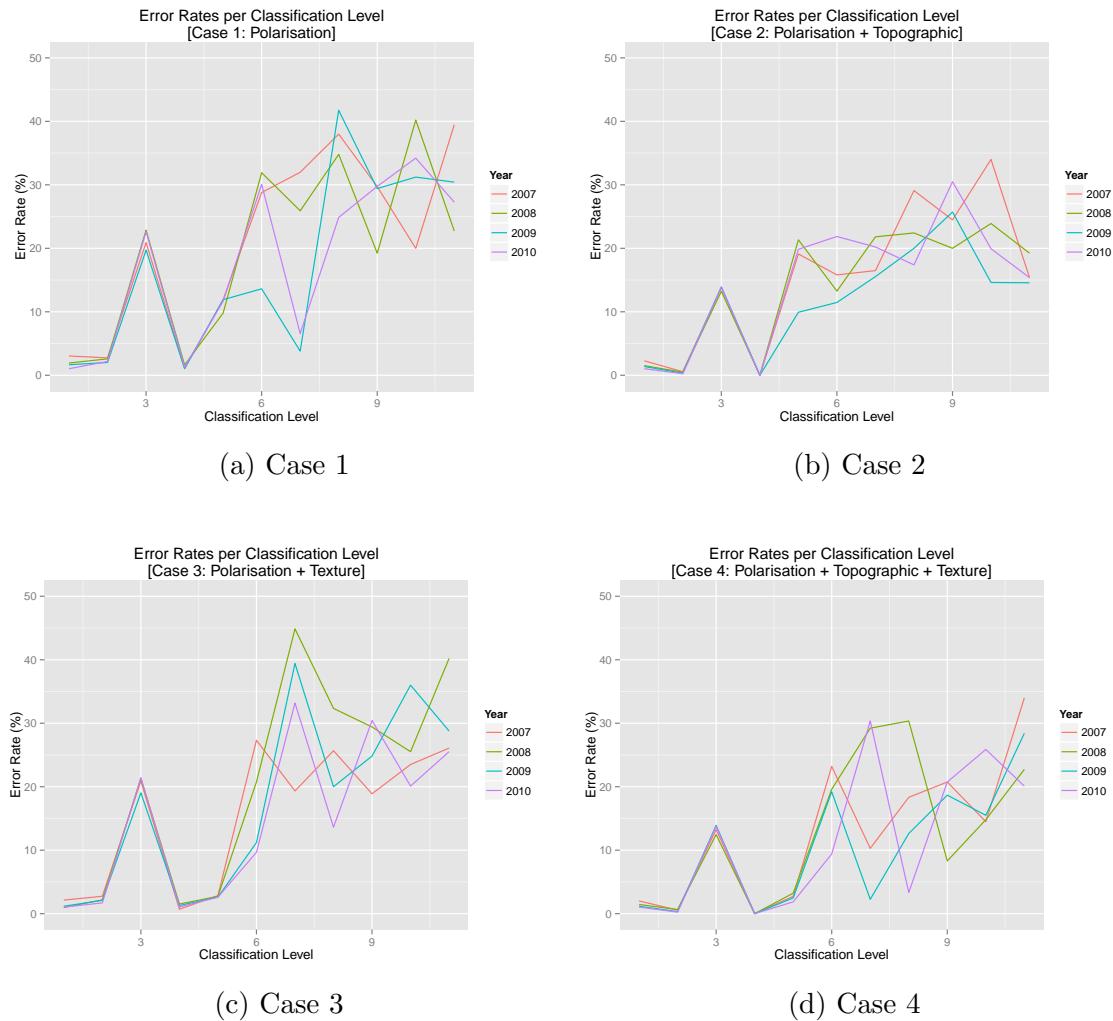


Fig. 4.13: Plots comparing error rates obtained for each year across classification levels per case.

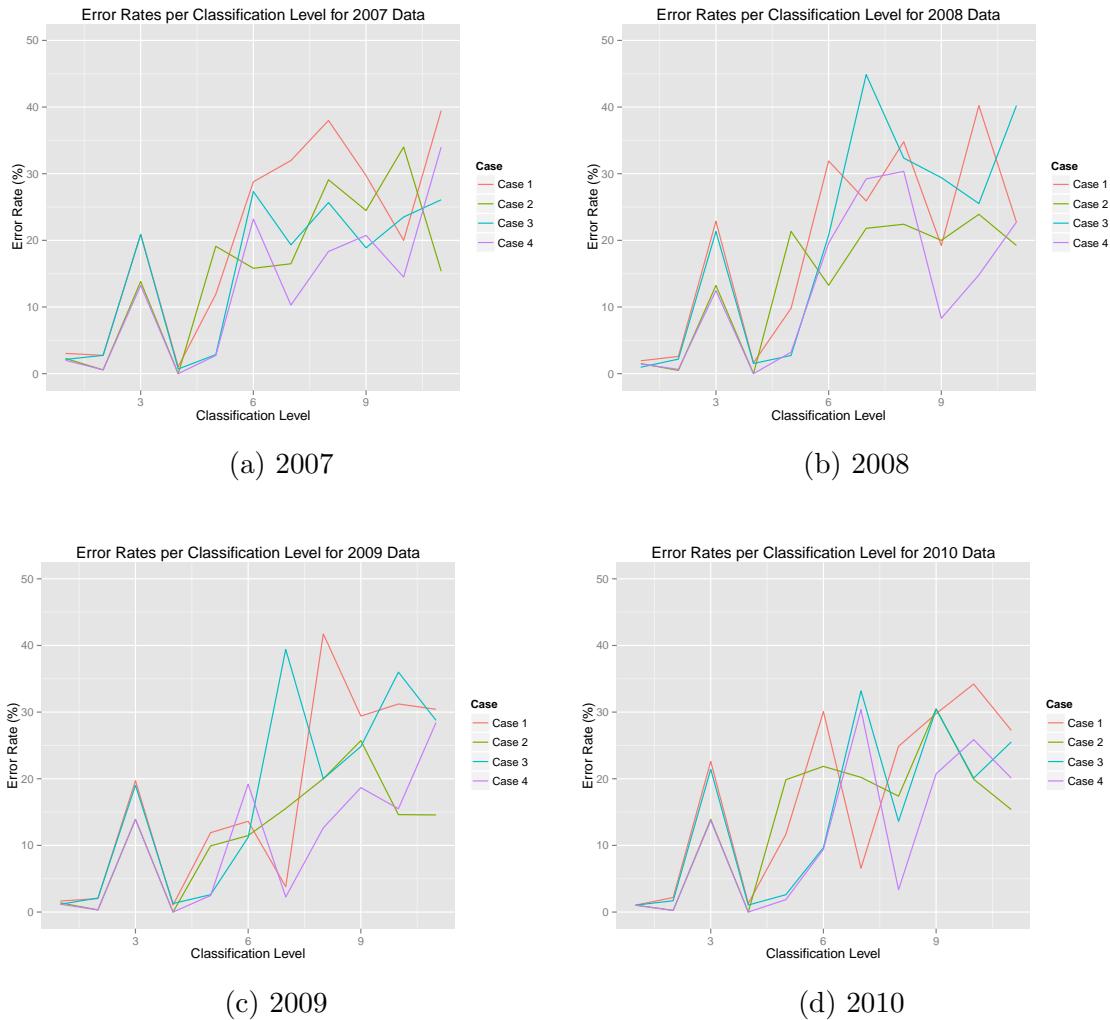


Fig. 4.14: Plots comparing error rates obtained for each case across classification levels per year.

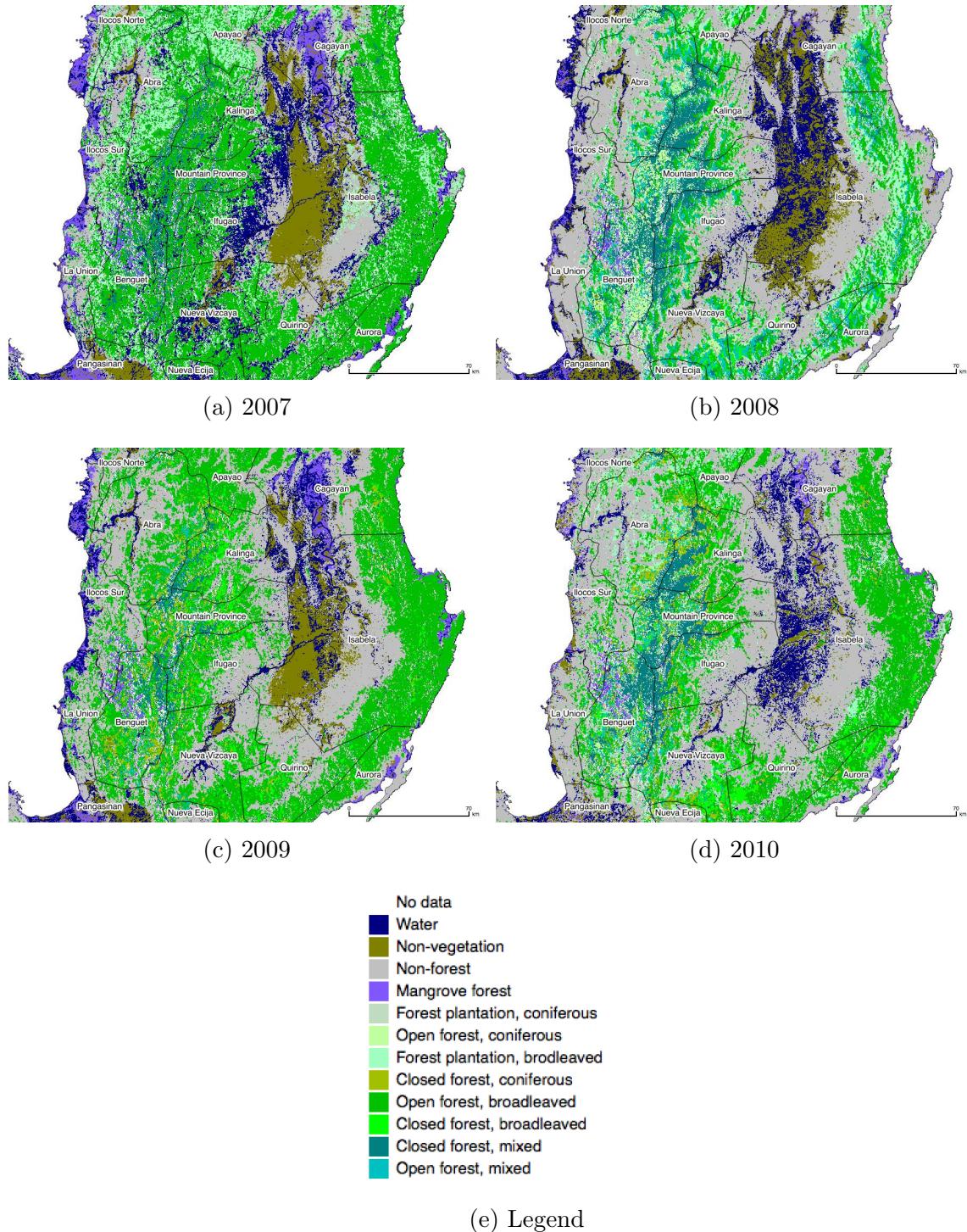


Fig. 4.15: Land cover and forest type maps generated from annual ALOS/PALSAR mosaics using decision tree classification.

Chapter 5

Discussion

Sustainable forest management needs accurate information on the extent of forests, of which remote sensing is an effective tool for mapping and monitoring of forest cover. Land and forest cover maps of the Philippines in the past were mainly generated using optical data, but spaceborne radar sensors have been promising as a viable and accurate data source. This study evaluated the suitability of ALOS/PALSAR mosaic data for mapping and monitoring of forest cover types in the Philippines by assessing the following indicators, particularly: the temporal consistency of multi-year PALSAR mosaic data; the presence and consistency of a multi-level classification hierarchy based on a combination of polarimetric, topographic, and texture variables; and accuracy of forest type classification using decision trees.

This study demonstrated the temporal consistency of globally available PALSAR mosaic data. This makes ALOS/PALSAR data suitable for regular and periodic forest monitoring purposes, appropriate for developing operational national forest monitoring systems, and engaging in performance-based incentive schemes such as

Reducing Emissions from Deforestation and Forest Degradation (REDD+), wherein data consistency is a key requirement. This consistency was demonstrated by analysing the distribution of radar backscatter vis-a-vis data acquisition dates of PALSAR mosaic data acquired in a single year but the data strips used to put the mosaic together were taken at different seasons, and of PALSAR mosaic data acquired along the same data strip or path but at different years.

The existence of a multi-level classification hierarchy of forest types developed using hierarchical cluster analysis was also demonstrated in the study, which was consistent across annual PALSAR mosaics. This was demonstrated by high Bakers Gamma indices, which indicated that the hierarchical clustering of forest types were similar and consistent across years, and reinforced the temporal consistency of PALSAR mosaic data. Using polarimetric variables showed the highest consistencies between pairwise dendograms. The inclusion of topographic or texture variables—while still resulting into high indices—suggested that the variables marginally decreases the consistency of the multi-level classification hierarchy over multi-year PALSAR mosaics. The most consistent multi-level classification hierarchy of forest types guided subsequent forest type classification using decision trees.

Hierarchical multi-level classification provided a suitable approach for quantifying the degree of accuracy in discriminating different forest cover types. Employing a decision tree classification approach identified the most relevant image layers and feature attributes for classifying different classes. Combining polarimetric, topographic, and texture attributes yielded lowest classification error rates compared to use of polarimetric data alone, or the combination of either polarimetric and topographic or

polarimetric and texture variables in the classification.

However, even with the best combination of feature attributes used in study, high error rates were obtained in the classification of forest types following the multi-level hierarchical classification. In turn, forest cover types were mapped incorrectly in the resulting annual forest cover maps, and area statistics varied considerably among forest types across years.

In the separate assessment following a heuristic classification approach, misclassification error rates were lower in separating closed and open canopy forests compared to separating between broadleaved, coniferous, and mixed forests. This was in contrast to the findings of Dobson et al. (1996), particularly in their investigation of the capability of ERS-1 and JERS-1 SAR imagery using a knowledge-based classification with hierarchical decision rules to differentiate structural land cover categories, of which they concluded that L-band was capable of differentiating forest types based on their structural characteristics, particularly as either broadleaved or coniferous, but not canopy attributes like closed or open canopy cover.

Also, within the heuristic approach, the inclusion of topographic variables in either forest groups affected the classification accuracy by yielding lower error rates as compared to including texture variables, which had no effect on improving classification accuracy and was no different than using polarimetric variables alone. This result was similar to the findings of Li et al. (2012) where texture attributes from L-band SAR data were less important than polarimetric attributes.

Nevertheless, the error rates obtained in the heuristic classification approach that aimed to gauge the separation of either closed or open canopy forests; or broadleaved,

coniferous, and mixed forests were comparably much higher than using the multi-level hierarchical classification approach, which suggests that using a hierarchical clustering analysis to find natural groupings of forest types in the data yielded better classification accuracies compared to a heuristic method. Also, in the case of discriminating mangroves from other forest types, apparent classification errors (false positive) were produced as a result of the classifier determining elevation as the most parsimonious attribute for its discrimination (from closer inspection of the resulting decision trees).

The low accuracies achieved from discriminating different forest cover types may be attributed to the low sensitivity of data layers and feature attributes used in distinguishing the physical characteristics of the forest types being mapped. It may also be attributed to the incompatibility of discriminating forest attributes by the PALSAR sensor with the forest categories based on the FAO Global Forest Resources Assessment classification scheme. L-band radar backscatter was poorly sensitive to canopy attributes (e.g., closed canopy and open canopy) that characterise the FAO Global Forest Resources Assessment forest cover types. While other studies have indicated better performance of L-band radar backscatter at discriminating forest cover types characterised by forest structural attributes such as broadleaved and coniferous, the error rates obtained in this study suggested otherwise and also showed that error rates obtained for forest types described by forest canopy attributes were slightly lower than the error rates obtained for forests types described by structural attributes.

Chapter 6

Conclusions and Recommendations

This study found multi-year ALOS/PALSAR mosaic data to be temporally consistent and capable of classifying between forest and non-forest areas; hence making it suitable tool for mapping and monitoring forest extent. Unlike fully polarimetric SAR data which require more complex image processing, the mosaic data is less complicated and is provided to the public free of charge, which makes it a viable data source for operational, wall-to-wall forest mapping and monitoring. In terms of discriminating forest types, ALOS/PALSAR was not suitable for discriminating between forest types based on canopy characteristics or a forest typology such as the FAO Global Forest Resource Assessment classification scheme.

Cluster analysis provided information on structure and hierarchy of forest types in aid of classification. The analysis can similarly be applied for other classification systems to reveal hierarchy and structure of land cover types, which can guide subsequent classification processes. Decision trees were successfully used in past research for classifying land and forest cover types using SAR data, and even outperformed

other statistical classification techniques. A decision tree classifier was also applied in this study, which provided information on classification accuracies to assess the degree to which different forest types can be discriminated using a combination of feature attributes. A combination of hierarchical cluster analysis and decision tree classification techniques is recommended as a viable approach for determining forest hierarchy by finding structure in the data and using the output dendrogram to guide the subsequent decision tree classification.

However, while the multi-level hierarchical classification approach is recommended, low classification accuracies (or high error rates) were obtained, which may be primarily due to the incompatibility of the sensitivities of the PALSAR sensor to forest structural attributes with the forest classification scheme used, particularly the forest types of the FAO Global Forest Resources Assessment. Hence, a different forest classification system such as the one devised by Fernando et al. (2008), which described Philippine forest formations typified based on their physical attributes, particularly physiognomy, elevation, and soil substrate may be explored as a plausible forest type classification scheme, and a topic of subsequent research for mapping Philippine forest cover types using SAR data.

An object-based approach also allowed the use of multiple sources of data to increase the dimensionality of the input data for classification. Topographic and texture feature attributes helped to decrease classification error rates in contrast to using solely polarimetric attributes for classifying land and forest cover types. In turn, the availability of multiple data sources was maximised by the decision tree classifier through its selection of the most parsimonious feature attributes to improve

the classification outcome. A subject of future research can be to test the sensitivity of specific texture measures to forest type classification, and exploring the effects of varying image segmentation parameters in terms of assessing temporal consistency, multi-level classification hierarchy, and classification accuracy.

Discriminating and mapping forest cover types is a key piece of information to support forest management. Combining multi-frequency SAR data such as C-band and L- band radar sensors is suggested as the subject of future research to improve the discrimination of forest types. C-band data such as Sentinel-1A sensor, which is sensitive to forest canopy components can complement L-band PALSAR data, which is more sensitive to forest structure components. Since SAR overcomes limitations of optical data, combining the data from these two SAR platforms have the potential to provide spatially explicit information in support of an operational, wall-to-wall forest mapping and monitoring system.

Bibliography

- Achard, F, R Beuchle, P Mayaux, HJ Stibig, C Bodart, A Brink, S Carboni, B Desclée, F Donnay, HD Eva, A Lupi, R Raši, R Seliger, and D Simonetti (2014). Determination of tropical deforestation rates and related carbon losses from 1990 to 2010. *Global Change Biology* **20**(8), 2540–2554. ISSN: 1365-2486. DOI: 10.1111/gcb.12605.
- Achard, F, HJ Stibig, HD Eva, EJ Lindquist, A Bouvet, O Arino, and P Mayaux (2010). Estimating tropical deforestation from Earth observation data. *Carbon Management* **1**(2), 271–287. ISSN: 1758-3004. DOI: 10.4155/cmt.10.30.
- Ahlqvist, O (2008). In search of classification that supports the dynamics of science: the FAO Land Cover Classification System and proposed modifications. *Environment and Planning B: Planning and Design* **35**(1), 169–186. ISSN: 0265-8135, 1472-3417. DOI: 10.1068/b3344.
- Almeida-Filho, R, A Rosenqvist, YE Shimabukuro, and R Silva-Gomez (2007). Detecting deforestation with multitemporal L-band SAR imagery: a case study in western Brazilian Amazônia. *International Journal of Remote Sensing* **28**(6), 1383–1390. ISSN: 0143-1161. DOI: 10.1080/01431160600754591.

Almeida-Filho, R, YE Shimabukuro, A Rosenqvist, and GA Sánchez (2009). Using dual-polarized ALOS PALSAR data for detecting new fronts of deforestation in the Brazilian Amazônia. *International Journal of Remote Sensing* **30**(14), 3735–3743. doi: 10.1080/01431160902777175.

Arino, O, D Gross, F Ranera, M Leroy, P Bicheron, C Brockman, P Defourny, C Vancutsem, F Achard, L Durieux, L Bourg, J Latham, AD Gregorio, R Witt, M Herold, J Sambale, S Plummer, and JL Weber (2007). GlobCover: ESA service for global land cover from MERIS. In: *2007 IEEE International Geoscience and Remote Sensing Symposium*, pp.2412–2415. doi: 10.1109/IGARSS.2007.4423328.

Avtar, R, H Sawada, W Takeuchi, and G Singh (2012). Characterization of forests and deforestation in Cambodia using ALOS/PALSAR observation. *Geocarto International* **27**(2), 119–137. ISSN: 1010-6049. doi: 10.1080/10106049.2011.626081.

Baatz, M and A Schäpe (2000). Multi-resolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: *Angewandte Geographische Informationsverarbeitung XII*. Ed. by J Strobl, T Blaschke, and G Griesebner. Heidelberg, Germany, pp.12–23.

Bajorski, P (2011). *Statistics for Imaging, Optics, and Photonics*. 1st ed. Hoboken, N.J: Wiley. ISBN: 978-0-470-50945-6.

Baker, FB (1974). Stability of two hierarchical grouping techniques case I: sensitivity to data errors. *Journal of the American Statistical Association* **69**(346), 440–445. ISSN: 0162-1459. doi: 10.1080/01621459.1974.10482971.

- Bankoff, G (2007). One island too many: reappraising the extent of deforestation in the Philippines prior to 1946. *Journal of Historical Geography* **33**(2), 314–334. ISSN: 0305-7488. doi: DOI:10.1016/j.jhg.2006.06.021.
- Bartholomé, E and AS Belward (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing* **26**(9), 1959–1977. ISSN: 0143-1161. doi: 10.1080/01431160412331291297.
- Benz, UC, P Hofmann, G Willhauck, I Lingenfelder, and M Heynen (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing* **58**(3–4), 239–258. ISSN: 0924-2716. doi: 10.1016/j.isprsjprs.2003.10.002.
- Berk, RA (2008). “Classification and Regression Trees (CART)”. In: *Statistical Learning from a Regression Perspective*. Springer Series in Statistics. DOI: 10.1007/978-0-387-77501-2_3. Springer New York, pp.1–65. ISBN: 978-0-387-77500-5 978-0-387-77501-2.
- Bhatta, B (2013). *Research Methods in Remote Sensing*. SpringerBriefs in Earth Sciences. Dordrecht: Springer Netherlands. ISBN: 978-94-007-6593-1 978-94-007-6594-8.
- Blanco, JA and YH Lo, eds. (2012). *Forest Ecosystems - More than Just Trees*. InTech. ISBN: 978-953-51-0202-1.
- Blaschke, T, GJ Hay, M Kelly, S Lang, P Hofmann, E Addink, R Queiroz Feitosa, F van der Meer, H van der Werff, F van Coillie, and D Tiede (2014). Geographic object-based image analysis – towards a new paradigm. *ISPRS Journal of Photogrammetry*

and Remote Sensing **87**, 180–191. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2013.09.014.

Boyd, DS and FM Danson (2005). Satellite remote sensing of forest resources: three decades of research development. *Progress in Physical Geography* **29**(1), 1–26. DOI: 10.1191/0309133305pp432ra.

Breiman, L, J Friedman, CJ Stone, and RA Olshen (1984). *Classification and Regression Trees*. Taylor & Francis. ISBN: 978-0-412-04841-8.

Burnett, C and T Blaschke (2003). A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling* **168**(3), 233–249. ISSN: 0304-3800. DOI: 10.1016/S0304-3800(03)00139-X.

Campbell, JB and RH Wynne (2011). *Introduction to Remote Sensing*. 5th ed. New York, USA: The Guilford Press. ISBN: 978-1-60918-176-5.

Canty, MJ (2014). *Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for ENVI/IDL and Python*. 3rd ed. New York: CRC Press. ISBN: 978-1-4665-7037-5.

Cartus, O, M Santoro, and J Kellndorfer (2012). Mapping forest aboveground biomass in the Northeastern United States with ALOS PALSAR dual-polarization L-band. *Remote Sensing of Environment* **124**, 466–478. ISSN: 0034-4257. DOI: 10.1016/j.rse.2012.05.029.

Chazdon, RL, PHS Brancalion, L Laestadius, A Bennett-Curry, K Buckingham, C Kumar, J Moll-Rocek, ICG Vieira, and SJ Wilson (2016). When is a forest a forest? Forest concepts and definitions in the era of forest and landscape restoration. *Ambio*. ISSN: 0044-7447, 1654-7209. DOI: 10.1007/s13280-016-0772-y.

- Chokkalingam, U, AP Carandang, JM Pulhin, RD Lasco, RJJ Peras, and T Toma, eds. (2006). *One Century of Forest Rehabilitation in the Philippines: Approaches, Outcomes and Lessons*. Bogor, Indonesia: Center for International Forestry Research. ISBN: 979-24-4643-5.
- Costanza, R, R de Groot, P Sutton, S van der Ploeg, SJ Anderson, I Kubiszewski, S Farber, and RK Turner (2014). Changes in the global value of ecosystem services. *Global Environmental Change* **26**, 152–158. ISSN: 0959-3780. doi: 10.1016/j.gloenvcha.2014.04.002.
- Dahdouh-Guebas, F (2002). The use of remote sensing and GIS in the sustainable management of tropical coastal ecosystems. *Environment, Development and Sustainability* **4**(2). 10.1023/A:1020887204285, 93–112. ISSN: 1387-585X.
- Defourny, P and S Bontemps (2012). “Revisiting Land-Cover Mapping Concepts”. In: *Remote Sensing of Land Use and Land Cover*. Ed. by C Giri. Remote Sensing Applications Series. CRC Press, pp.49–64. ISBN: 978-1-4200-7074-3.
- DeFries, RS, M Hansen, JRG Townshend, and R Sohlberg (1998). Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. *International Journal of Remote Sensing* **19**(16), 3141–3168. ISSN: 0143-1161. doi: 10.1080/014311698214235.
- Di Gregorio, A (2005). *Land Cover Classification System (LCCS), version 2: Classification System and Concepts*. Tech. rep. 8. Rome, Italy: Food and Agriculture Organization of the United Nations.

- Dobson, MC, L Pierce, and FT Ulaby (1996). Knowledge-based land-cover classification using ERS-1/JERS-1 SAR composites. *IEEE Transactions on Geoscience and Remote Sensing* **34**(1), 83–99. ISSN: 0196-2892. doi: 10.1109/36.481896.
- Dong, J, X Xiao, S Sheldon, C Biradar, G Zhang, N Dinh Duong, M Hazarika, K Wikantika, W Takeuhci, and B Moore III (2014). A 50-m forest cover map in Southeast Asia from ALOS/PALSAR and its application on forest fragmentation assessment. *PLoS ONE* **9**(1), e85801. doi: 10.1371/journal.pone.0085801.
- Durden, SL, JJ Van Zyl, and HA Zebker (1989). Modeling and observation of the radar polarization signature of forested areas. *IEEE Transactions on Geoscience and Remote Sensing* **27**(3), 290–301.
- Eder, JF (1990). Deforestation and detribalization in the Philippines: the Palawan case. *Population and Environment* **12**(2), 99–115. ISSN: 0199-0039, 1573-7810. doi: 10.1007/BF01261477.
- Evans, TL, M Costa, WM Tomas, and AR Camilo (2014). Large-scale habitat mapping of the Brazilian Pantanal wetland: a synthetic aperture radar approach. *Remote Sensing of Environment* **155**, 89–108. ISSN: 0034-4257. doi: 10.1016/j.rse.2013.08.051.
- Evans, TL and M Costa (2013). Landcover classification of the Lower Nhecolândia subregion of the Brazilian Pantanal Wetlands using ALOS/PALSAR, RADARSAT-2 and ENVISAT/ASAR imagery. *Remote Sensing of Environment* **128**, 118–137. ISSN: 0034-4257. doi: 10.1016/j.rse.2012.09.022.
- Evans, TL, M Costa, K Telmer, and TSF Silva (2010). Using ALOS/PALSAR and RADARSAT-2 to map land cover and seasonal inundation in the Brazilian Pan-

- tanal. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **3**(4), 560–575.
- FAO (2005). *The 2005 Global Forest Resource Assessment: Country Report, Philippines*. Tech. rep. FRA2005/202. Rome, Italy: Food and Agriculture Organization of the United Nations, p. 46.
- FAO (2010). *The 2010 Global Forest Resource Assessment: Country Report, Philippines*. Tech. rep. FRA2010/164. Rome, Italy: Food and Agriculture Organization of the United Nations, p. 66.
- FAO (2015a). *Forests and Poverty Reduction*. <http://www.fao.org/forestry/livelelihoods/en/> (visited on 03/15/2015).
- FAO (2015b). *Global Forest Resource Assessment 2015: How Are the World's Forests Changing?* Tech. rep. Rome, Italy: Food and Agriculture Organization of the United Nations, p. 56. <http://www.fao.org/forest-resources-assessment/en/>.
- Fernando, ES, MH Suh, J Lee, and DK Lee (2008). *Forest Formations of the Philippines*. English. Seoul, South Korea: ASEAN-Korea Environmental Cooperation Unit (AKECU). ISBN: 978-89-92239-40-0 93530.
- Fisher, P (1997). The pixel: a snare and a delusion. *International Journal of Remote Sensing* **18**(3), 679–685. ISSN: 0143-1161. DOI: 10.1080/014311697219015.
- Flores De Santiago, F, JM Kovacs, and P Lafrance (2013). An object-oriented classification method for mapping mangroves in Guinea, West Africa, using multipolarized ALOS PALSAR L-band data. *International Journal of Remote Sensing* **34**(2), 563–586. ISSN: 0143-1161. DOI: 10.1080/01431161.2012.715773.

FMB (2015). *Forestry Statistics at a Glance*. <http://forestry.denr.gov.ph/statglance.pdf> (visited on 07/17/2015).

Foody, GM (2003). Remote sensing of tropical forest environments: towards the monitoring of environmental resources for sustainable development. *International Journal of Remote Sensing* **24**(20), 4035–4046. doi: 10.1080/0143116031000103853.

Fowlkes, EB and CL Mallows (1983). A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association* **78**(383), 553–569. ISSN: 0162-1459. doi: 10.1080/01621459.1983.10478008.

Franklin, S (2001). *Remote Sensing for Sustainable Forest Management*. CRC Press. ISBN: 978-1-56670-394-9 978-1-4200-3285-7.

Freeman, A and SL Durden (1998). A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing* **36**(3), 963–973.

Friedl, MA and CE Brodley (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment* **61**(3), 399–409. ISSN: 0034-4257. doi: 10.1016/S0034-4257(97)00049-7.

Galili, T (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, btv428. ISSN: 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btv428.

Galili, T, G Simpson, GJ Jefferis, M Gallota, J Renaudie, The R. Core Team, K Hornik, U Ligges, AN Spiess, S Horvath, P Langfelder, skullkey, M Van Der Loo, A de Vries, and Y Benjamini (2015). *dendextend: Extending R's Dendrogram Functionality*. <http://cran.r-project.org/web/packages/dendextend/index.html>.

Gallardo-Cruz, JA, JA Meave, EJ González, EE Lebrija-Trejos, MA Romero-Romero, EA Pérez-García, R Gallardo-Cruz, JL Hernández-Stefanoni, and C Martorell (2012). Predicting tropical dry forest successional attributes from space: is the key hidden in image texture? *PLoS ONE* **7**(2), e30506. doi: 10.1371/journal.pone.0030506.

Geist, HJ and EF Lambin (2002). Proximate causes and underlying driving forces of tropical deforestation tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. *BioScience* **52**(2), 143–150. ISSN: 0006-3568, 1525-3244. doi: 10.1641/0006-3568(2002)052[0143:PCAUDF]2.0.CO;2.

Geofabrik GmbH and OpenStreetMap Contributors (2015). *Philippines*. <http://download.geofabrik.de/asia/philippines.html> (visited on 04/27/2015).

Gibbs, HK, AS Ruesch, F Achard, MK Clayton, P Holmgren, N Ramankutty, and JA Foley (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proceedings of the National Academy of Sciences* **107**(38), 16732–16737. ISSN: 0027-8424, 1091-6490. doi: 10.1073/pnas.0910275107.

Giri, CP (2012). “Brief Overview of Remote Sensing of Land Cover”. In: *Remote Sensing of Land Use and Land Cover*. Remote Sensing Applications Series. CRC Press, pp.3–12. ISBN: 978-1-4200-7074-3.

Hair, JF, WC Black, BJ Babin, and RE Anderson (2013). *Multivariate Data Analysis: Pearson New International Edition*. 7th ed. Pearson.

Hamdan, O, H Khali Aziz, and I Mohd Hasmadi (2014). L-band ALOS PALSAR for biomass estimation of Matang Mangroves, Malaysia. *Remote Sensing of Environment* **155**, 69–78. ISSN: 0034-4257. DOI: 10.1016/j.rse.2014.04.029.

Hansen, MC, RS Defries, JRG Townshend, and R Sohlberg (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing* **21**(6-7), 1331–1364. DOI: 10.1080 / 014311600210209.

Hansen, MC, PV Potapov, R Moore, M Hancher, SA Turubanova, A Tyukavina, D Thau, SV Stehman, SJ Goetz, TR Loveland, A Kommareddy, A Egorov, L Chini, CO Justice, and JRG Townshend (2013). High-resolution global maps of 21st-century forest cover change. *Science* **342**(6160), 850–853. DOI: 10.1126/science.1244693.

Hansen, M, R Dubayah, and R DeFries (1996). Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing* **17**(5), 1075–1081. ISSN: 0143-1161. DOI: 10.1080/01431169608949069.

Happ, PN, RS Ferreira, C Bentes, GAOP Costa, and RQ Feitosa (2010). Multi-resolution segmentation: a parallel approach for high resolution image segmentation in multi-core architectures. In: *ISPRS Archives*. Vol. XXXVIII. Ghent, Belgium: International Society of Photogrammetry and Remote Sensing, pp.345–349.

Haralick, RM, K Shanmugam, and I Dinstein (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics* **3**(6), 610–621. DOI: 10.1109/TSMC.1973.4309314.

- Hay, GJ and G Castilla (2008). “Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline”. In: *Object-Based Image Analysis*. Ed. by T Blaschke, S Lang, and GJ Hay. Lecture Notes in Geoinformation and Cartography. DOI: 10.1007/978-3-540-77058-9_4. Springer Berlin Heidelberg, pp.75–89. ISBN: 978-3-540-77057-2 978-3-540-77058-9.
- Herold, M and A Di Gregorio (2012). “Evaluating Land-Cover Legends Using the UN Land-Cover Classification System”. In: *Remote Sensing of Land Use and Land Cover*. Remote Sensing Applications Series. CRC Press, pp.65–90. ISBN: 978-1-4200-7074-3.
- Hoekman, DH, M Vissers, and N Wielaard (2010). PALSAR wide-area mapping of Borneo: methodology and map validation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **3**(4), 605–617. ISSN: 1939-1404. DOI: 10.1109/JSTARS.2010.2070059.
- Hong, SH, HO Kim, S Wdowinski, and E Feliciano (2015). Evaluation of polarimetric SAR decomposition for classifying wetland vegetation types. *Remote Sensing* **7**(7), 8563–8585. DOI: 10.3390/rs70708563.
- Horning, N, JA Robinson, E Sterling, W Turner, and S Spector (2010). *Remote Sensing for Ecology and Conservation*. New York, USA: Oxford University Press.
- Imhoff, M (1995). A theoretical analysis of the effect of forest structure on synthetic aperture radar backscatter and the remote sensing of biomass. *IEEE Transactions on Geoscience and Remote Sensing* **33**(2), 341–352. ISSN: 0196-2892. DOI: 10.1109/36.377934.

- Johansen, K, R Bartolo, and S Phinn (2010). SPECIAL FEATURE – Geographic Object-Based Image Analysis. *Journal of Spatial Science* **55**(1), 3–7. doi: 10.1080/14498596.2010.494653.
- Johnson, SC (1967). Hierarchical clustering schemes. *Psychometrika* **32**(3), 241–254. ISSN: 0033-3123, 1860-0980. doi: 10.1007/BF02289588.
- Joshi, N, ET Mitchard, N Woo, J Torres, J Moll-Rocek, A Ehammer, Murray Collins, MR Jepsen, and R Fensholt (2015). Mapping dynamics of deforestation and forest degradation in tropical forests using radar satellite data. *Environmental Research Letters* **10**(3), 034014. ISSN: 1748-9326. doi: 10.1088/1748-9326/10/3/034014.
- Kalensky, ZD (1998). AFRICOVER Land Cover Database and Map of Africa. *Canadian Journal of Remote Sensing* **24**(3), 292–297. ISSN: 0703-8992. doi: 10.1080/07038992.1998.10855250.
- Kaufman, L and PJ Rousseeuw (1990). “Introduction”. In: *Finding Groups in Data*. John Wiley & Sons, Inc., pp.1–67. ISBN: 978-0-470-31680-1.
- Keenan, RJ, GA Reams, F Achard, JVd Freitas, A Grainger, and E Lindquist (2015). Dynamics of global forest area: results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management* **352**. Changes in Global Forest Resources from 1990 to 2015, 9–20. ISSN: 0378-1127. doi: <http://dx.doi.org/10.1016/j.foreco.2015.06.014>.
- Kellndorfer, JM, L Pierce, MC Dobson, and FT Ulaby (1998). Toward consistent regional-to-global-scale vegetation characterization using orbital SAR systems. *IEEE Transactions on Geoscience and Remote Sensing* **36**(5), 1396–1411. ISSN: 0196-2892. doi: 10.1109/36.718844.

- Kindermann, G, I McCallum, S Fritz, and M Obersteiner (2008). A global forest growing stock, biomass and carbon map based on FAO statistics. *Silva Fennica* **42**(3). ISSN: 22424075. DOI: 10.14214/sf.244.
- Kuenzer, C, P Leinenkugel, M Vollmuth, and S Dech (2014). Comparing global land-cover products – implications for geoscience applications: an investigation for the trans-boundary Mekong Basin. *International Journal of Remote Sensing* **35**(8), 2752–2779. ISSN: 0143-1161. DOI: 10.1080/01431161.2014.890305.
- Kummer, DM (1992a). Measuring forest decline in the Philippines: an exercise in historiography. *Forest & Conservation History* **36**(4), 185–189. ISSN: 1046-7009, DOI: 10.2307/3983680.
- Kummer, DM (1992b). Remote sensing and tropical deforestation: a cautionary note from the Philippines. *Photogrammetric Engineering and Remote Sensing* **58**(10), 1469–1471.
- Kuplich, TM, PJ Curran, and PM Atkinson (2005). Relating SAR image texture to the biomass of regenerating tropical forests. *International Journal of Remote Sensing* **26**(21), 4829–4854. DOI: 10.1080/01431160500239107.
- Kurvanen, L and MT Hallikainen (1999). Textural information of multitemporal ERS-1 and JERS-1 SAR images with applications to land and forest type classification in boreal zone. *IEEE Transactions on Geoscience and Remote Sensing* **37**(2), 680–689. ISSN: 0196-2892. DOI: 10.1109/36.752185.
- Lachowski, HM, DL Dietrich, R Umali, R Aquino, and V Basa (1979). Landsat assisted forest land-cover assessment of the Philippine Islands. *Photogrammetric Engineering and Remote Sensing* **45**(10), 1387–1391. ISSN: 0099-1112.

- Lasco, RD and FB Pulhin (2000). Forest land use change in the Philippines and climate change mitigation. *Mitigation and Adaptation Strategies for Global Change* **5**(1), 81–97. ISSN: 1381-2386, 1573-1596. DOI: 10.1023/A:1009629220978.
- Le Toan, T, A Beaudoin, J Riom, and D Guyon (1992). Relating forest biomass to SAR data. *IEEE Transactions on Geoscience and Remote Sensing* **30**(2), 403–411.
- Lee, JS, L Jurkevich, P Dewaele, P Wambacq, and A Oosterlinck (1994). Speckle filtering of synthetic aperture radar images: a review. *Remote Sensing Reviews* **8**(4), 313–340. ISSN: 0275-7257. DOI: 10.1080/02757259409532206.
- Lee, JS (1980). Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-2**(2), 165–168.
- Lee, JS (1981). Speckle analysis and smoothing of synthetic aperture radar images. *Computer Graphics and Image Processing* **17**(1), 24–32. ISSN: 0146-664X. DOI: 10.1016/S0146-664X(81)80005-6.
- Li, G, D Lu, E Moran, L Dutra, and M Batistella (2012). A comparative analysis of ALOS PALSAR L-band and RADARSAT-2 C-band data for land-cover classification in a tropical moist region. *ISPRS Journal of Photogrammetry and Remote Sensing* **70**, 26–38. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2012.03.010.
- Liesenberg, V and R Gloaguen (2013). Evaluating SAR polarization modes at L-band for forest classification purposes in Eastern Amazon, Brazil. *International Journal of Applied Earth Observation and Geoinformation* **21**, 122–135. ISSN: 0303-2434. DOI: 10.1016/j.jag.2012.08.016.

Liu, DS, LR Iverson, and S Brown (1993). Rates and patterns of deforestation in the Philippines: application of geographic information system analysis. *Forest Ecology and Management* **57**(1-4), 1–16. ISSN: 0378-1127. doi: DOI:10.1016/0378-1127(93)90158-J.

Long, J, D Napton, C Giri, and J Graesser (2013). A mapping and monitoring assessment of the Philippines' mangrove forests from 1990 to 2010. *Journal of Coastal Research*. ISSN: 0749-0208. doi: 10.2112/JCOASTRES-D-13-00057.1.

Longépé, N, P Rakwatin, O Isoguchi, and M Shimada (2011). Assessment of ALOS PALSAR 50 m orthorectified FBD data for regional land cover classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* **49**(6), 1–16.

MacDicken, KG, P Sola, JE Hall, C Sabogal, M Tadoum, and Cd Wasseige (2015). Global progress toward sustainable forest management. *Forest Ecology and Management* **352**. Changes in Global Forest Resources from 1990 to 2015, 47–56. ISSN: 0378-1127. doi: <http://dx.doi.org/10.1016/j.foreco.2015.02.005>.

Mather, PM and M Koch (2011). *Computer Processing of Remotely-Sensed Images: An Introduction*. 4th ed. West Sussex, United Kingdom: Wiley-Blackwell. ISBN: 978-0-470-74238-9.

Mayaux, P, P Holmgren, F Achard, H Eva, HJ Stibig, and A Branthomme (2005). Tropical forest cover change in the 1990s and options for future monitoring. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **360**(1454), 373–384. doi: 10.1098/rstb.2004.1590.

Mermoz, S and T Le Toan (2016). Forest disturbances and regrowth assessment using ALOS PALSAR data from 2007 to 2010 in Vietnam, Cambodia and Lao PDR.

Remote Sensing **8**(3), 217. DOI: 10.3390/rs8030217.

Mermoz, S, M Réjou-Méchain, L Villard, T Le Toan, V Rossi, and S Gourlet-Fleury

(2015). Decrease of L-band SAR backscatter with biomass of dense forests. *Remote Sensing of Environment* **159**, 307–317. ISSN: 0034-4257. DOI: 10.1016/j.rse.2014.

12.019.

Michaelsen, J, DS Schimel, MA Friedl, FW Davis, and RC Dubayah (1994). Regression

tree analysis of satellite and terrain data to guide vegetation sampling and surveys.

Journal of Vegetation Science **5**(5), 673–686. ISSN: 1654-1103. DOI: 10.2307 / 3235882.

Mishra, P, D Singh, and Y Yamaguchi (2011). Land cover classification of PALSAR

images by knowledge based decision tree classifier and supervised classifiers based

on SAR observables. *Progress in Electromagnetics Research B* **30**, 47–70. DOI:

10.2528/PIERB11011405.

Mitchard, ETA, SS Saatchi, IH Woodhouse, G Nangendo, NS Ribeiro, M Williams,

CM Ryan, SL Lewis, TR Feldpausch, and P Meir (2009). Using satellite radar

backscatter to predict above-ground woody biomass: a consistent relationship

across four different African landscapes. *Geophysical Research Letters* **36**(23), 1–6.

Morales-Hidalgo, D, SN Oswalt, and E Somanathan (2015). Status and trends in

global primary forest, protected areas, and areas designated for conservation of

biodiversity from the Global Forest Resources Assessment 2015. *Forest Ecology*

- and Management* **352**. Changes in Global Forest Resources from 1990 to 2015, 68–77. ISSN: 0378-1127. DOI: <http://dx.doi.org/10.1016/j.foreco.2015.06.011>.
- Morel, AC, SS Saatchi, Y Malhi, NJ Berry, L Banin, D Burslem, R Nilus, and RC Ong (2011). Estimating aboveground biomass in forest and oil palm plantation in Sabah, Malaysian Borneo using ALOS PALSAR data. *Forest Ecology and Management* **262**(9), 1786–1798. ISSN: 0378-1127. DOI: [10.1016/j.foreco.2011.07.008](https://doi.org/10.1016/j.foreco.2011.07.008).
- Motohka, T, M Shimada, Y Uryu, and B Setiabudi (2014). Using time series PALSAR gamma nought mosaics for automatic detection of tropical deforestation: a test study in Riau, Indonesia. *Remote Sensing of Environment* **155**, 79–88. ISSN: 0034-4257. DOI: [10.1016/j.rse.2014.04.012](https://doi.org/10.1016/j.rse.2014.04.012).
- Munang, R, I Thiaw, J Thompson, D Ganz, EH Girvetz, and M Rivington (2011). *Sustaining Forests: Investing in Our Common Future*. http://www.unep.org/ecosystemmanagement/Portals/7/Documents/unep_policy_series/5thUNEPPolicySeries.pdf.
- Myers, N (1988). Tropical deforestation and remote sensing. *Forest Ecology and Management* **23**(2-3), 215–225. ISSN: 0378-1127. DOI: [10.1016/0378-1127\(88\)90083-7](https://doi.org/10.1016/0378-1127(88)90083-7).
- Nascimento, WR, PWM Souza-Filho, C Proisy, RM Lucas, and A Rosenqvist (2013). Mapping changes in the largest continuous Amazonian mangrove belt using object-based classification of multisensor satellite imagery. *Estuarine, Coastal and Shelf Science* **117**, 83–93. ISSN: 0272-7714. DOI: [10.1016/j.ecss.2012.10.005](https://doi.org/10.1016/j.ecss.2012.10.005).

- Oliver, CJ (2000). Rain forest classification based on SAR texture. *IEEE Transactions on Geoscience and Remote Sensing* **38**(2), 1095–1104. ISSN: 0196-2892. DOI: 10.1109/36.841988.
- Otukei, JR and T Blaschke (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*. Supplement Issue on "Remote Sensing for Africa – A Special Collection from the African Association for Remote Sensing of the Environment (AARSE)" **12, Supplement 1**, S27–S31. ISSN: 0303-2434. DOI: 10.1016/j.jag.2009.11.002.
- Otukei, JR, T Blaschke, and M Collins (2014). Using the Wishart maximum likelihood classifier for assessing the potential of TerraSAR-X and ALOS PALSAR data for land cover mapping. *International Journal of Image and Data Fusion* **5**(2), 138–151. ISSN: 1947-9832. DOI: 10.1080/19479832.2013.870933.
- Pal, M and PM Mather (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment* **86**(4), 554–565. ISSN: 0034-4257. DOI: 10.1016/S0034-4257(03)00132-9.
- Pan, Y, RA Birdsey, J Fang, R Houghton, PE Kauppi, WA Kurz, OL Phillips, A Shvidenko, SL Lewis, JG Canadell, P Ciais, RB Jackson, SW Pacala, AD McGuire, S Piao, A Rautiainen, S Sitch, and D Hayes (2011). A large and persistent carbon sink in the world's forests. *Science* **333**(6045), 988–993. DOI: 10.1126/science.1201609.

- Pan, Y, RA Birdsey, OL Phillips, and RB Jackson (2013). The structure, distribution, and biomass of the world's forests. *Annual Review of Ecology, Evolution, and Systematics* **44**(1), 593–622. DOI: 10.1146/annurev-ecolsys-110512-135914.
- Payn, T, JM Carnus, P Freer-Smith, M Kimberley, W Kollert, S Liu, C Orazio, L Rodriguez, LN Silva, and MJ Wingfield (2015). Changes in planted forests and future global implications. *Forest Ecology and Management* **352**. Changes in Global Forest Resources from 1990 to 2015, 57–67. ISSN: 0378-1127. DOI: <http://dx.doi.org/10.1016/j.foreco.2015.06.021>.
- Pearce, DW (2001). The economic value of forest ecosystems. *Ecosystem Health* **7**(4), 284–296. ISSN: 1526-0992. DOI: 10.1046/j.1526-0992.2001.01037.x.
- Pereira, RA, J Epting, D Juhn, O Coroza, L Miller, and F Maon (2006). Forest clearance and fragmentation in Palawan and Eastern Mindanao biodiversity corridors (1990-2000): a time sequential analysis of Landsat imagery. *Banwa* **3**(1 & 2), 130–147.
- Pierce, LE, FT Ulaby, K Sarabandi, and MC Dobson (1994). Knowledge-based classification of polarimetric SAR images. *IEEE Transactions on Geoscience and Remote Sensing* **32**(5), 1081–1086. ISSN: 0196-2892. DOI: 10.1109/36.312896.
- Pimentel, D, M McNair, L Buck, M Pimentel, and J Kamil (1997). The value of forests to world food security. *Human Ecology* **25**(1), 91–120. ISSN: 0300-7839, 1572-9915. DOI: 10.1023/A:1021987920278.
- Podani, J (2000). *Introduction to the Exploration of Multivariate Biological Data*. Leiden: Backhuys Publishers. ISBN: 978-90-5782-067-0.

- Podest, E and S Saatchi (2002). Application of multiscale texture in classifying JERS-1 radar data over tropical vegetation. *International Journal of Remote Sensing* **23**(7), 1487–1506. ISSN: 0143-1161. DOI: 10.1080/01431160110093000.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rahman, MM and JTS Sumantyo (2010). Mapping tropical forest cover and deforestation using synthetic aperture radar (SAR) images. *Applied Geomatics* **2**(3). 10.1007/s12518-010-0026-9, 113–121. ISSN: 1866-9298.
- Rahman, MM and JTS Sumantyo (2013). Retrieval of tropical forest biomass information from ALOS PALSAR data. *Geocarto International* **28**(5), 382–403. ISSN: 1010-6049. DOI: 10.1080/10106049.2012.710652.
- Rakwatin, P, N Longépé, O Isoguchi, M Shimada, Y Uryu, and W Takeuchi (2012). Using multiscale texture information from ALOS PALSAR to map tropical forest. *International Journal of Remote Sensing* **33**(24), 7727–7746. ISSN: 0143-1161, 1366-5901. DOI: 10.1080/01431161.2012.701349.
- Ripley, B (2015). *tree: Classification and Regression Trees*. <https://cran.r-project.org/web/packages/tree/index.html>.
- Ripley, BD (1996). *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press. ISBN: 978-0-511-81265-1.
- Rocha de Souza Pereira, F, M Kampel, and M Cunha-Lignon (2012). Mapping of mangrove forests on the southern coast of São Paulo, Brazil, using synthetic aperture radar data from ALOS/PALSAR. *Remote Sensing Letters* **3**(7), 567–576. ISSN: 2150-704X. DOI: 10.1080/01431161.2011.641511.

Rosenqvist, A, T Ogawa, M Shimada, and T Igarashi (2001). Initiating the ALOS Kyoto & Carbon Initiative. In: *Proceedings of the 2001 International Geoscience and Remote Sensing Symposium (IGARSS)*. Vol. 1. Sydney, Australia: IEEE, pp.546–548. doi: 10.1109/IGARSS.2001.976217.

Rosenqvist, A, M Shimada, BD Chapman, A Freeman, G De Grandi, SS Saatchi, and Y Rauste (2000). The Global Rain Forest Mapping project - A review. *International Journal of Remote Sensing* **21**(6-7), 1375–1387. doi: 10.1080/014311600210227.

Rosenqvist, A, M Shimada, BD Chapman, KC McDonald, GD De Grandi, H Jonsson, C Williams, Y Rauste, M Nilsson, D Sango, and M Matsumoto (2004). An overview of the JERS-1 SAR Global Boreal Forest Mapping (GBFM) project. In: *Proceedings of the 2004 International Geoscience and Remote Sensing Symposium (IGARSS)*. Vol. 2. Anchorage, Alaska: IEEE, pp.1033–1036. doi: 10.1109/IGARSS.2004.1368587.

Rosenqvist, A, M Shimada, N Ito, and M Watanabe (2007). ALOS PALSAR: a pathfinder mission for global-scale monitoring of the environment. *IEEE Transactions on Geoscience and Remote Sensing* **45**(11), 3307–3316. ISSN: 0196-2892. doi: 10.1109/TGRS.2007.901027.

Rosenqvist, A, M Shimada, RM Lucas, BD Chapman, P Paillou, LL Hess, and J Lowry (2010). The Kyoto & Carbon Initiative - a brief summary. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **3**(4), 551–553. ISSN: 1939-1404. doi: 10.1109/JSTARS.2010.2086270.

Running, SW, TR Loveland, LL Pierce, RR Nemani, and ER Hunt Jr. (1995). A remote sensing based vegetation classification logic for global land cover analysis.

Remote Sensing of Environment. Remote Sensing of Land Surface for Studies of Global Change **51**(1), 39–48. ISSN: 0034-4257. DOI: 10.1016/0034-4257(94)00063-S.

Saatchi, SS, B Nelson, E Podest, and J Holt (2000). Mapping land cover types in the Amazon Basin using 1 km JERS-1 mosaic. *International Journal of Remote Sensing* **21**(6-7), 1201–1234. ISSN: 0143-1161. DOI: 10.1080/014311600210146.

Sanden, JJ van der and DH Hoekman (1999). Potential of airborne radar to support the assessment of land cover in a tropical rain forest environment. *Remote Sensing of Environment* **68**(1), 26–40. ISSN: 0034-4257. DOI: 10.1016/S0034-4257(98)00099-6.

Sarker, MLR, J Nichol, B Ahmad, I Busu, and AA Rahman (2012). Potential of texture measurements of two-date dual polarization PALSAR data for the improvement of forest biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing* **69**, 146–166. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2012.03.002.

Sayed, S (2016). *Hierarchical Clustering*. <http://www.saedsayad.com/clustering-hierarchical.htm> (visited on 08/21/2014).

Shimada, M, T Itoh, T Motooka, M Watanabe, T Shiraishi, R Thapa, and R Lucas (2014). New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment* **155**, 13–31. ISSN: 0034-4257. DOI: 10.1016/j.rse.2014.04.014.

Shimada, M and T Ohtaki (2010). Generating large-scale high-quality SAR mosaic datasets: application to PALSAR data for global monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **3**(4), 637–656.

- Shimada, M, T Tadono, and A Rosenqvist (2010). Advanced Land Observing Satellite (ALOS) and monitoring global environmental change. *Proceedings of the IEEE* **98**(5), 780–799. ISSN: 0018-9219. DOI: 10.1109/JPROC.2009.2033724.
- Simard, M, G De Grandi, S Saatchi, and P Mayaux (2002). Mapping tropical coastal vegetation using JERS-1 and ERS-1 radar data with a decision tree classifier. *International Journal of Remote Sensing* **23**(7), 1461–1474. ISSN: 0143-1161. DOI: 10.1080/01431160110092984.
- Simard, M, SS Saatchi, and G De Grandi (2000). The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest. *IEEE Transactions on Geoscience and Remote Sensing* **38**(5), 2310–2321.
- Siry, JP, FW Cubbage, and MR Ahmed (2005). Sustainable forest management: global trends and opportunities. *Forest Policy and Economics* **7**(4), 551–561. ISSN: 1389-9341. DOI: 10.1016/j.foreco.2003.09.003.
- Sloan, S and JA Sayer (2015). Forest Resources Assessment of 2015 shows positive global trends but forest loss and degradation persist in poor tropical countries. *Forest Ecology and Management* **352**. Changes in Global Forest Resources from 1990 to 2015, 134–145. ISSN: 0378-1127. DOI: <http://dx.doi.org/10.1016/j.foreco.2015.06.013>.
- Sokal, RR and PHA Sneath (1963). *Principles of Numerical Taxonomy*. W. H. Freeman.
- Strobl, C, J Malley, and G Tutz (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* **14**(4), 323–348. ISSN: 1939-1463, 1082-989X. DOI: 10.1037/a0016973.

- Suzuki, R and H Shimodaira (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**(12), 1540–1542. ISSN: 1367-4803, 1460-2059. DOI: 10.1093/bioinformatics/btl117.
- Suzuki, R and H Shimodaira (2014). *pvcust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling*. <http://cran.r-project.org/web/packages/pvclust/index.html> (visited on 04/16/2015).
- Swedish Space Corporation (1988). *Mapping of the Natural Conditions of the Philipines*. Final Report FSF101. Solna, Sweden: Swedish Space Corporation, p. 73.
- Thapa, RB, T Itoh, M Shimada, M Watanabe, M Takeshi, and T Shiraishi (2014). Evaluation of ALOS PALSAR sensitivity for characterizing natural forest cover in wider tropical areas. *Remote Sensing of Environment* **155**, 32–41. ISSN: 0034-4257. DOI: 10.1016/j.rse.2013.04.025.
- Thapa, RB, M Shimada, M Watanabe, T Motohka, and T Shiraishi (2013). The tropical forest in south east Asia: monitoring and scenario modeling using synthetic aperture radar data. *Applied Geography* **41**, 168–178. ISSN: 0143-6228. DOI: 10.1016/j.apgeog.2013.04.009.
- Thapa, RB, M Watanabe, T Motohka, and M Shimada (2015). Potential of high-resolution ALOS–PALSAR mosaic texture for aboveground forest carbon tracking in tropical region. *Remote Sensing of Environment* **160**, 122–133. ISSN: 0034-4257. DOI: 10.1016/j.rse.2015.01.007.
- Thiel, C, C Thiel, T Riedel, and C Schmullius (2008). “Object based classification of SAR data for the delineation of forest cover maps and the detection of deforestation – A viable procedure and its application in GSE Forest Monitoring”. In: *Object-*

- Based Image Analysis*. Ed. by T Blaschke, S Lang, and GJ Hay. Lecture Notes in Geoinformation and Cartography. DOI: 10.1007/978-3-540-77058-9_18. Springer Berlin Heidelberg, pp.327–343. ISBN: 978-3-540-77057-2 978-3-540-77058-9.
- Ulaby, FT, F Kouyate, B Brisco, and THL Williams (1986). Textural information in SAR images. *IEEE Transactions on Geoscience and Remote Sensing* **24**(2), 235–245. ISSN: 0196-2892. DOI: 10.1109/TGRS.1986.289643.
- Van Zyl, JJ (1989). Unsupervised classification of scattering behavior using radar polarimetry data. *IEEE Transactions on Geoscience and Remote Sensing* **27**(1), 36–45.
- Walker, WS, CM Stickler, JM Kellndorfer, KM Kirsch, and DC Nepstad (2010). Large-area classification and mapping of forest and land cover in the Brazilian Amazon: a comparative analysis of ALOS/PALSAR and Landsat data sources. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **3**(4), 594–604.
- Whittle, M, S Quegan, Y Uryu, M Stüewe, and K Yulianto (2012). Detection of tropical deforestation using ALOS-PALSAR: A Sumatran case study. *Remote Sensing of Environment* **124**, 83–98. ISSN: 0034-4257. DOI: 10.1016/j.rse.2012.04.027.
- Wickham, H and W Chang (2015). *ggplot2: An Implementation of the Grammar of Graphics*. <http://cran.r-project.org/web/packages/ggplot2/index.html> (visited on 04/16/2015).
- Woodhouse, IH, ETA Mitchard, M Brolly, D Maniatis, and CM Ryan (2012). Radar backscatter is not a 'direct measure' of forest biomass. *Nature Climate Change* **2**(8), 556–557. ISSN: 1758-678X, 1758-6798. DOI: 10.1038/nclimate1601.

- Wu, J (1999). Hierarchy and scaling: extrapolating information along a scaling ladder. *Canadian Journal of Remote Sensing* **25**(4), 367–380. ISSN: 0703-8992. DOI: 10.1080/07038992.1999.10874736.
- Wu, X and V Kumar, eds. (2009). *The Top Ten Algorithms in Data Mining*. 1st ed. Boca Raton: Chapman and Hall/CRC. ISBN: 978-1-4200-8964-6.

Appendices

Appendix A

Supplementary: Literature Review

This Literature Review appendix presents the hierarchical structure of land cover classification adopted for the 2003/2010 NFA in the Philippines.

Table A.1: Hierarchical structure of land cover classification adopted for the 2003/2010 NFA in the Philippines (Source: FAO, 2005, 2010).

Land cover/forest type	Definition
Closed forest, broadleaved	Land spanning more than 0.5 ha with trees higher than 5 m and a canopy cover of more than 10%, or trees able to reach these thresholds in situ. It does not include land that is predominantly under agricultural or urban land use.
Closed forest, coniferous	
Closed forest, mixed	
Open forest, broadleaved	
Open forest, coniferous	
Open forest, mixed	
Mangrove forest	
Forest plantation, broadleaved	
Forest plantation, coniferous	
Forest plantation, mixed	
Other wooded land, shrubs	Land not classified as “Forest”, spanning more than 0.5 ha; with trees higher than 5 m and a canopy cover of 5-10%, or trees able to reach these thresholds in situ; or with a combined cover of shrubs, bushes and trees above 10%. It does not include land that is predominantly under agricultural or urban land use.
Other wooded land, fallow	
Other wooded land, wooded grassland	
Other land, barren land	All land that is not classified as “Forest” or “Other wooded land”.
Other land, grassland	
Other land, marshland	
Other land, annual crop	
Other land, perennial crop	
Other land, built-up area	
Inland water	Inland water bodies generally include major rivers, lakes and water reservoirs.
Fishpond	

Appendix B

Supplementary: Methods

This Methods appendix presents the land area statistics of forest cover types per province based on NAMRIA 2010 land cover map, and images of segmentation tests and the selected segmentation parameters.

Table B.1: Land area statistics of forest cover types per province based on NAMRIA 2010 land cover map.

Province	FCFB	FCFC	FCFM	FOFB	FOFC	FOFM	FFPB	FFPC	LVMG	NF	Total (ha)
Abra	27,025	9,812	6,479	54,884	35,508	11,119	1,874			250,940	397,640
Apayao	95,704			17,353			10,023			59,469	182,549
Aurora	51,933			60,265					511	49,888	162,598
Benguet	122	3,023	51	12,055	75,580	28,786		8		135,870	255,495
Cagayan	160,433			89,908			4,459			295,973	550,772
Ifugao	12,170	671	851	84,327	4,379					163,577	265,974
Ilocos Norte	884			17,003						83,313	101,199
Ilocos Sur		25	53	20,270	1,229	7,277	2,154	793	211	239,236	271,248
Isabela	69,444			307,852			254		723	662,263	1,040,536
Kalinga	40,305	22	8,561	39,113	8,949	1,912				186,971	285,833
La Union				2,558		2,165	1,037		120	139,801	145,681
Mt Province	17,955		9,523	20,072	32,211	7,505				110,635	197,900
N. Ecija	4,310			11,824		567				44,038	60,739
N. Vizcaya	119,583	2,351	555	57,812	5,028	5,994	1,307		217,853	410,482	
Pangasinan	181			5,160	924	280	812		556	304,127	312,040
Quirino	85,630			40,728			147			141,634	268,139
TOTAL	685,677	15,905	26,072	841,185	163,806	65,604	22,067	802	2,122	3,085,588	4,908,827
%	13.97	0.32	0.53	17.14	3.34	1.34	0.45	0.02	0.04	62.86	100.00

Note: FCFB: Closed forest, broadleaved; FCFC: Closed forest, coniferous; FCFM: Closed forest, mixed; FOFB: Open forest, broadleaved; FOFC: Open forest, coniferous; FOFM: Open forest, mixed; FFPB: Forest plantation, broadleaved; FFPC: Forest plantation, coniferous; LVMG: Mangrove forest; NF: Non-Forest

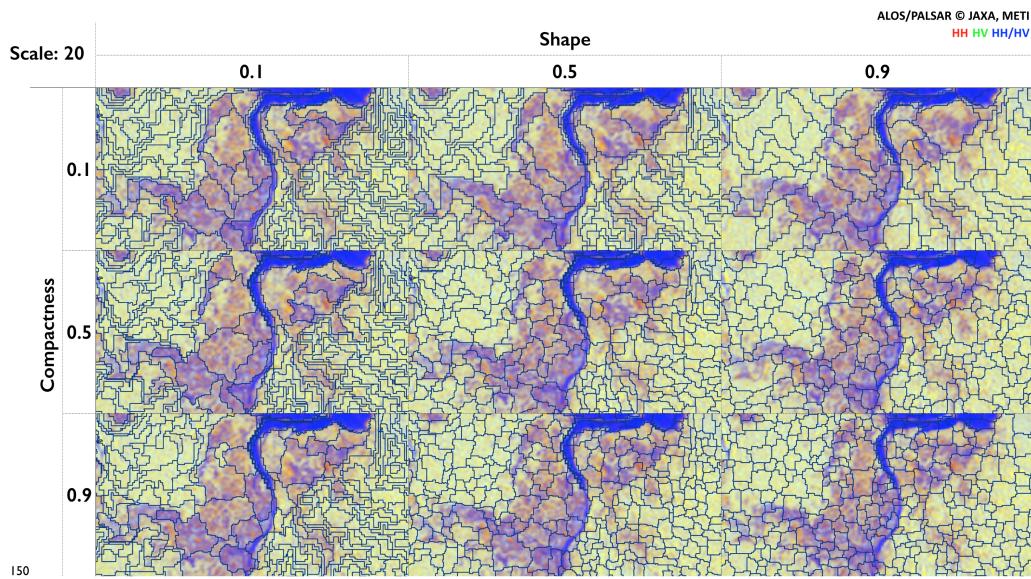


Fig. B.1: Test segmentation images using a scale value of 20 with shape and compactness values set at 0.1, 0.5, and 0.9. Image weights set at 1 for radar and topographic data layers.

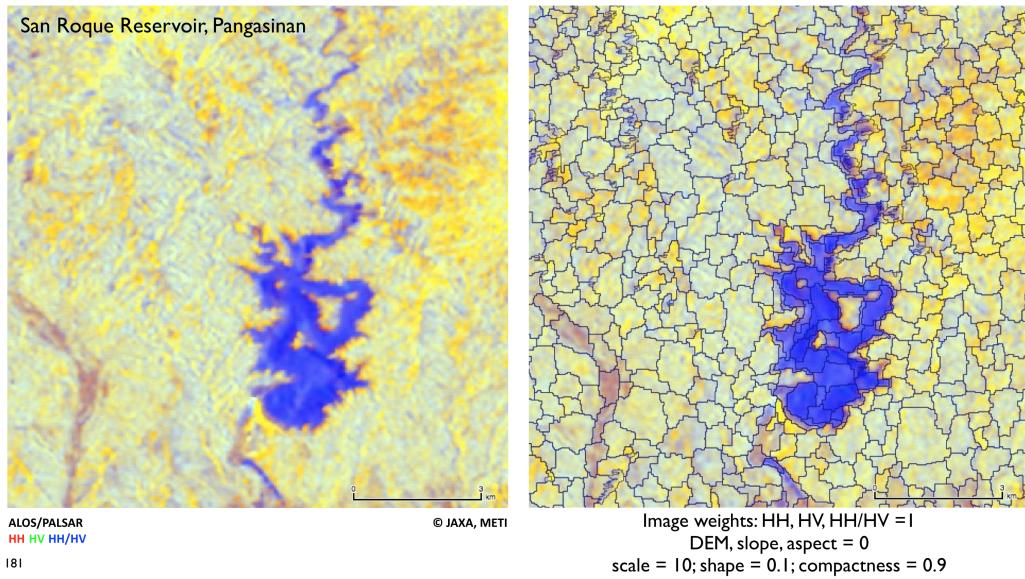


Fig. B.2: Selected segmentation parameters using a scale value of 10, shape value of 0.1, and compactness value at 0.9. Image weights set at 1 for radar data layers and 0 for topographic data layers.

Appendix C

Supplementary: Results

This Results appendix presents the classification trees and cross-validation graphs generated by R tree package.

C.1 Classification trees and cross-validation graphs

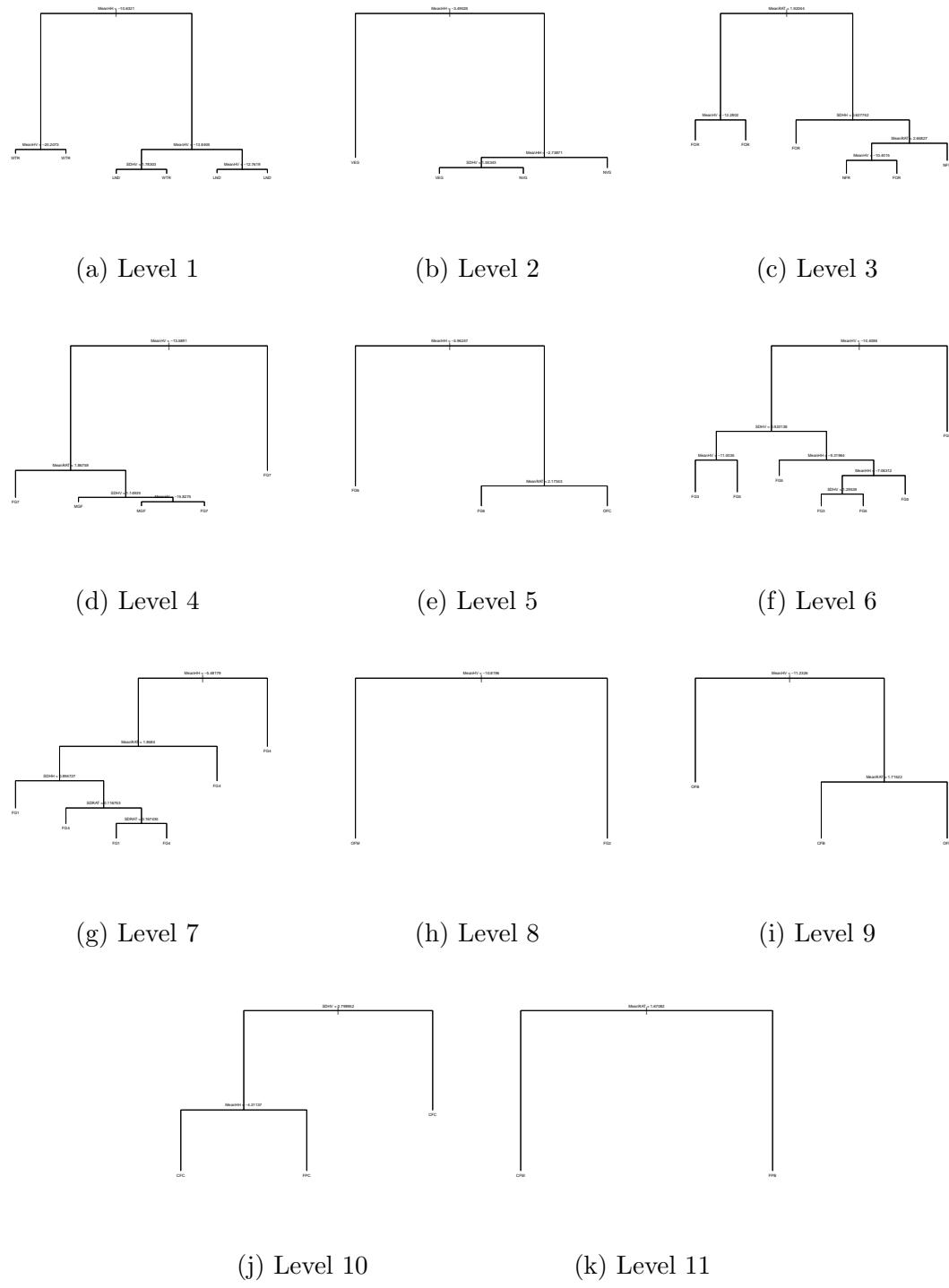


Fig. C.1: The Case 1 2007 classification trees generated by R tree package.

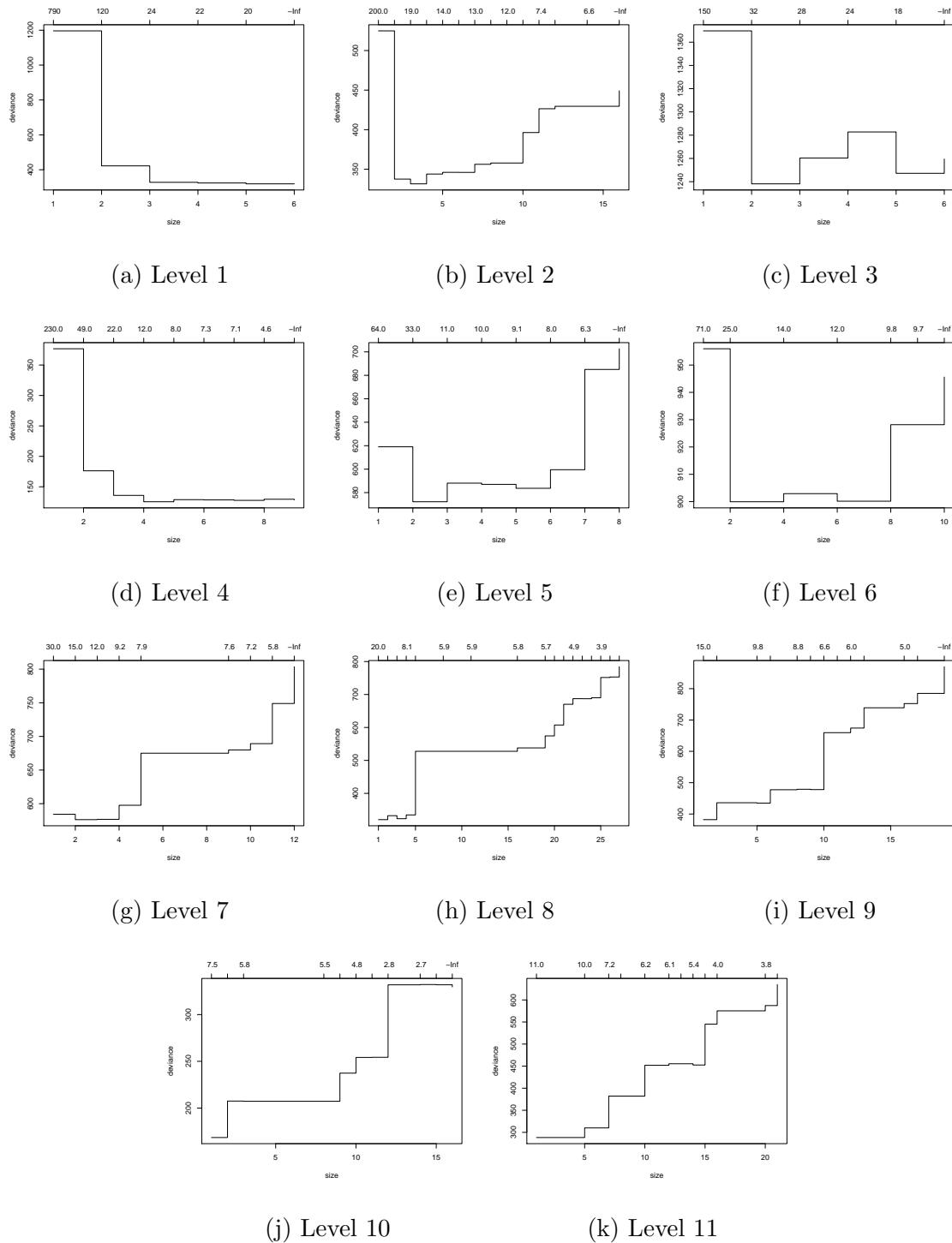


Fig. C.2: The Case 1 2007 cross-validation graphs generated by R tree package.

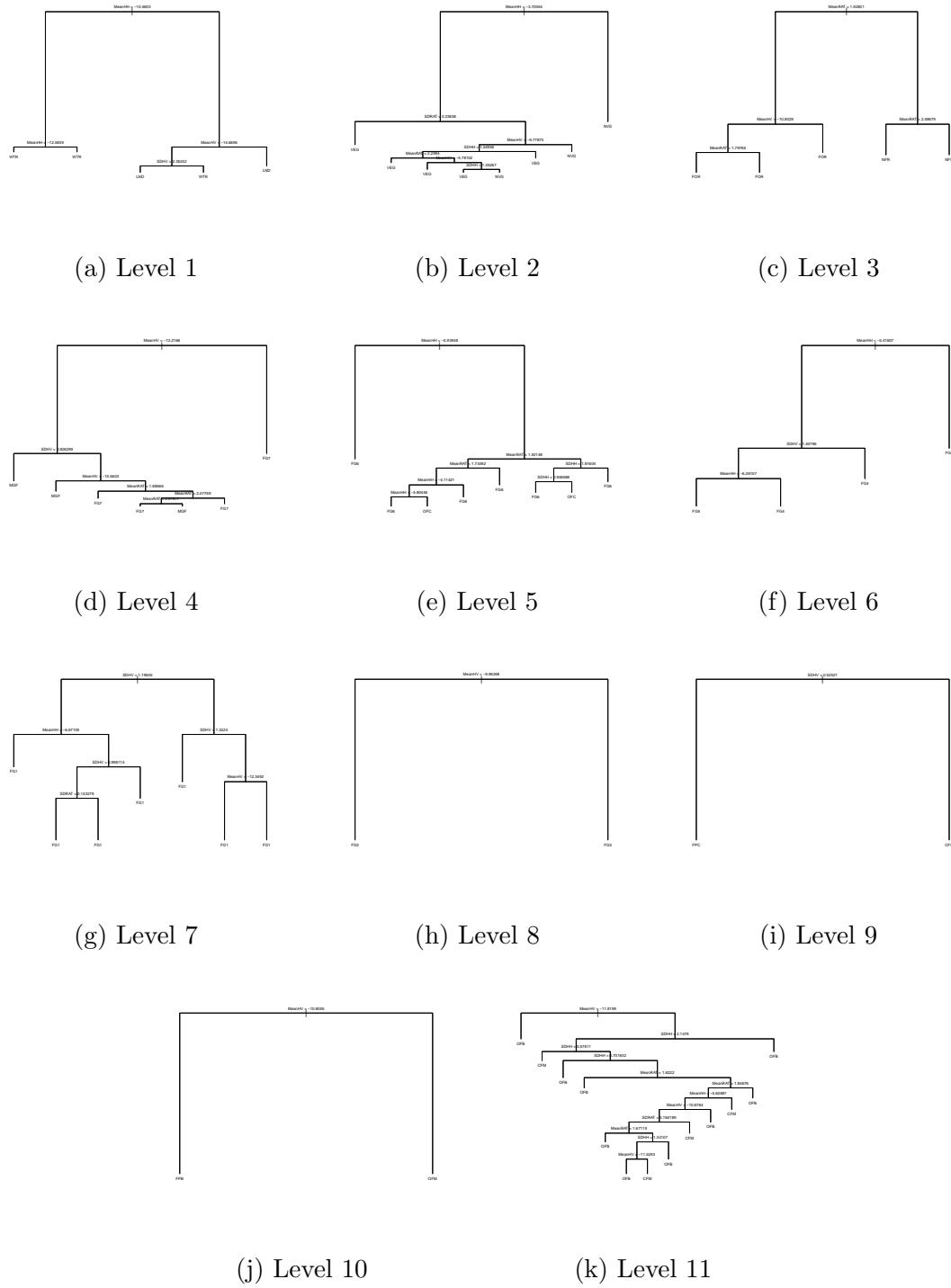


Fig. C.3: The Case 1 2008 classification trees generated by R tree package.

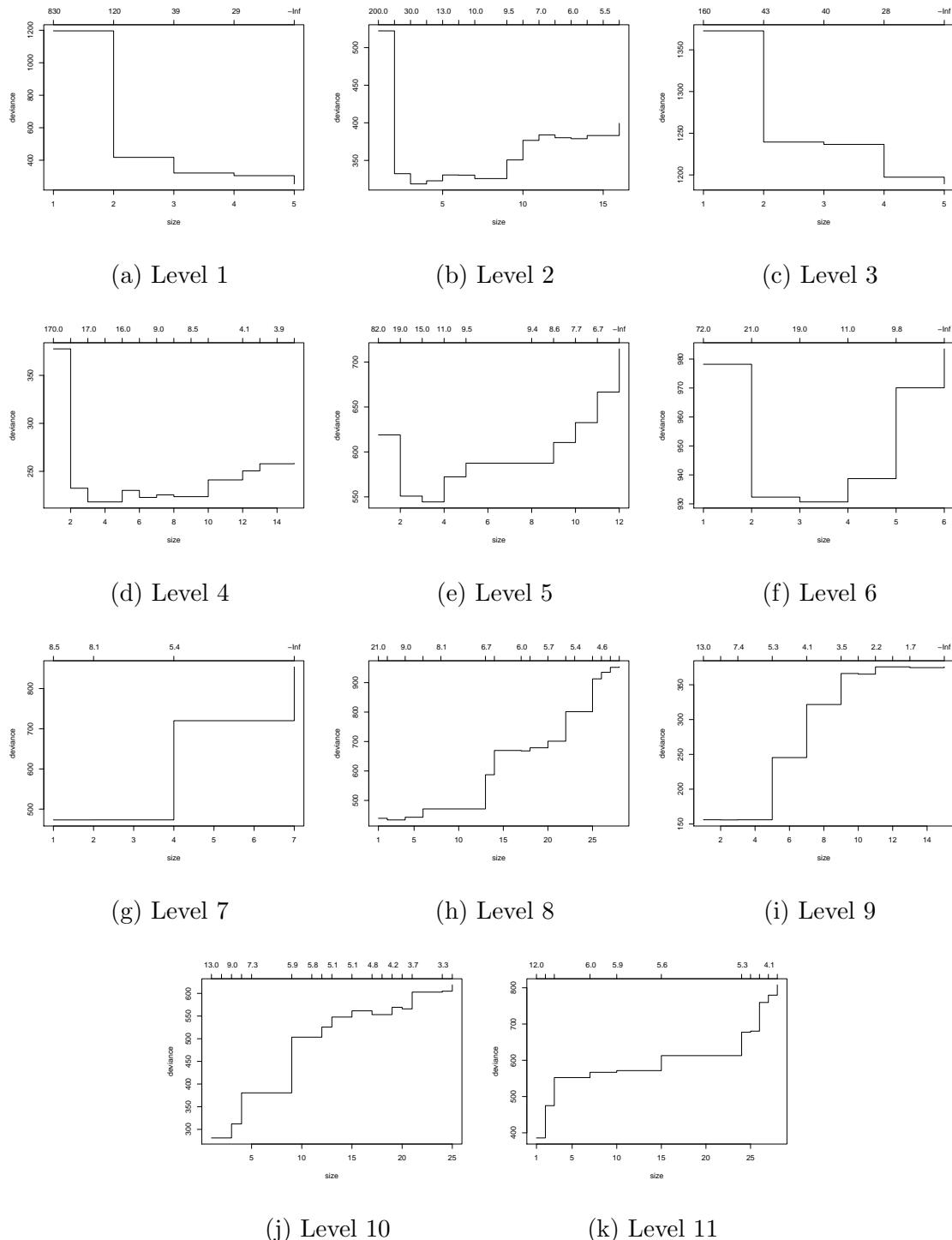


Fig. C.4: The Case 1 2008 cross-validation graphs generated by R tree package.

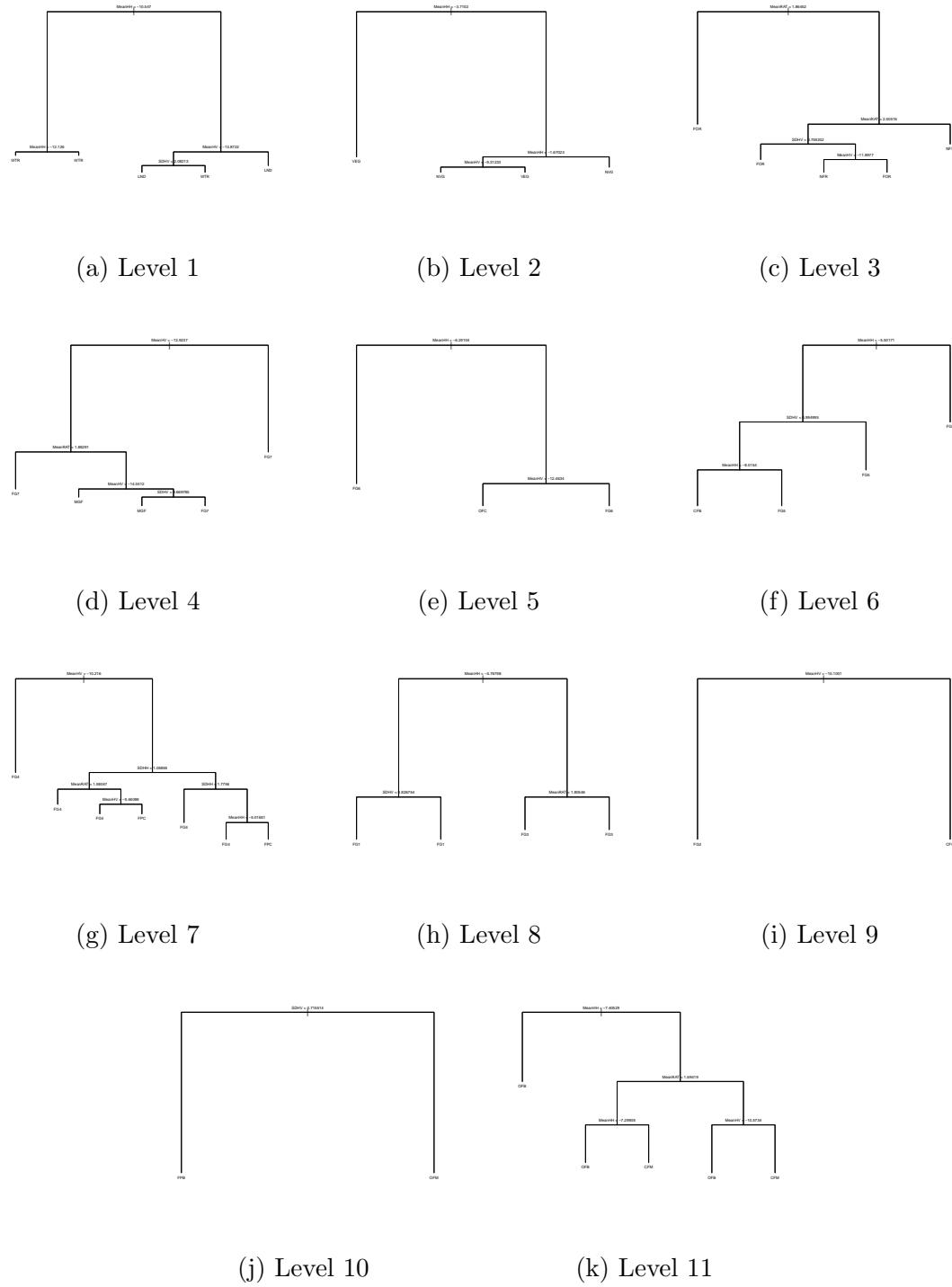


Fig. C.5: The Case 1 2009 classification trees generated by R tree package.

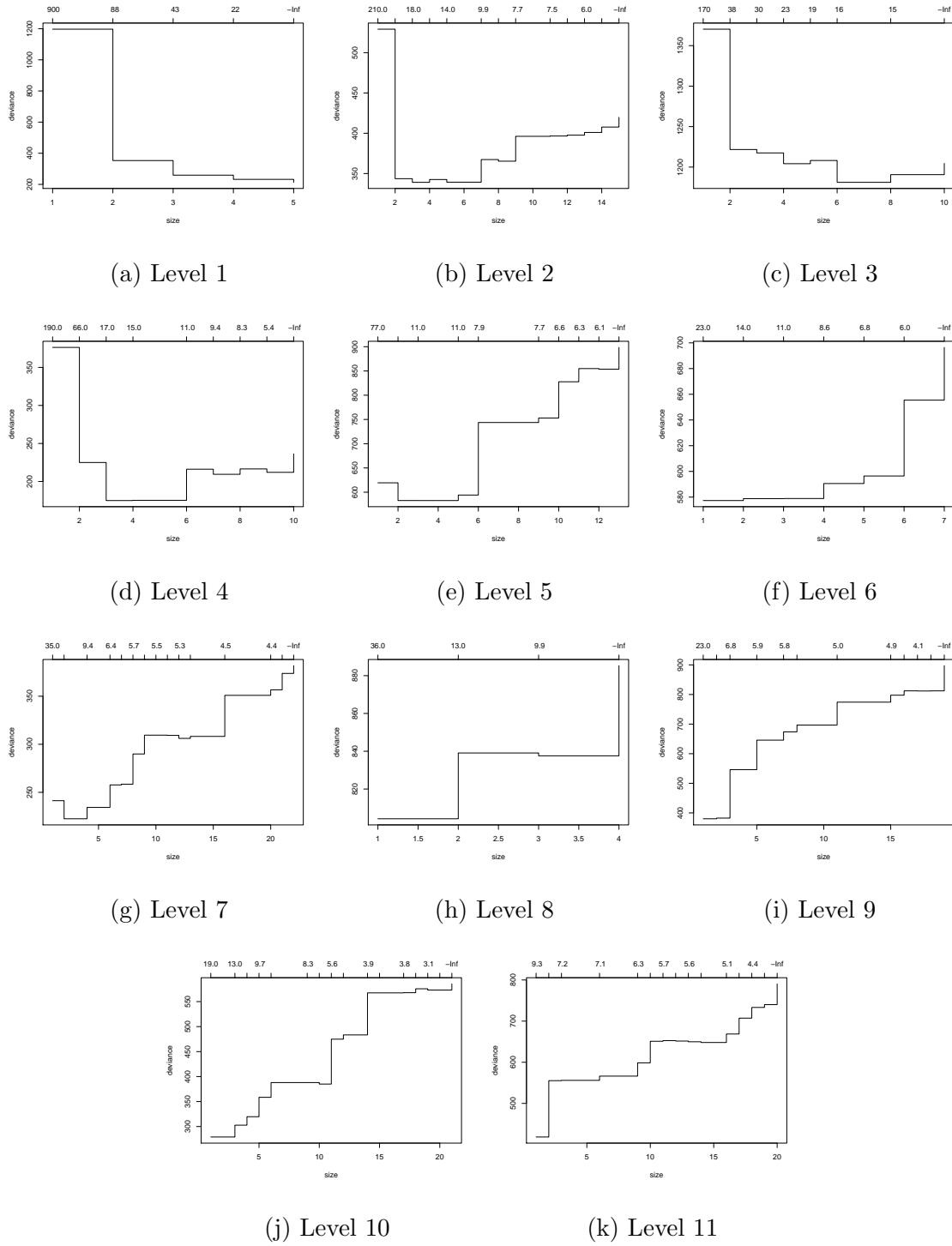


Fig. C.6: The Case 1 2009 cross-validation graphs generated by R tree package.

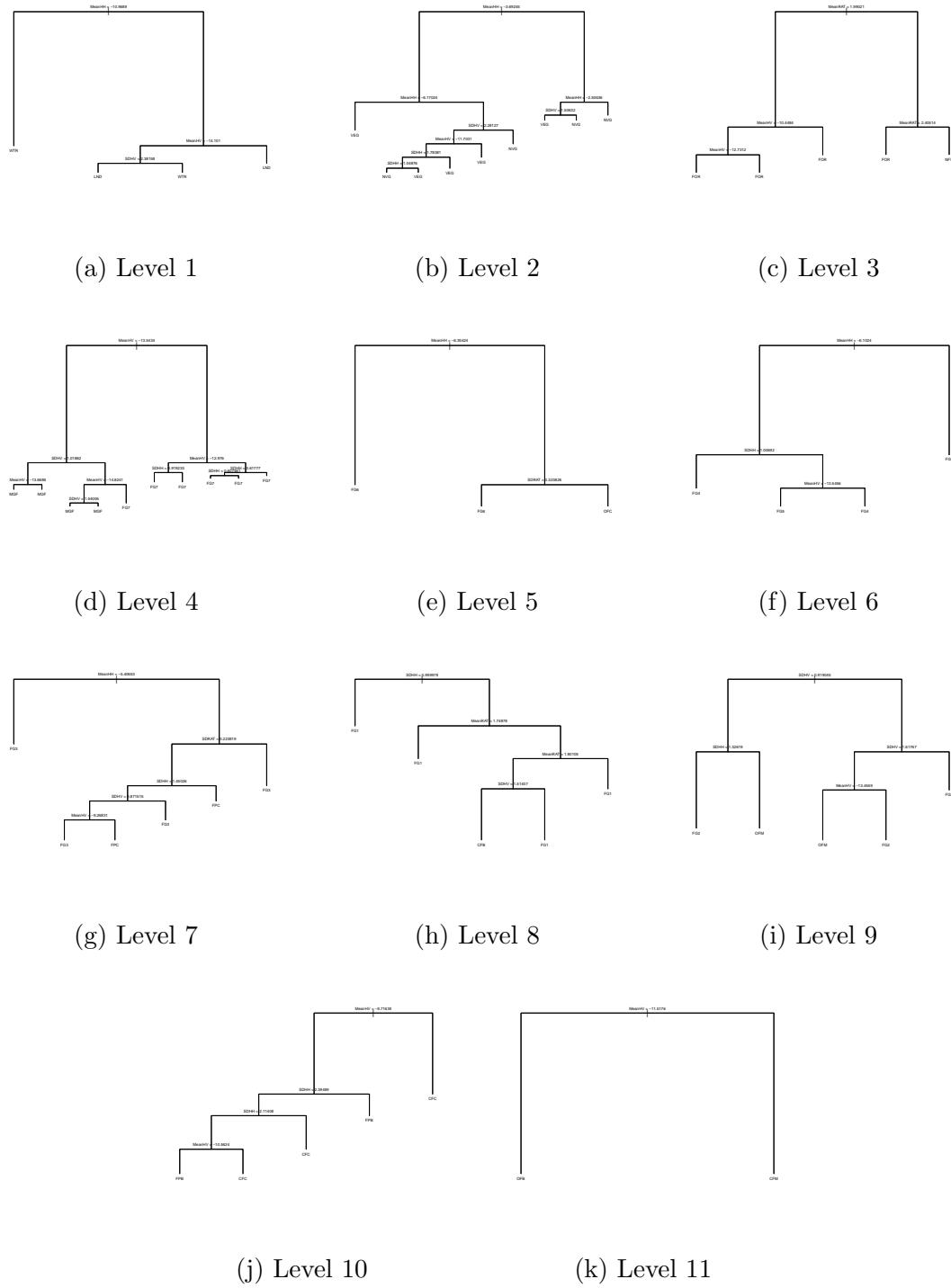


Fig. C.7: The Case 1 2010 classification trees generated by R tree package.

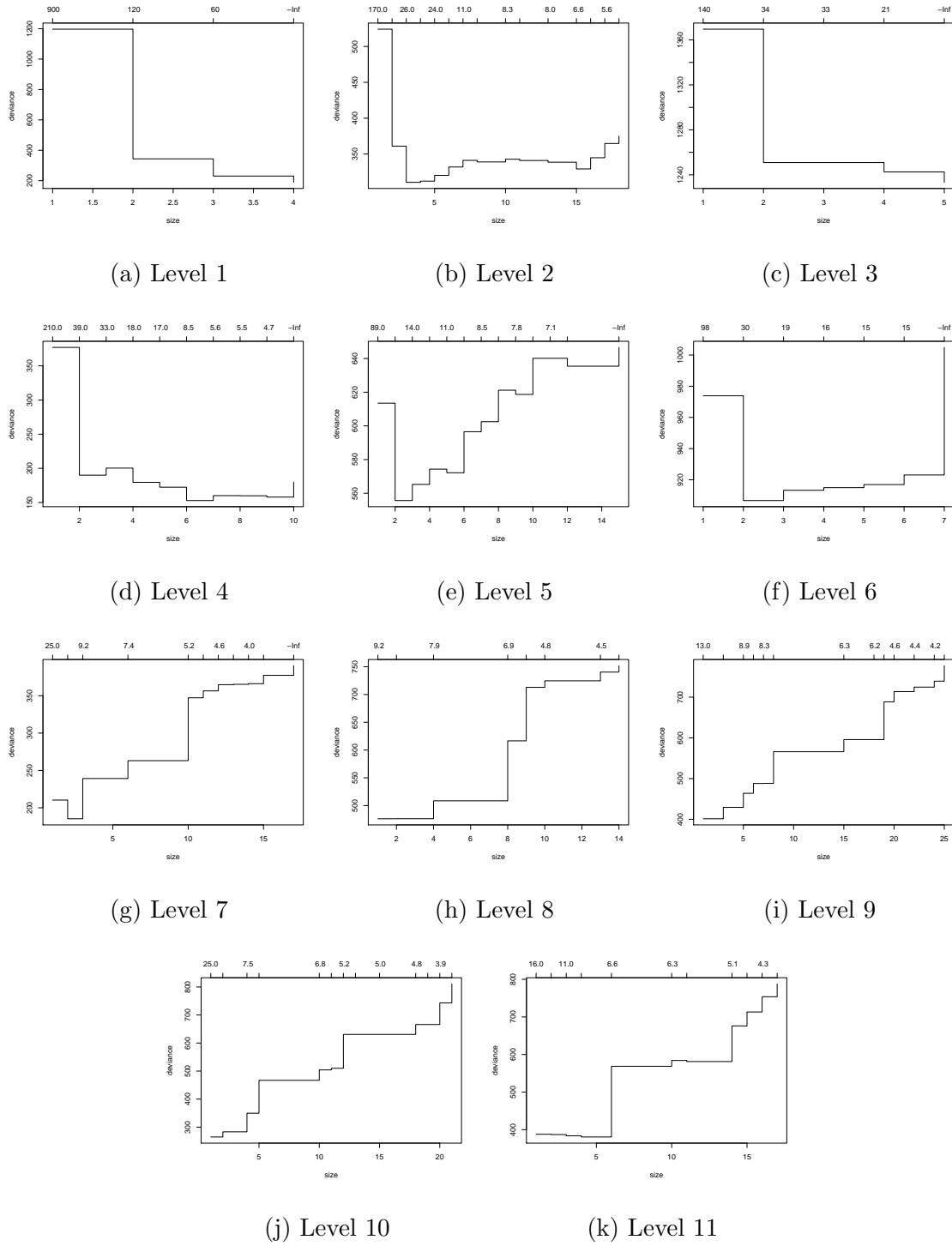


Fig. C.8: The Case 1 2010 cross-validation graphs generated by R tree package.

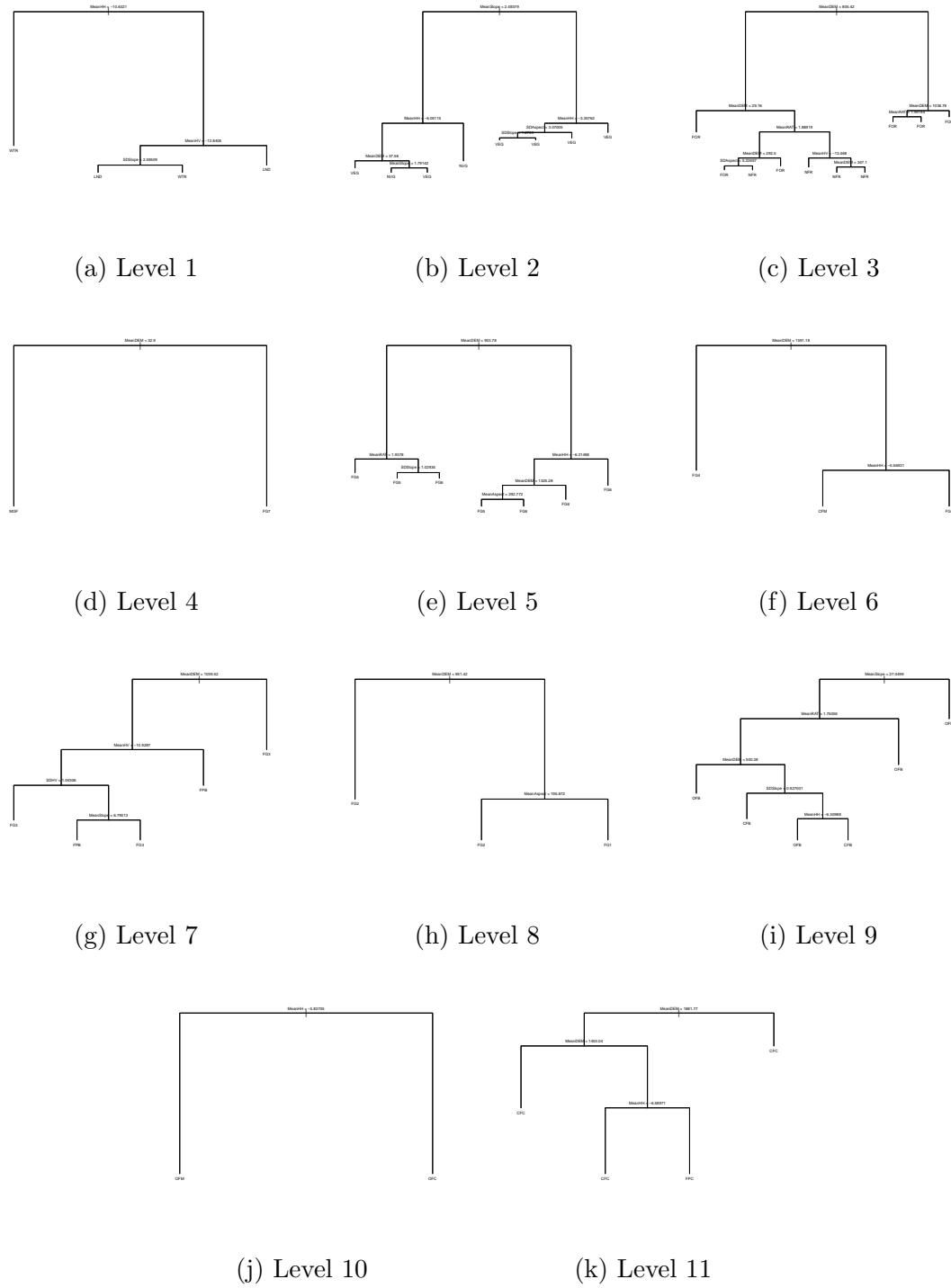


Fig. C.9: The Case 2 2007 classification trees generated by R tree package.

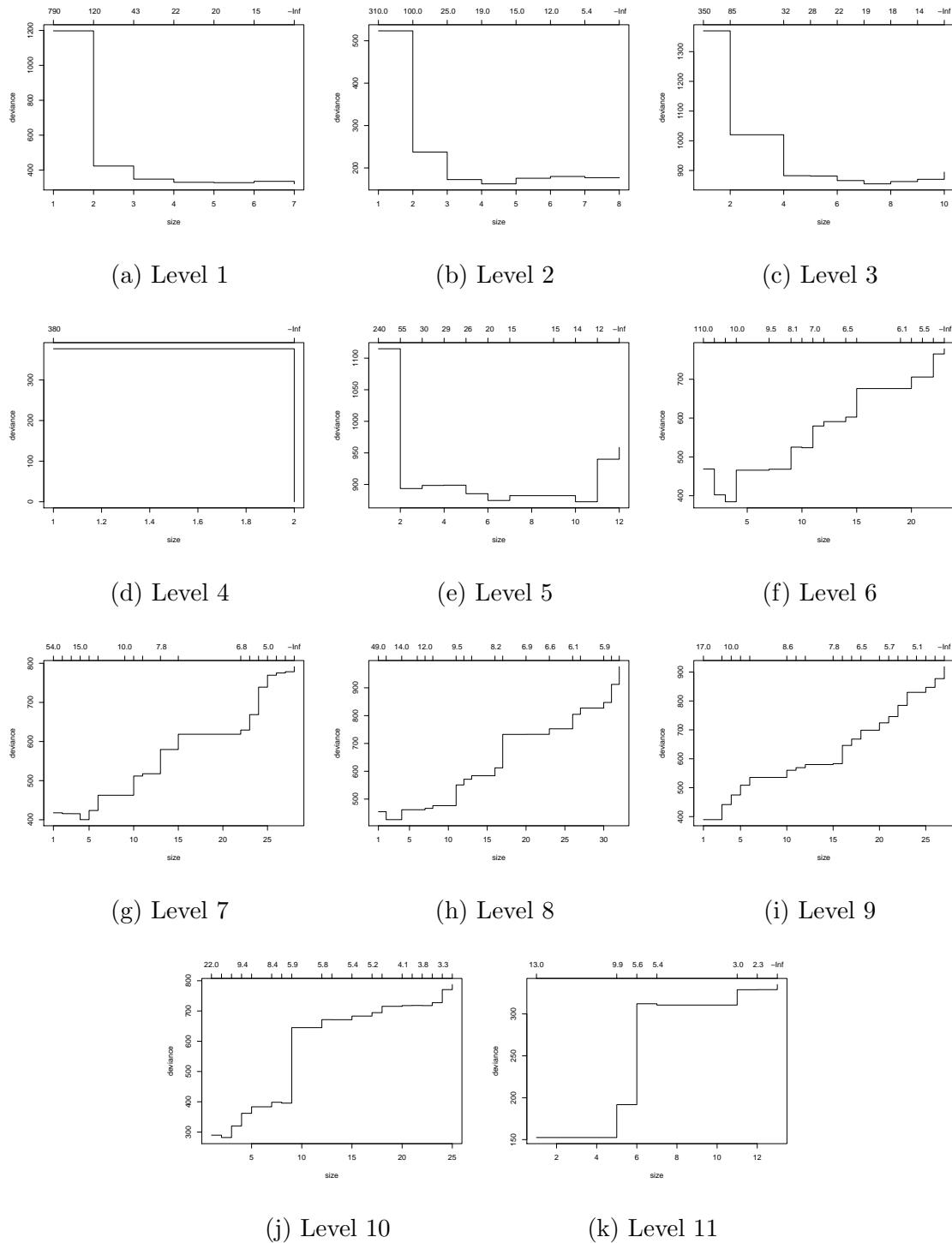


Fig. C.10: The Case 2 2007 cross-validation graphs generated by R tree package.

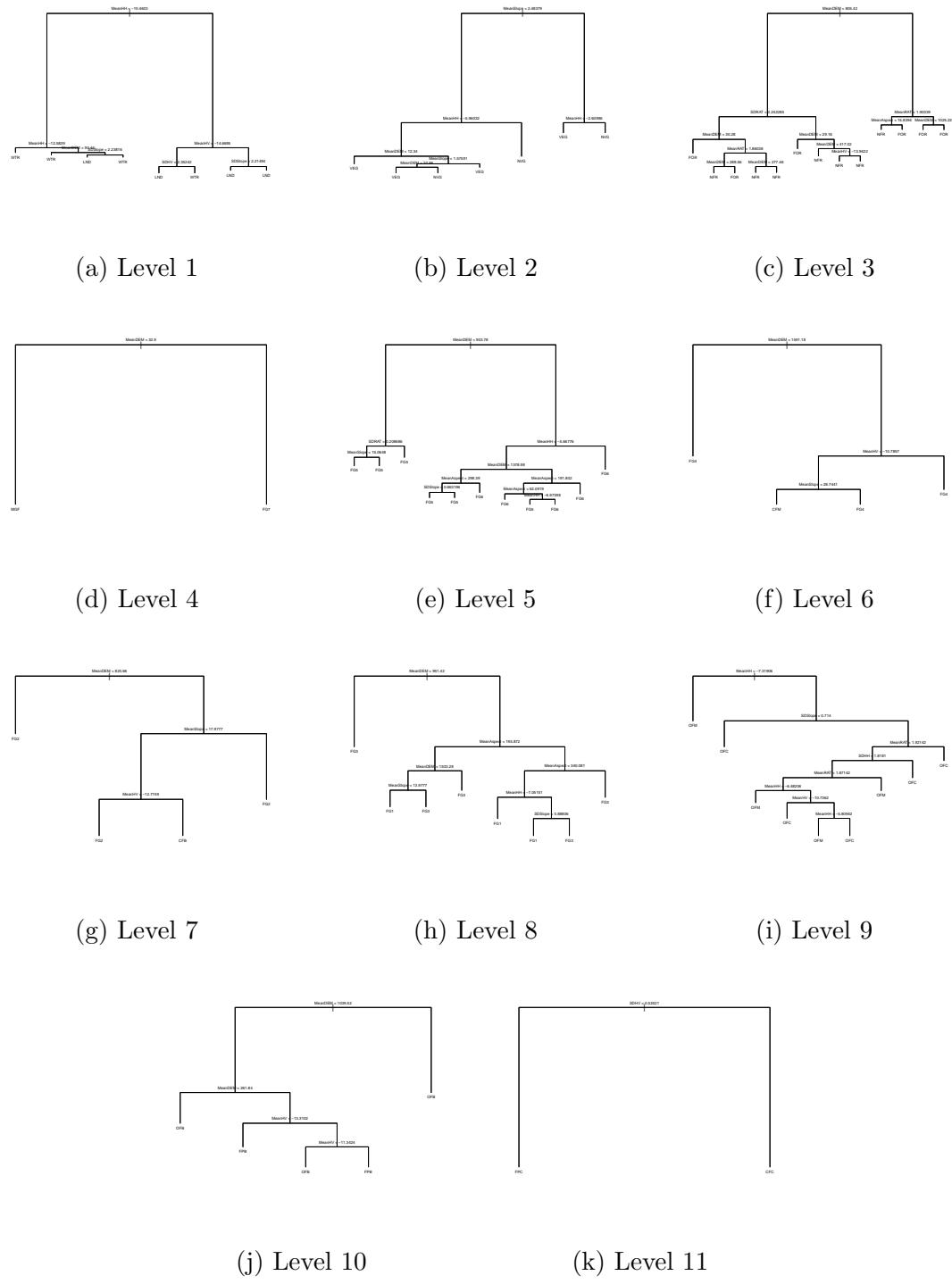


Fig. C.11: The Case 2 2008 classification trees generated by R tree package.

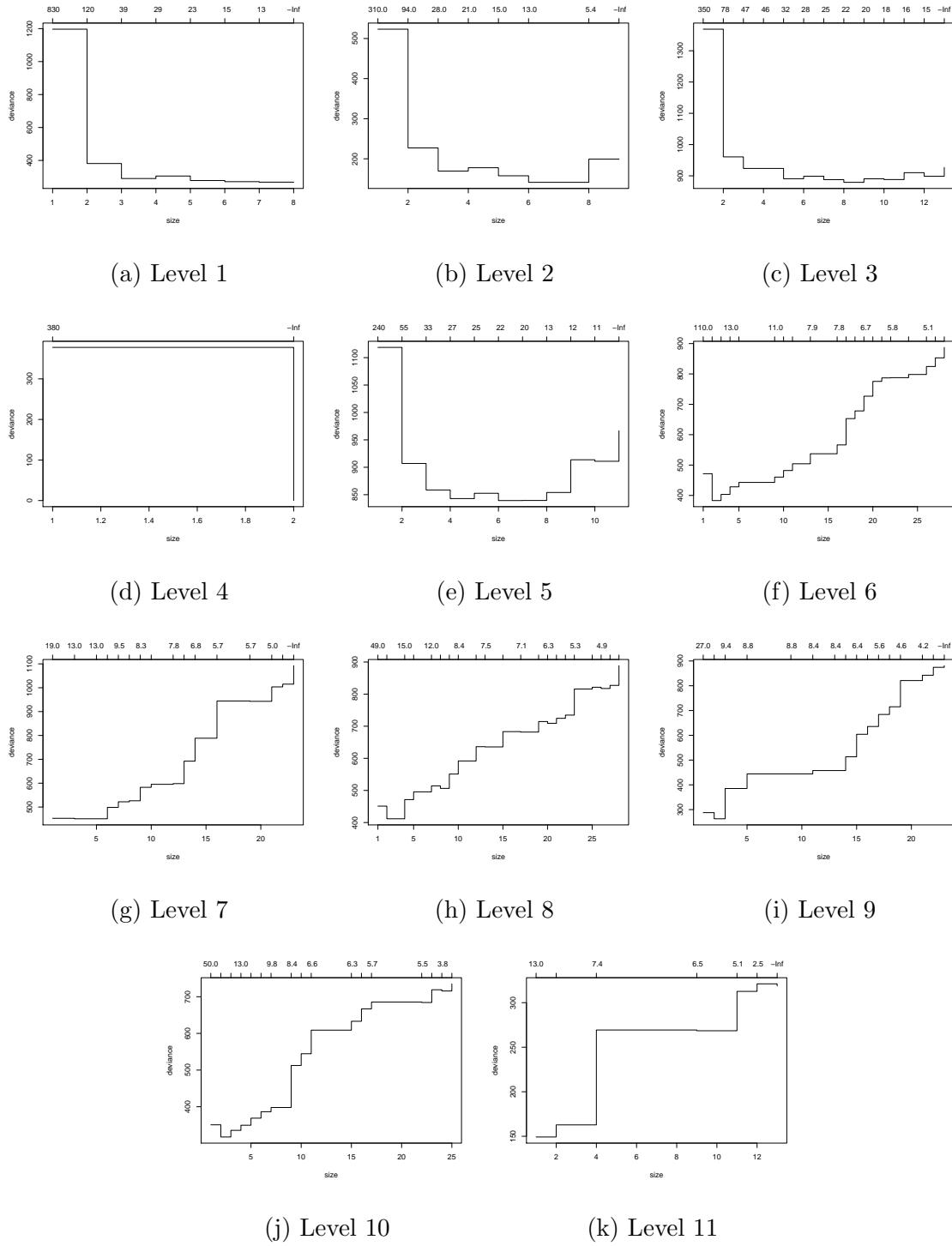


Fig. C.12: The Case 2 2008 cross-validation graphs generated by R tree package.

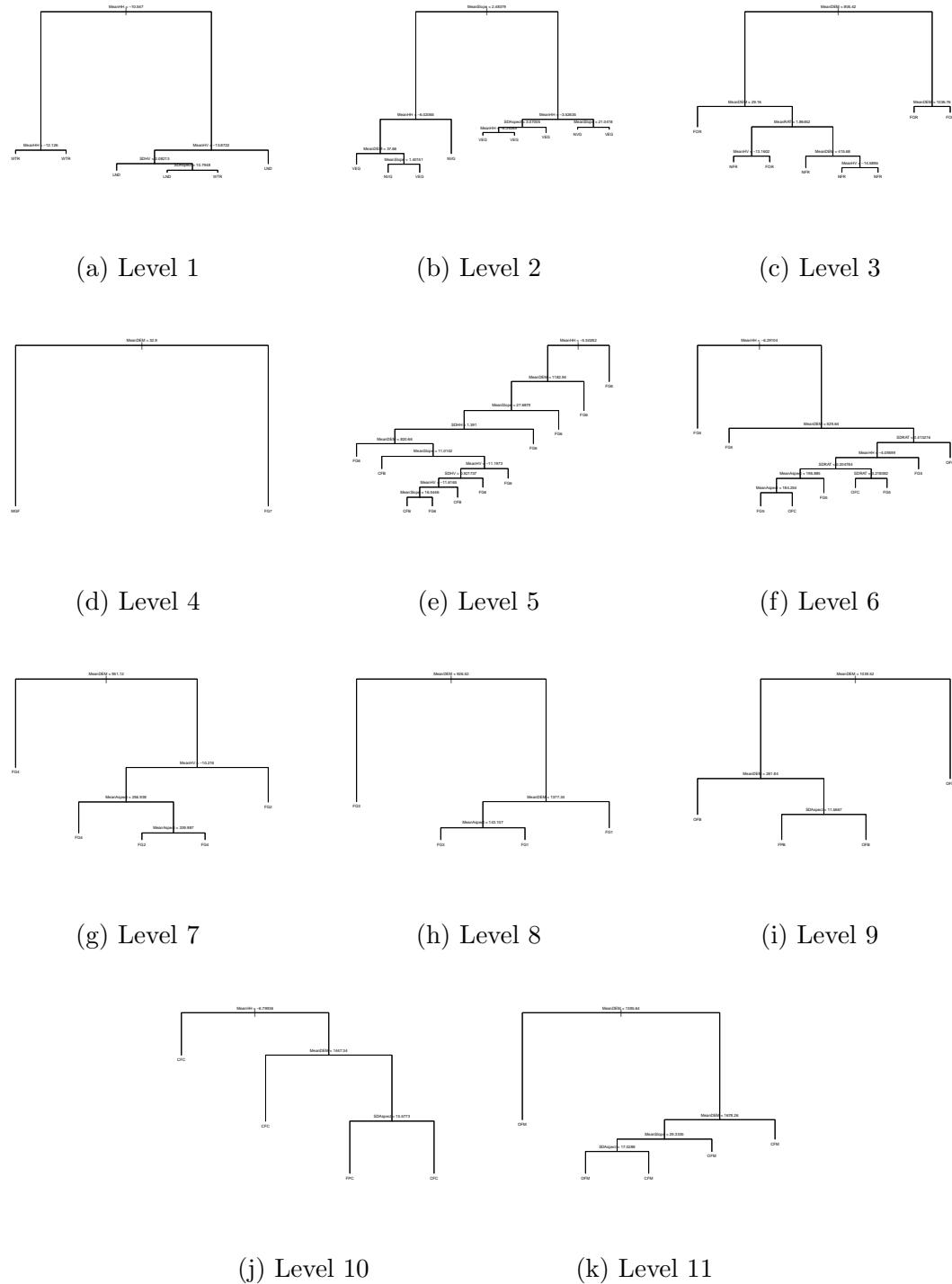


Fig. C.13: The Case 2 2009 classification trees generated by R tree package.

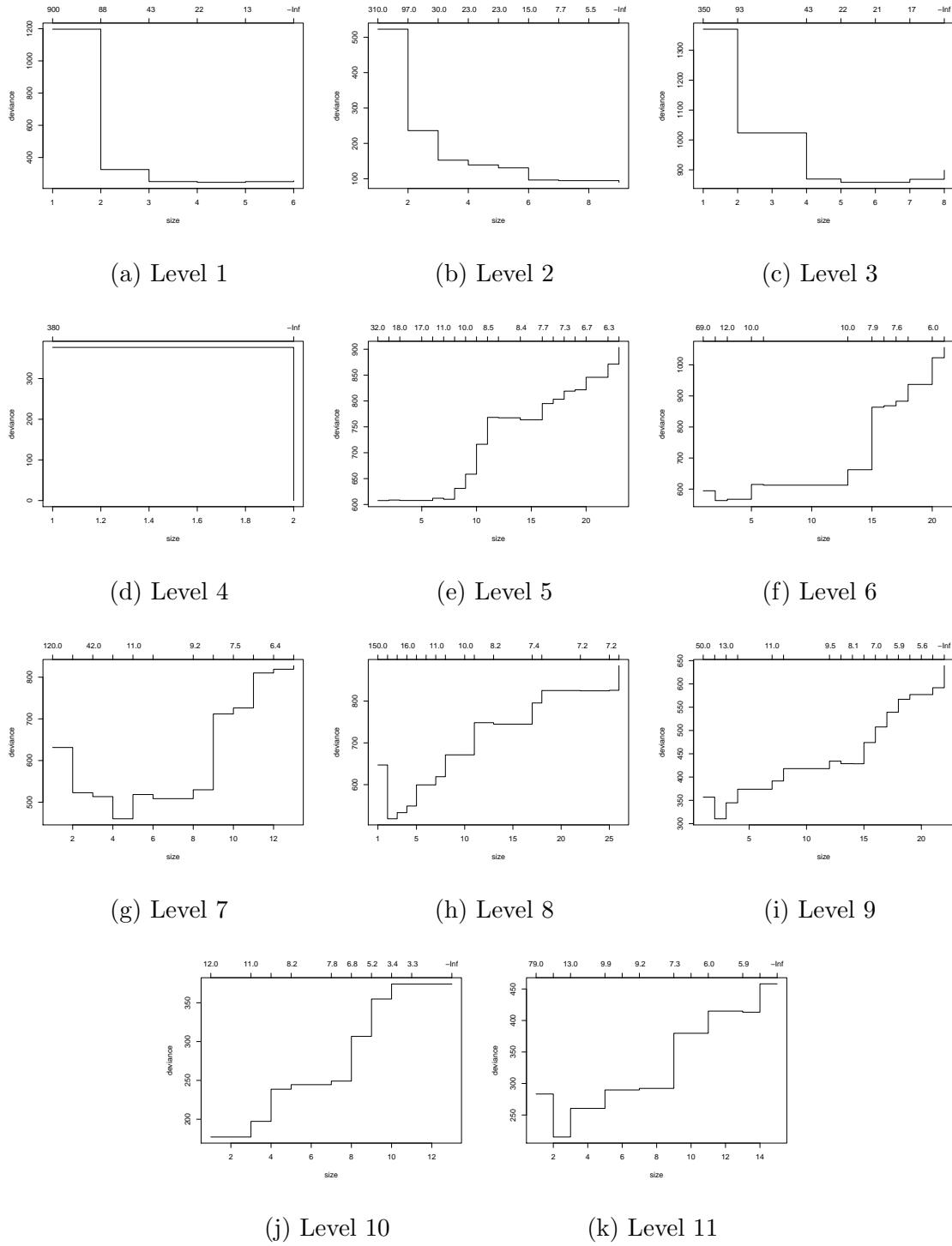


Fig. C.14: The Case 2 2009 cross-validation graphs generated by R tree package.

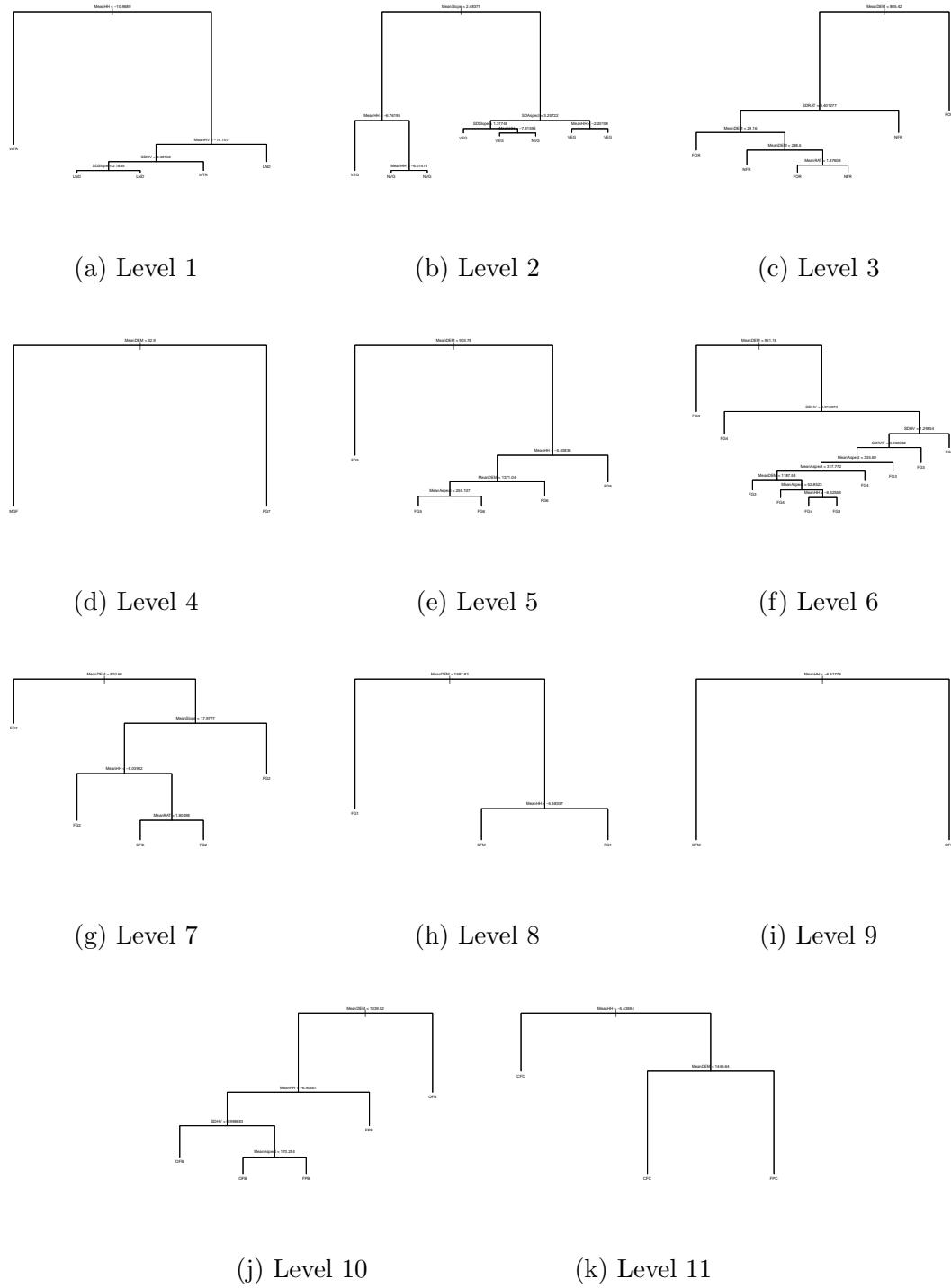


Fig. C.15: The Case 2 2010 classification trees generated by R tree package.

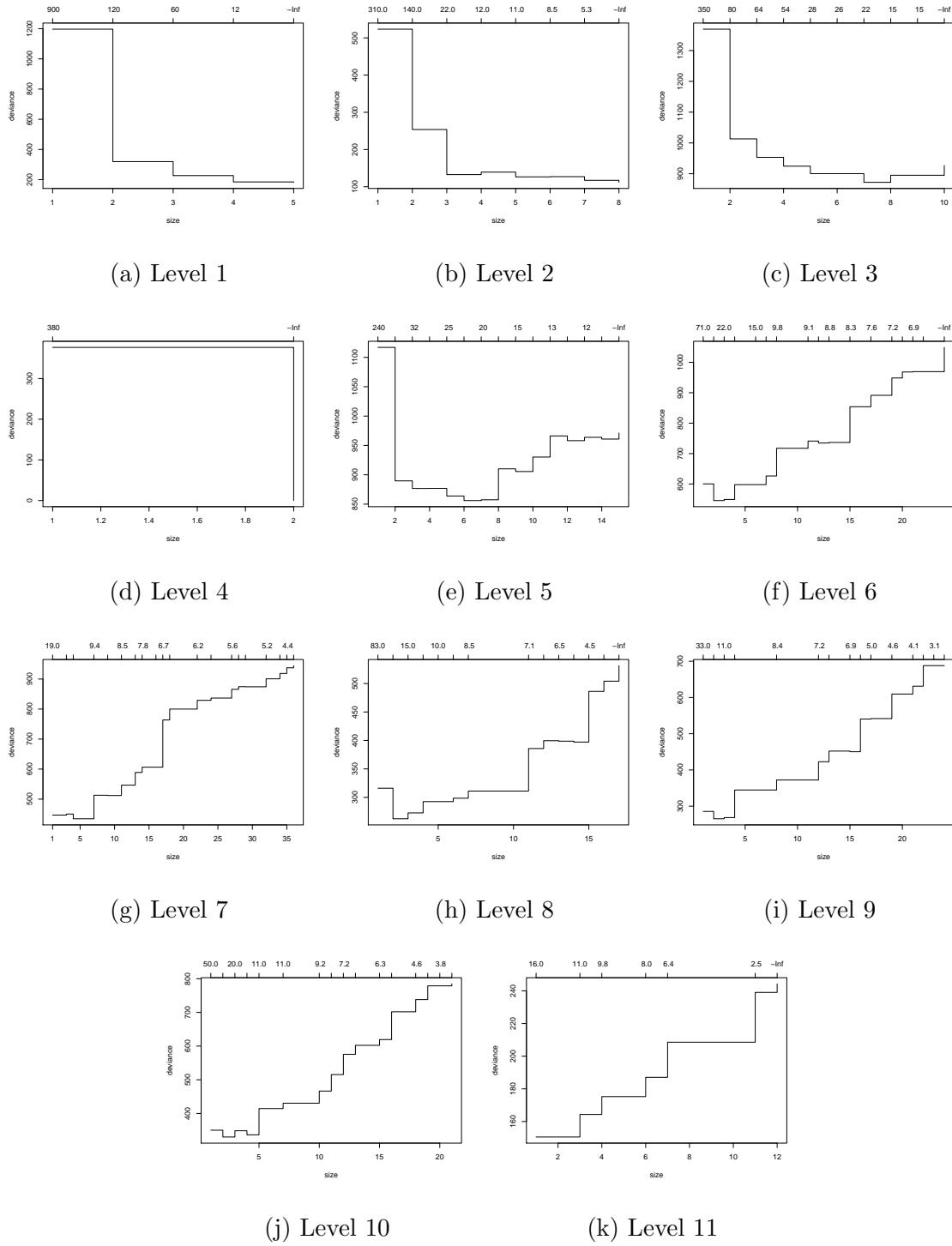


Fig. C.16: The Case 2 2010 cross-validation graphs generated by R tree package.

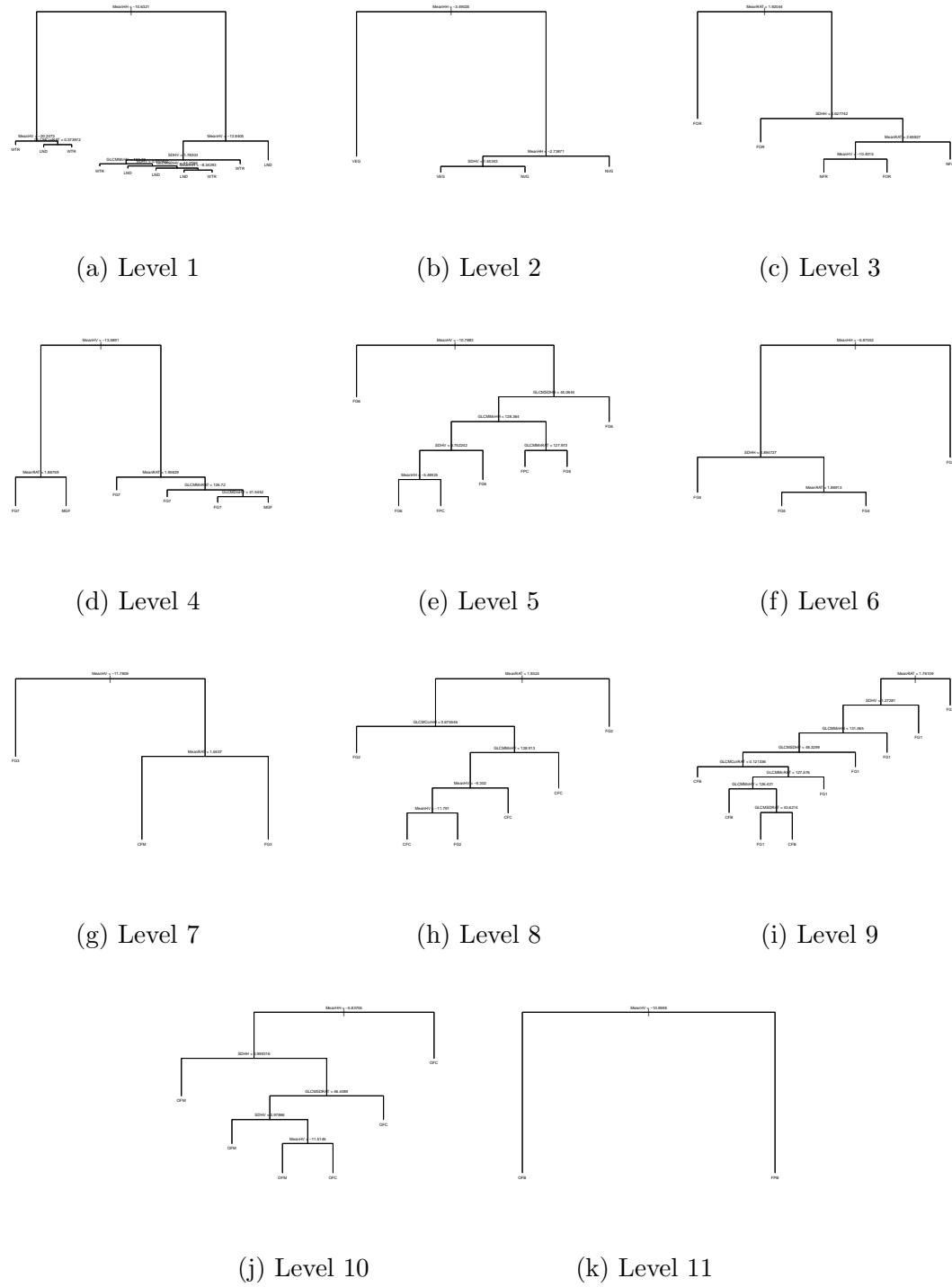


Fig. C.17: The Case 3 2007 classification trees generated by R tree package.

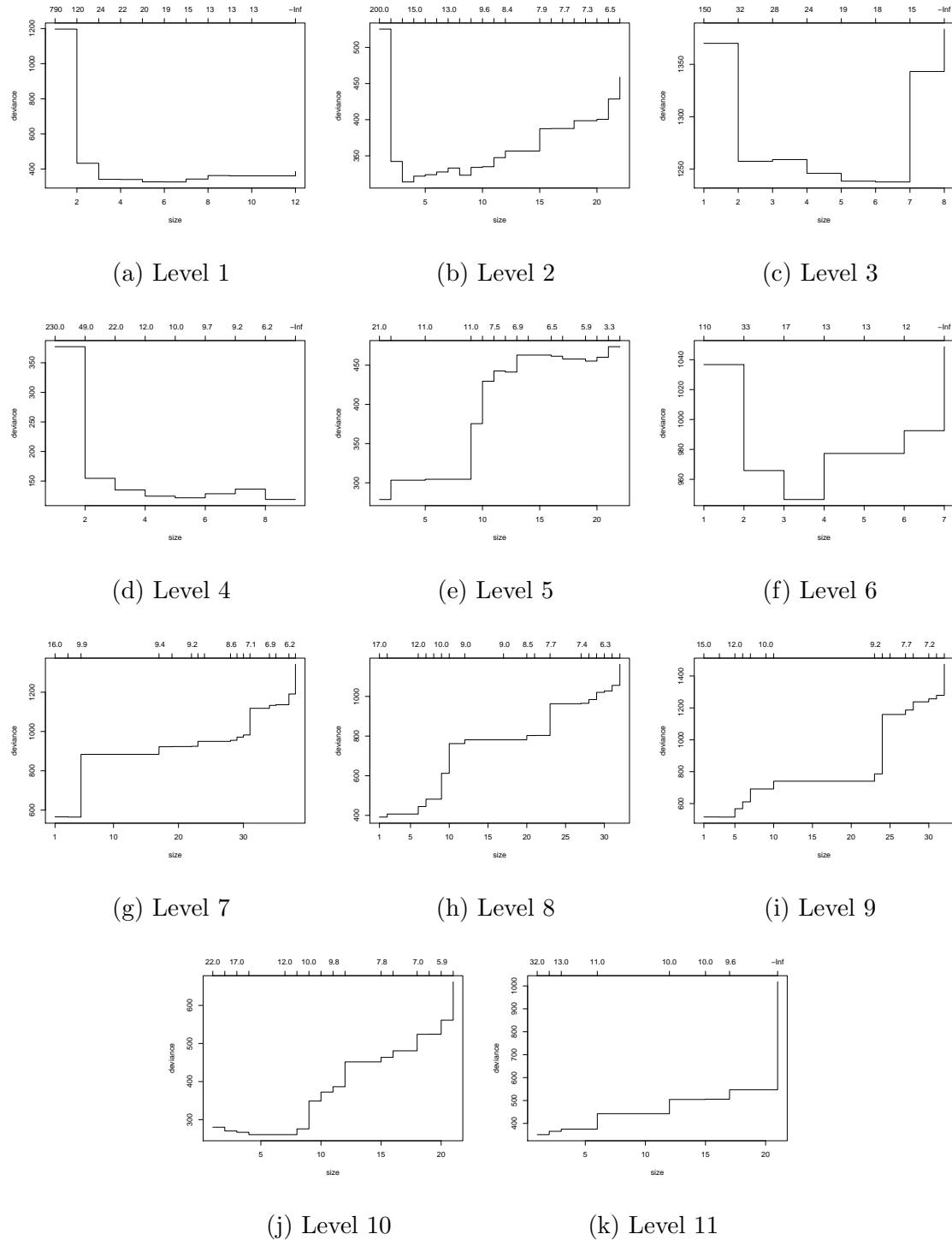


Fig. C.18: The Case 3 2007 cross-validation graphs generated by R tree package.

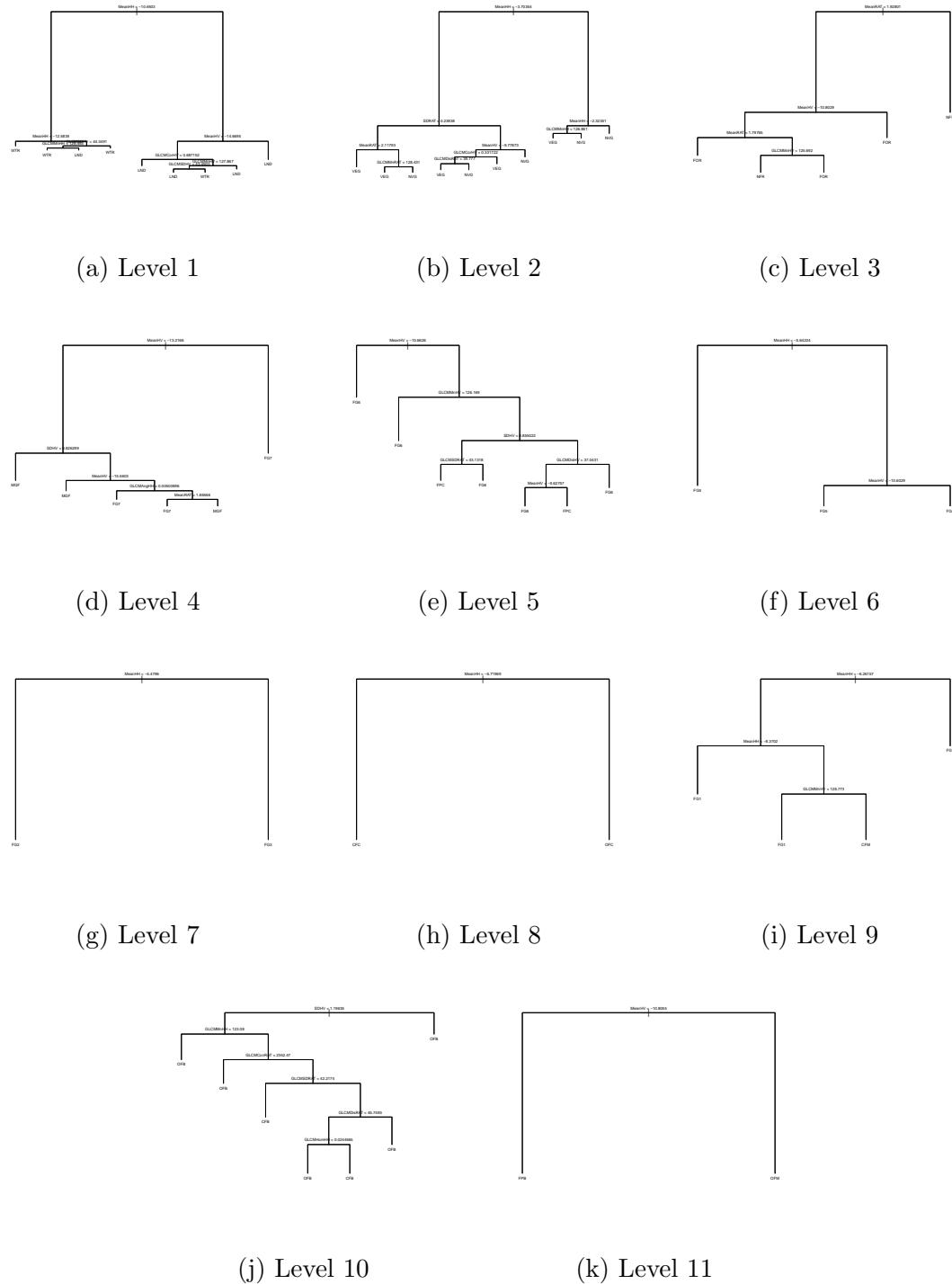


Fig. C.19: The Case 3 2008 classification trees generated by R tree package.

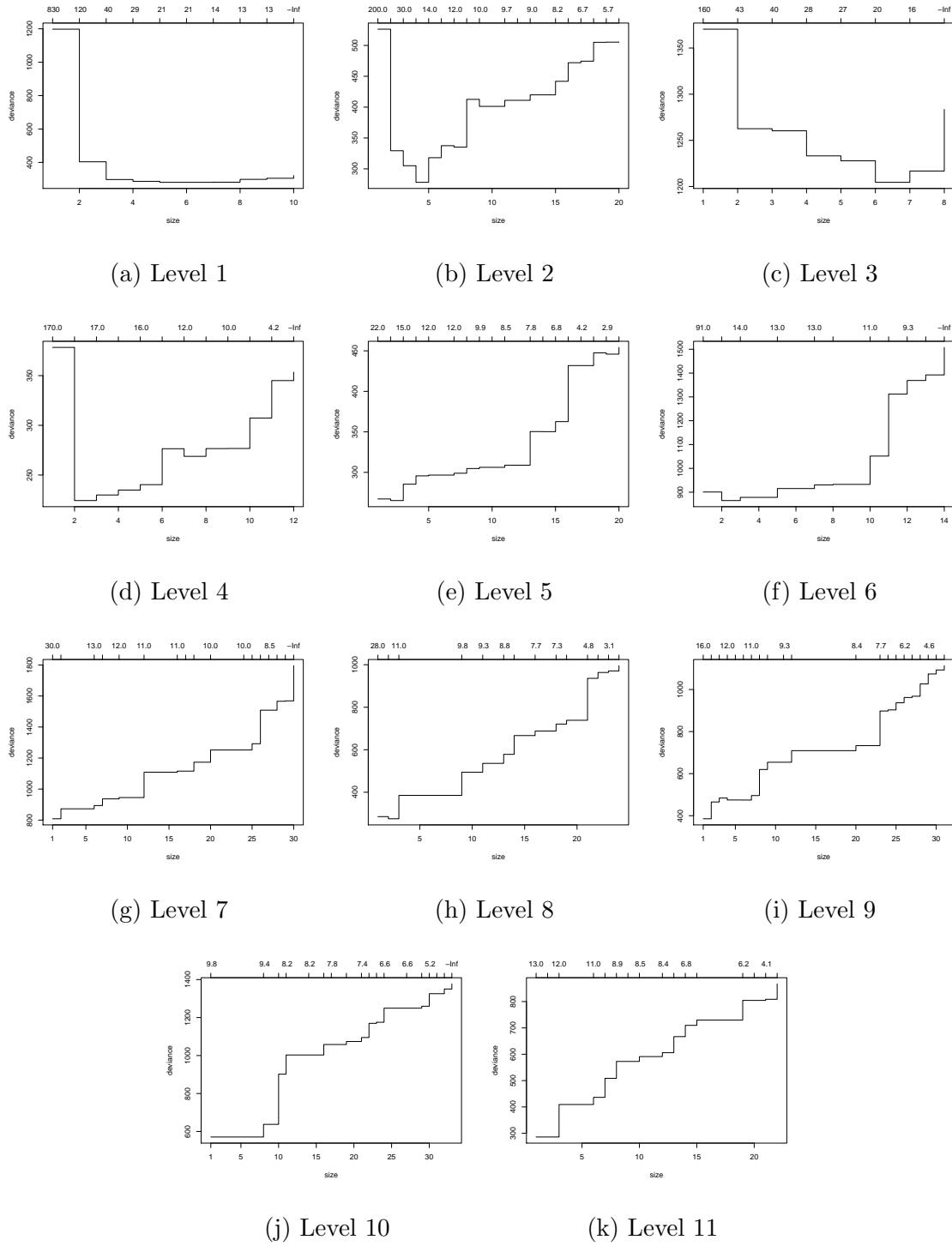


Fig. C.20: The Case 3 2008 cross-validation graphs generated by R tree package.

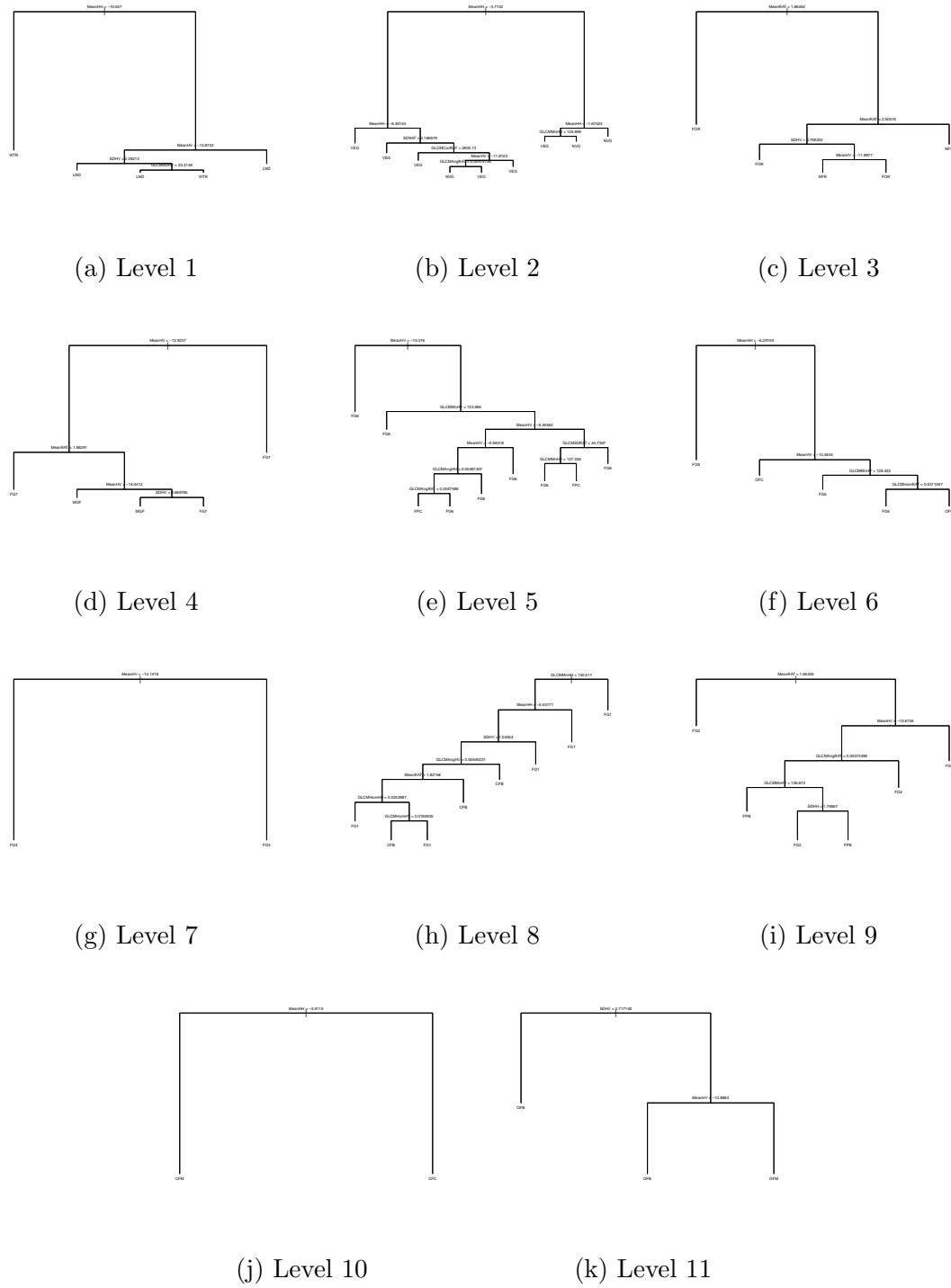


Fig. C.21: The Case 3 2009 classification trees generated by R tree package.

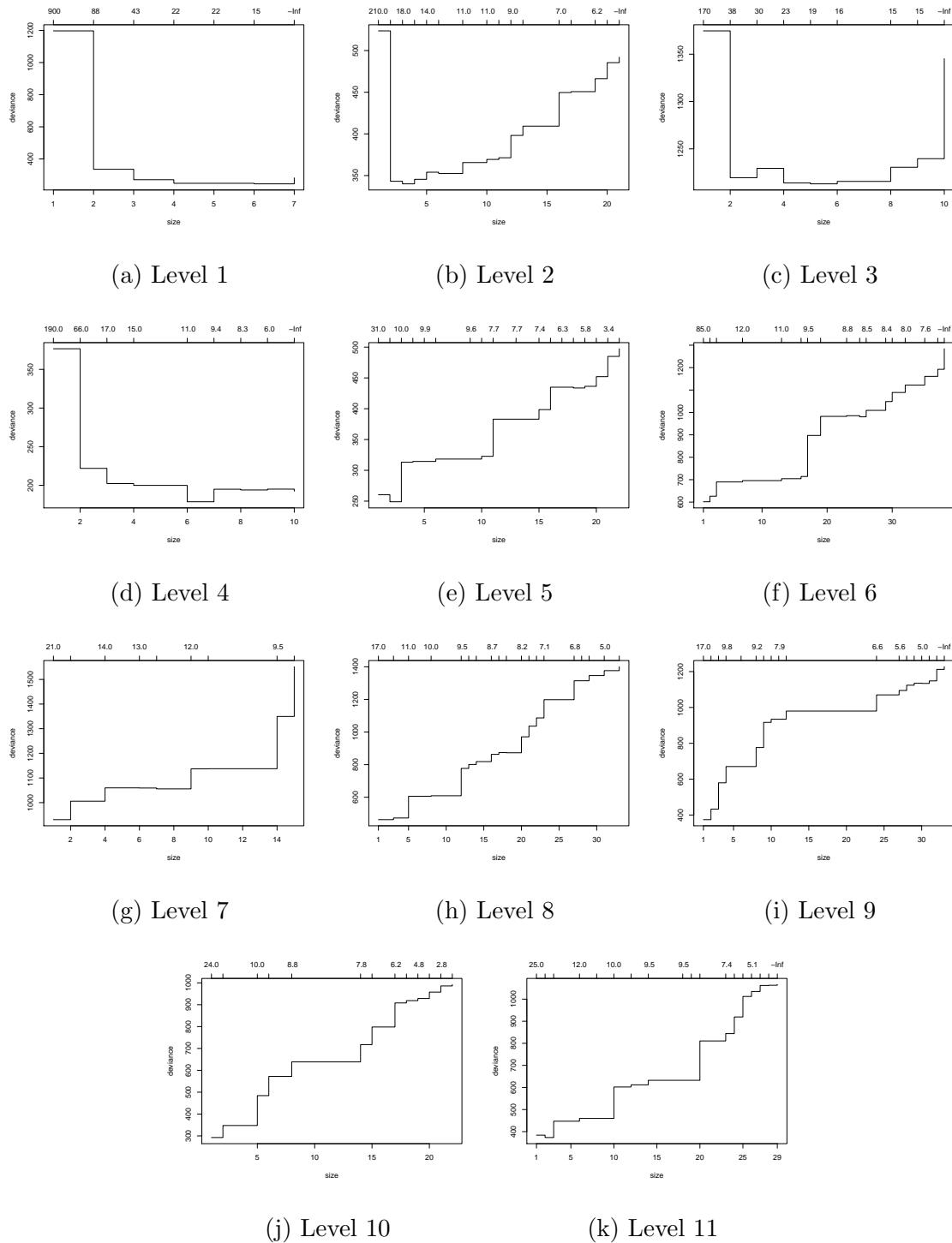


Fig. C.22: The Case 3 2009 cross-validation graphs generated by R tree package.

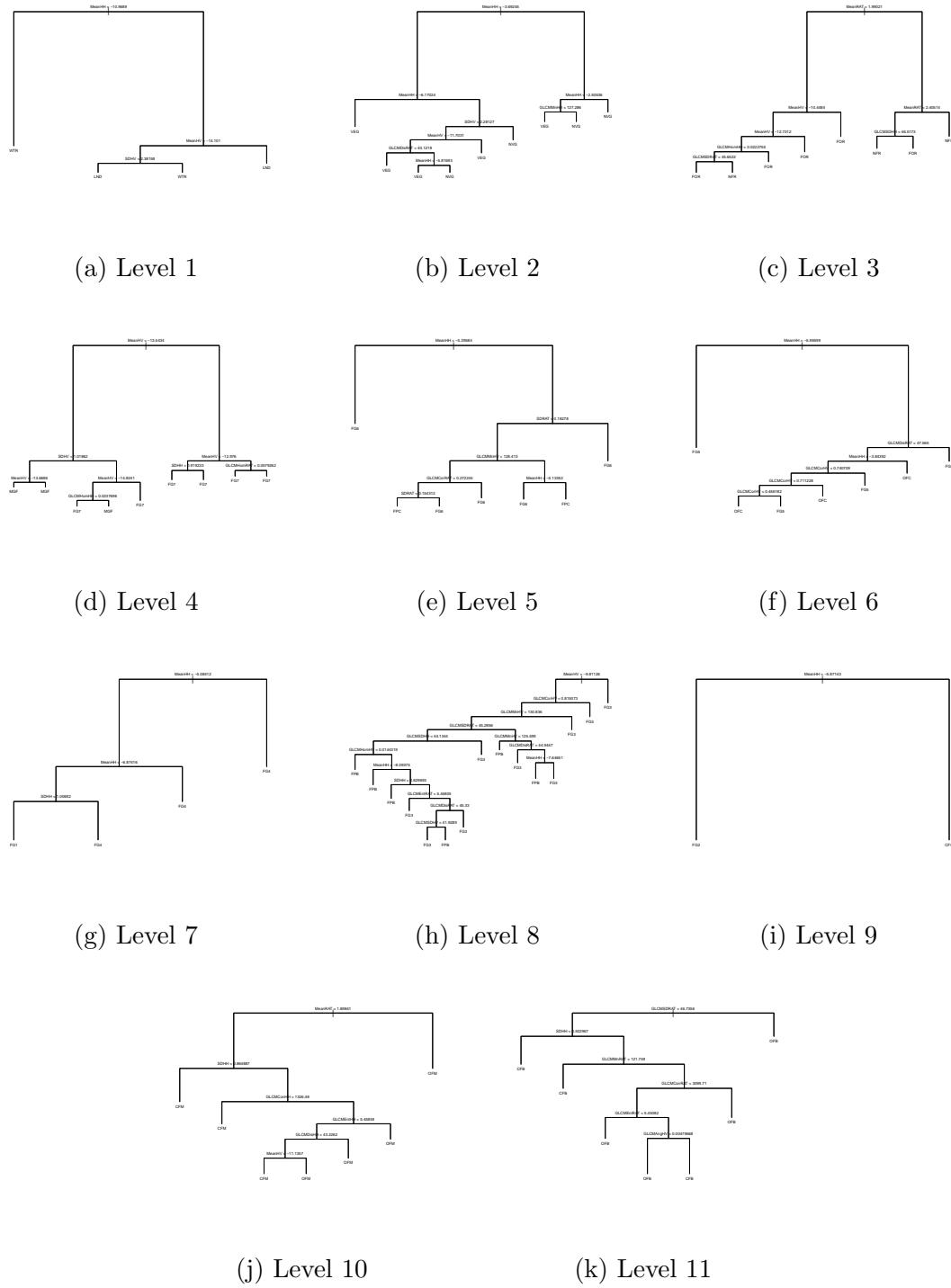


Fig. C.23: The Case 3 2010 classification trees generated by R tree package.

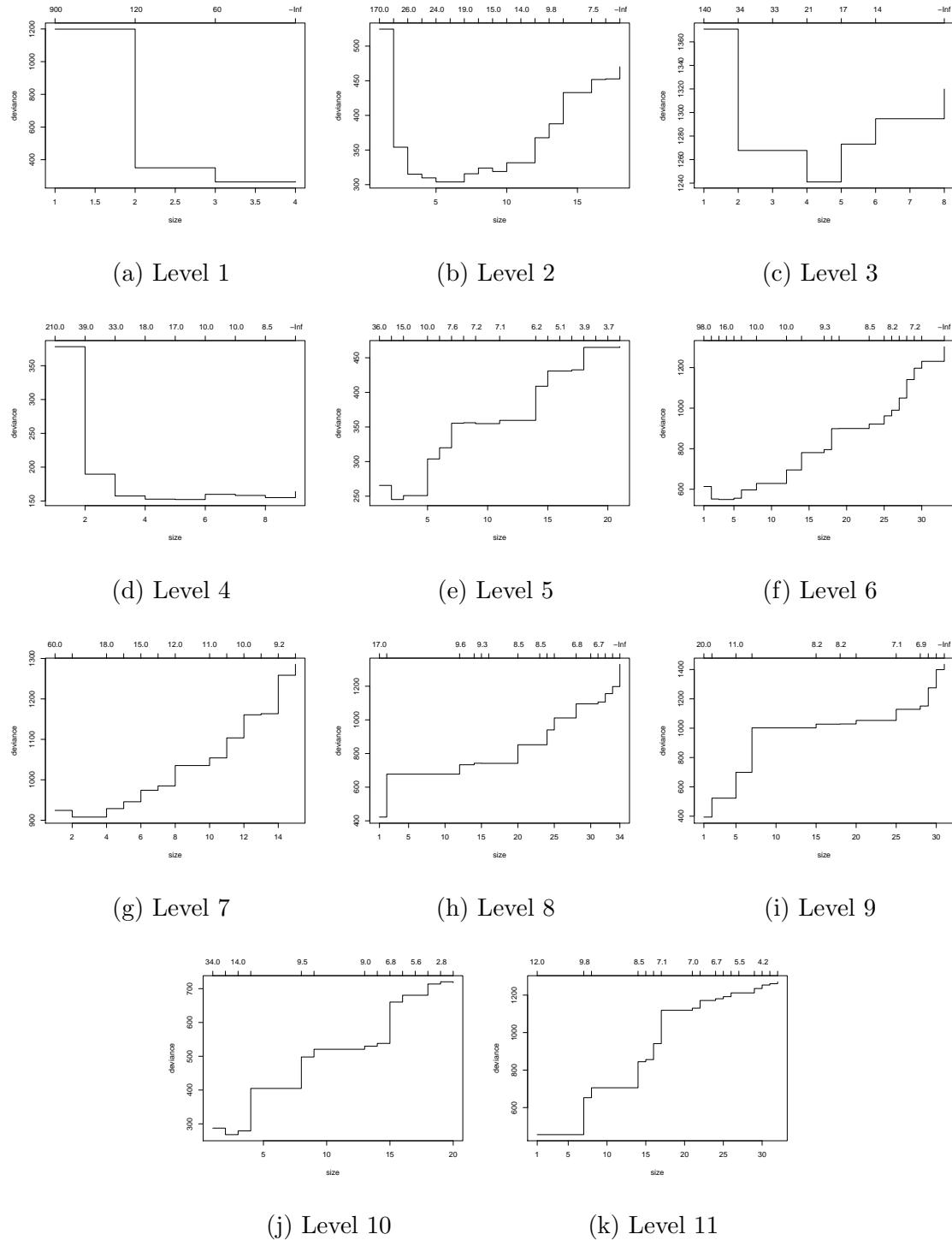


Fig. C.24: The Case 3 2010 cross-validation graphs generated by R tree package.

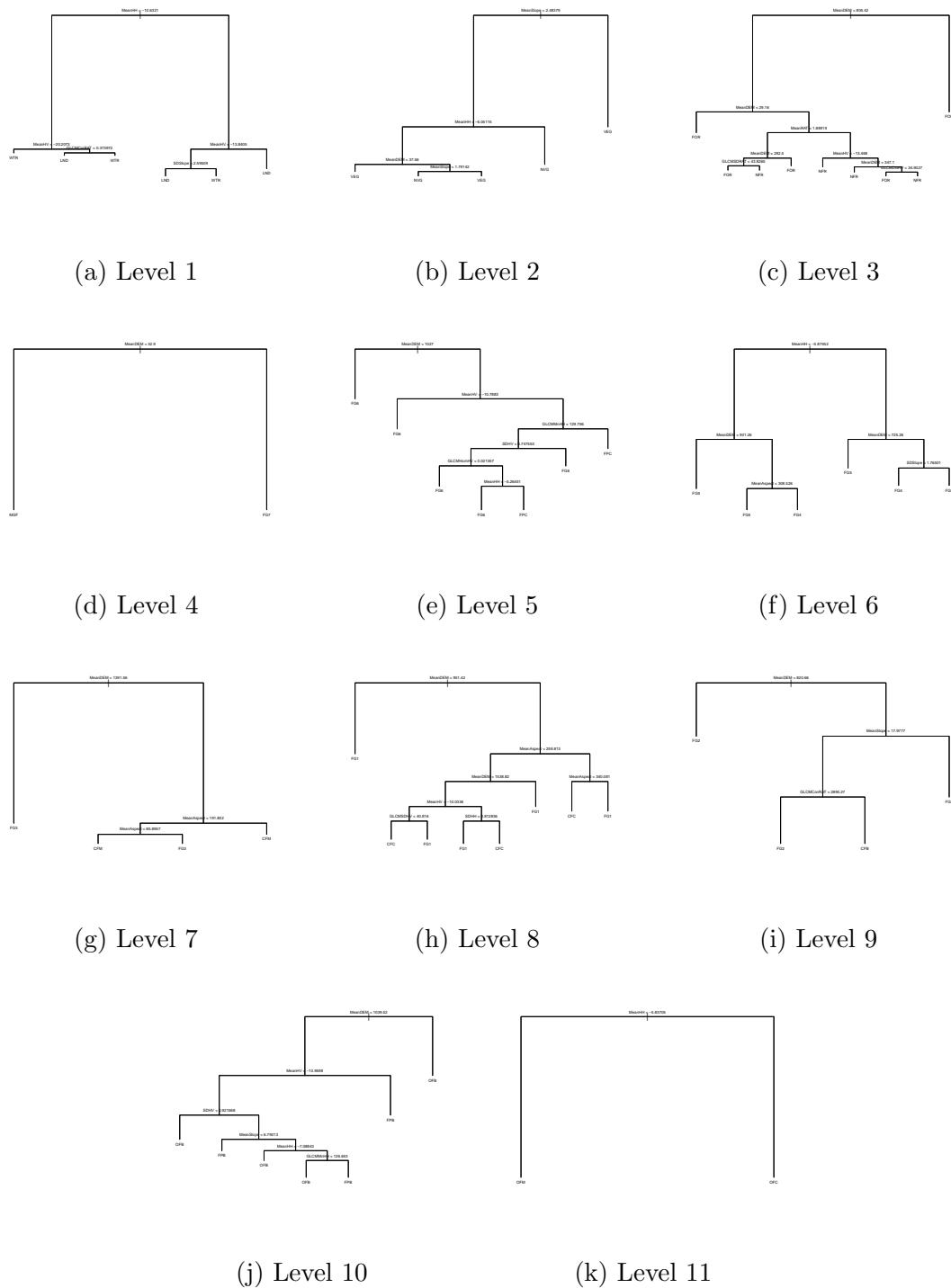


Fig. C.25: The Case 4 2007 classification trees generated by R tree package.

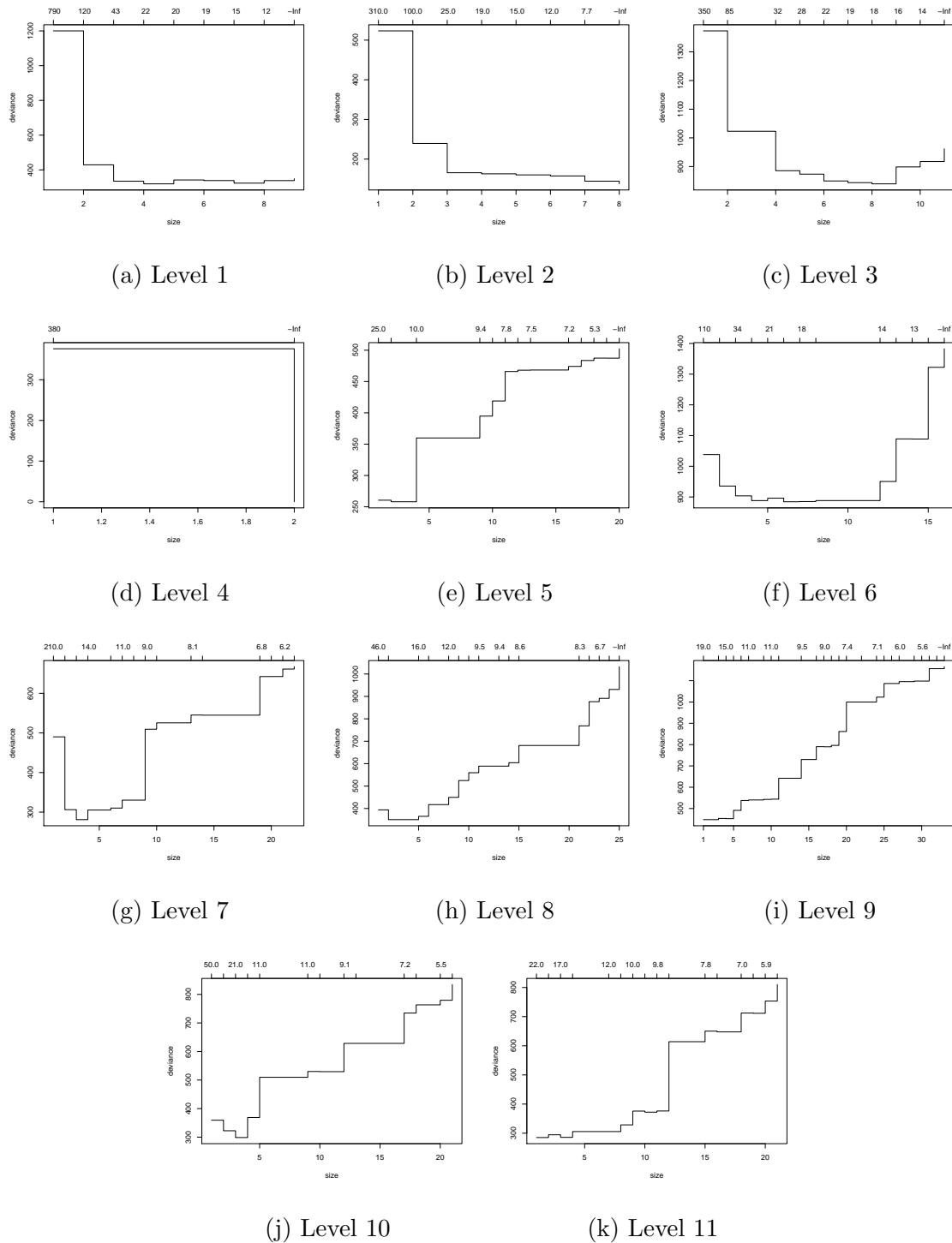


Fig. C.26: The Case 4 2007 cross-validation graphs generated by R tree package.

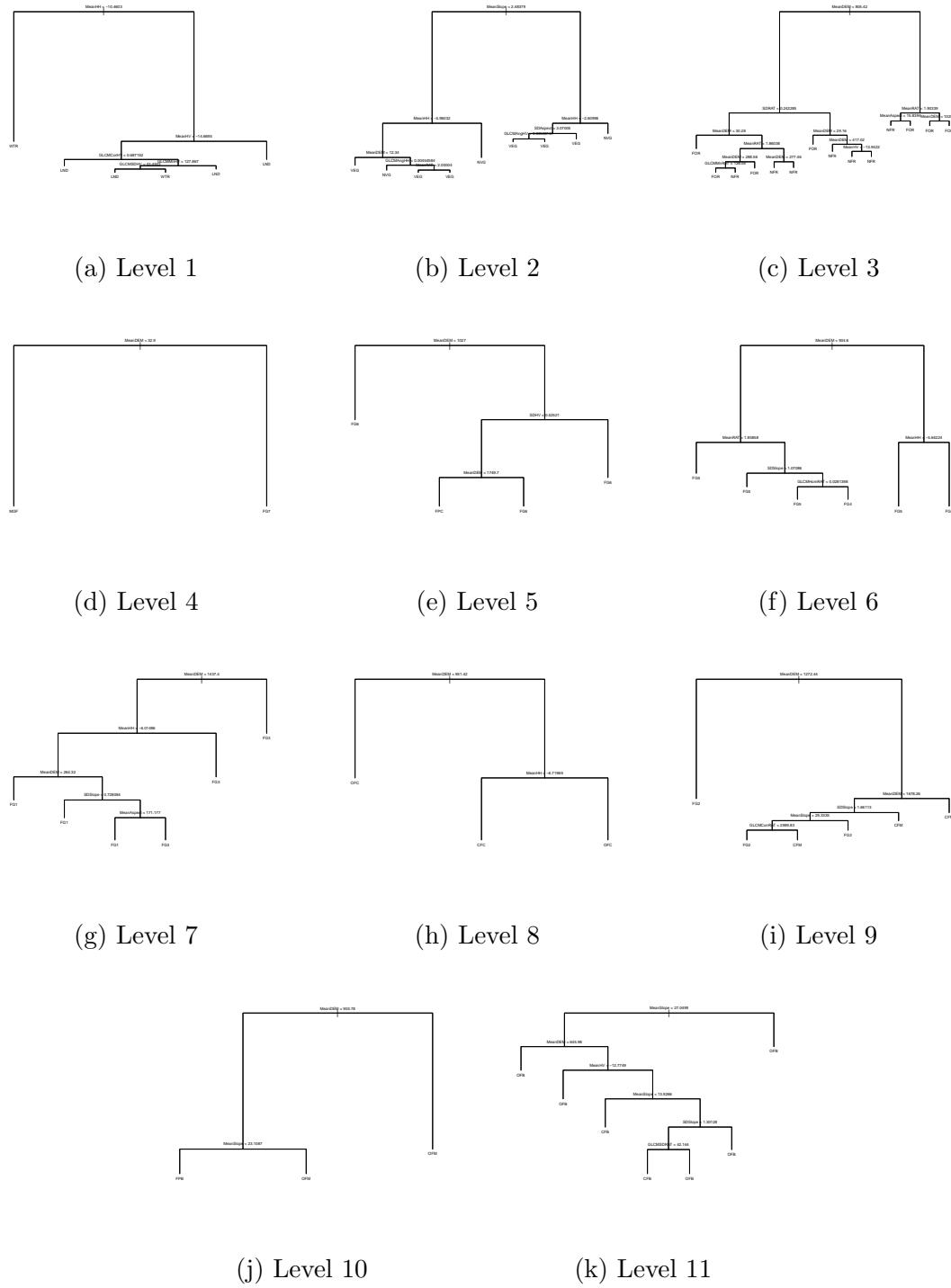


Fig. C.27: The Case 4 2008 classification trees generated by R tree package.

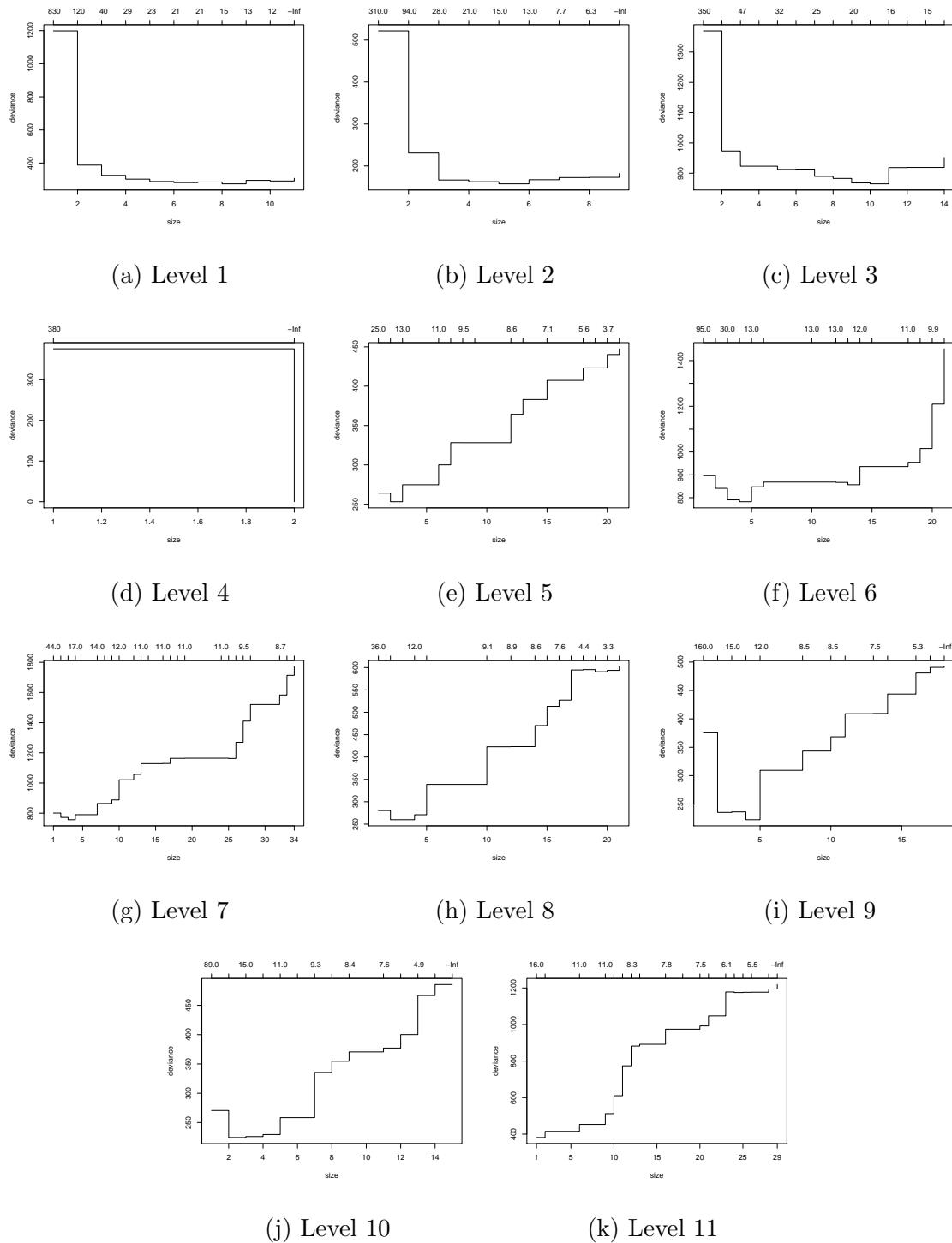


Fig. C.28: The Case 4 2008 cross-validation graphs generated by R tree package.

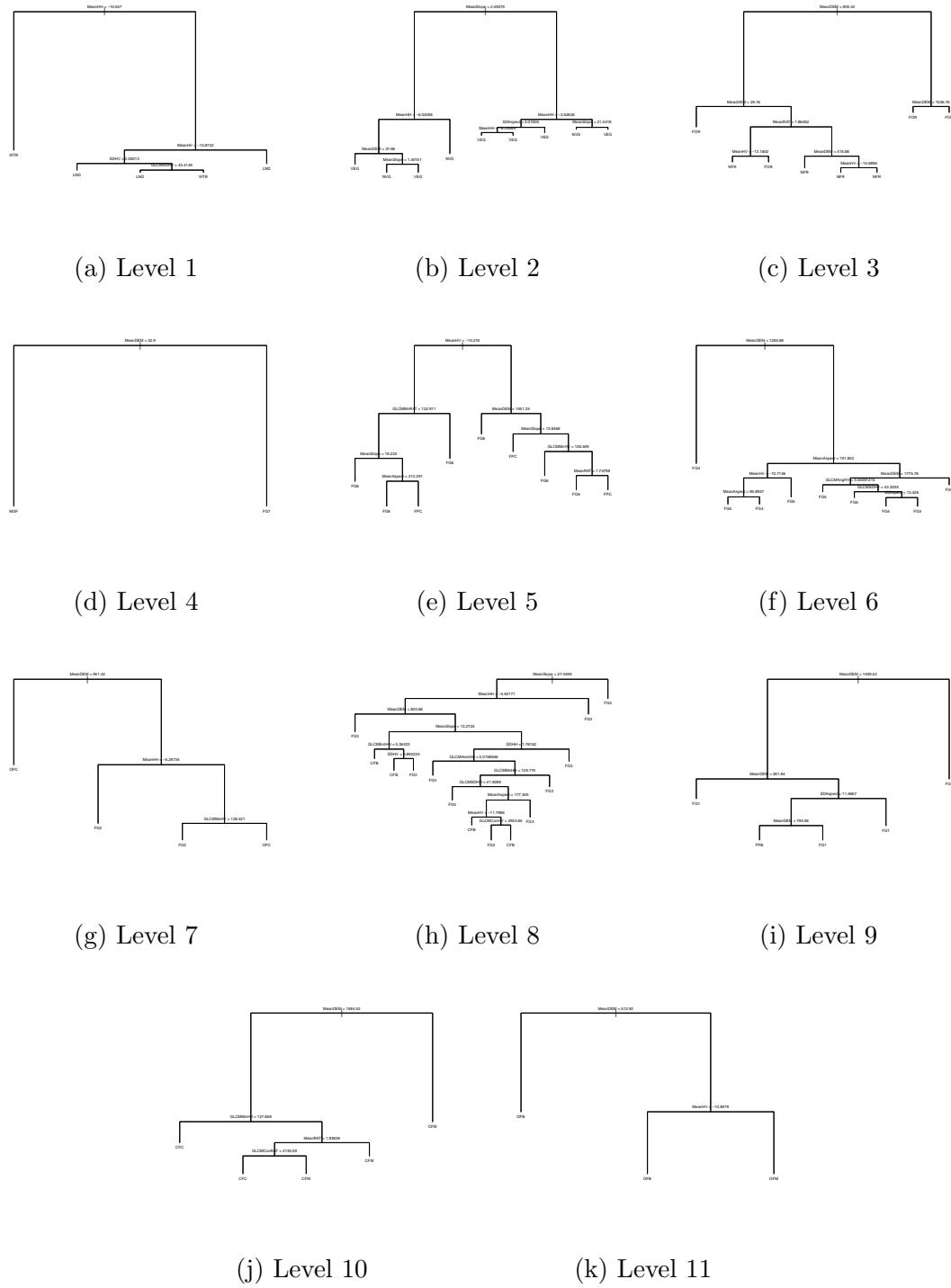


Fig. C.29: The Case 4 2009 classification trees generated by R tree package.

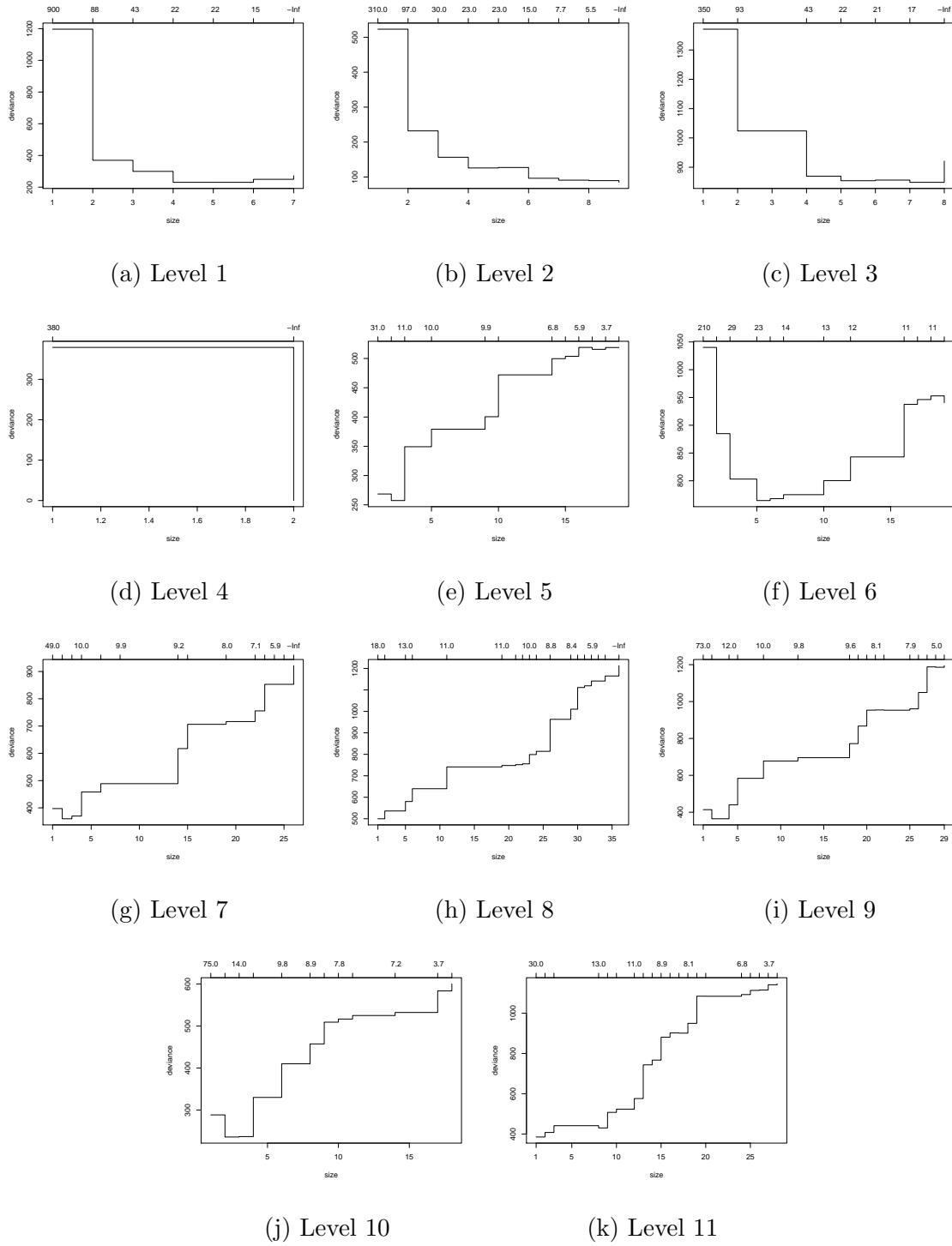


Fig. C.30: The Case 4 2009 cross-validation graphs generated by R tree package.

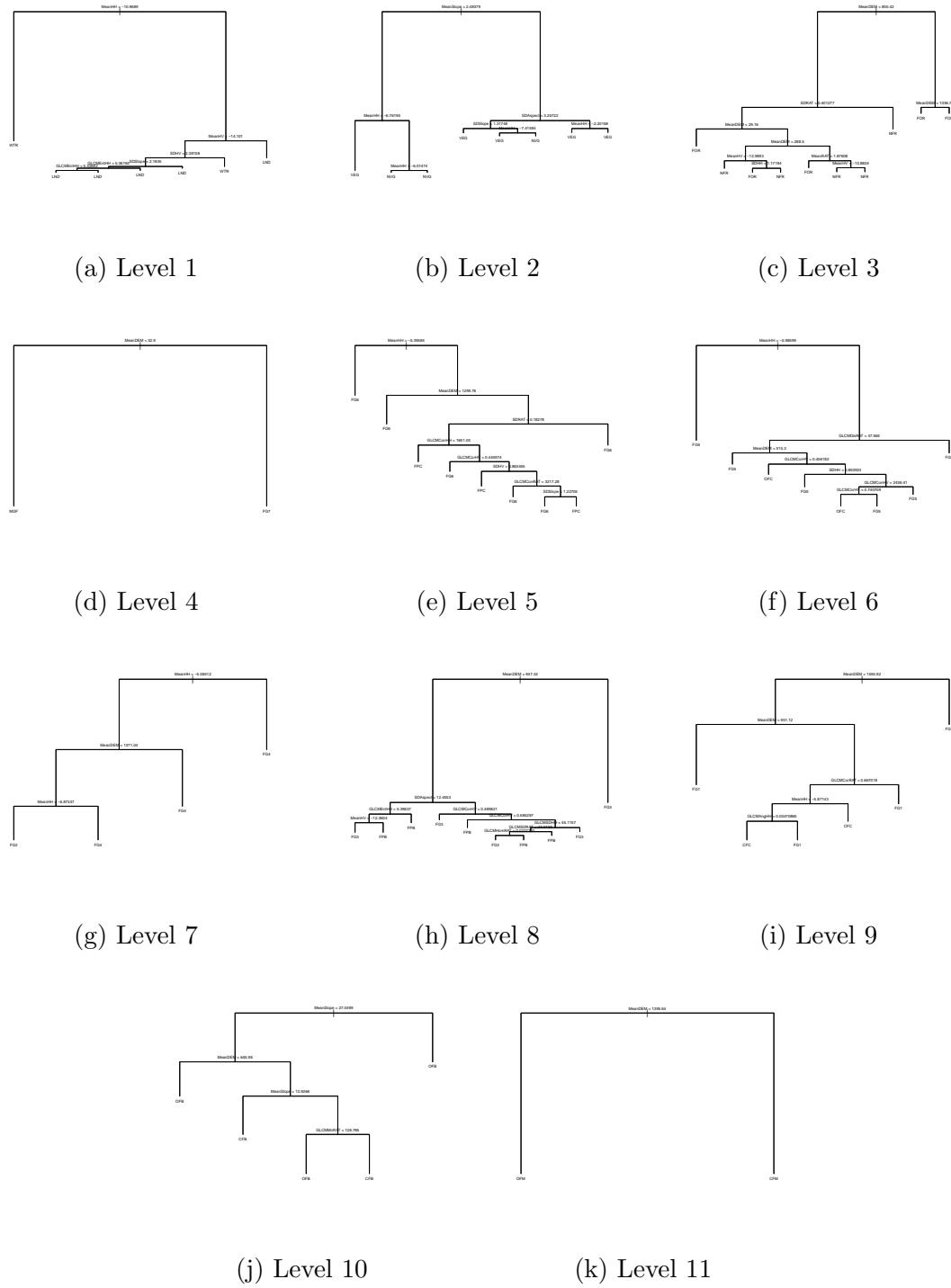


Fig. C.31: The Case 4 2010 classification trees generated by R tree package.

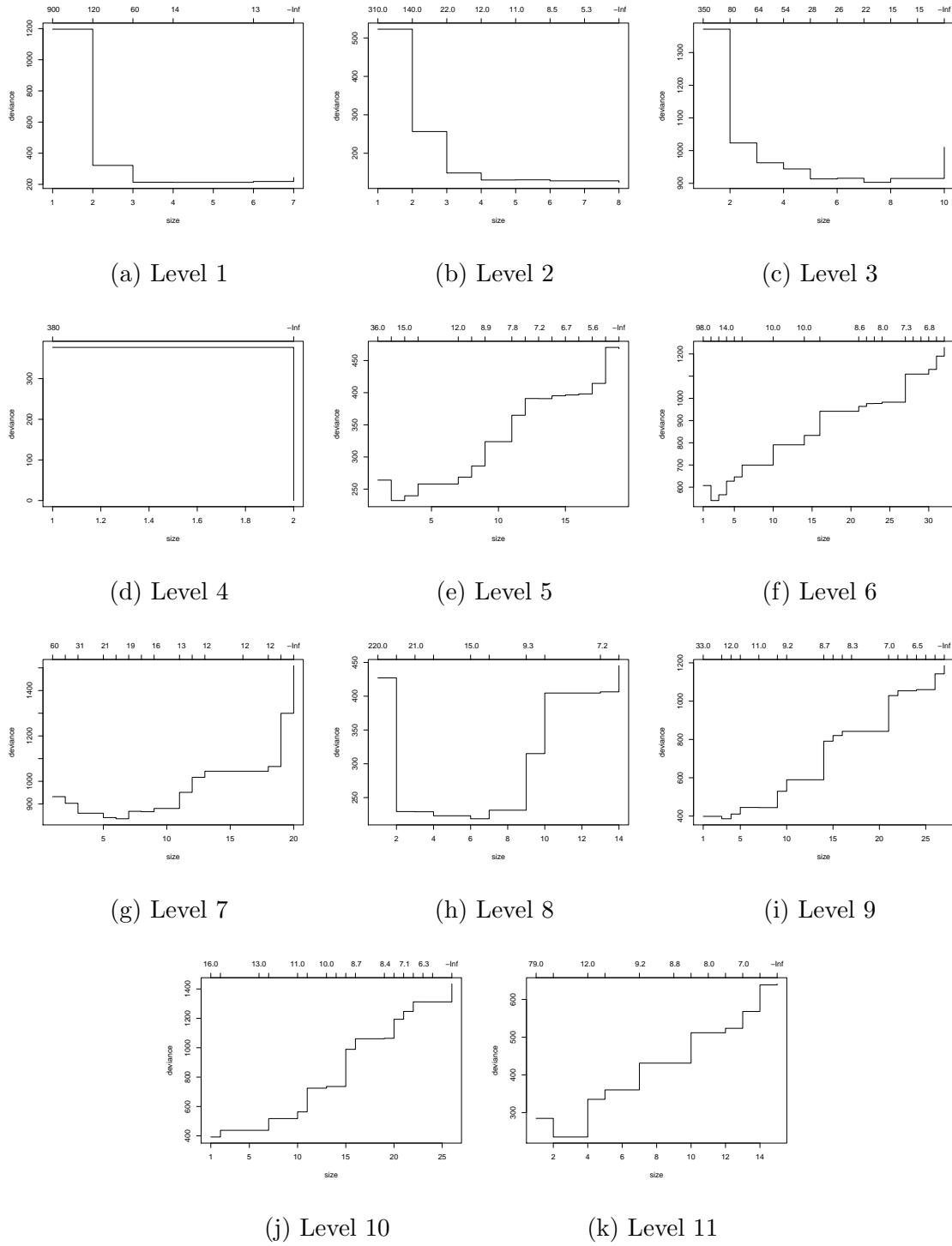


Fig. C.32: The Case 4 2010 cross-validation graphs generated by R tree package.

C.2 Classification error rates for each case

Table C.1: Summary table of classification error rates for Case 1: Polarisation.

Level	2007		2008		2009		2007	
	Class	Error Rate (%)						
1	LND, WTR	3.039	LND, WTR	1.934	LND, WTR	1.657	LND, WTR	1.036
2	VEG, NVG	2.583	VEG, NVG	2.502	VEG, NVG	2.018	VEG, NVG	2.098
3	FOR, NFR	20.9	FOR, NFR	22.87	FOR, NFR	18.6	FOR, NFR	22.61
4	MGF, FG7	1.053	MGF, FG7	1.287	MGF, FG7	1.053	MGF, FG7	1.287
5	OFC, FG6	11.66	OFC, FG6	9.801	OFC, FG6	10.92	OFC, FG6	10.79
6	FG3, FG5	27.94	FG4, FG5	31.91	CFB, FG5	13.62	FG4, FG5	30.07
7	FG1, FG4	29.59	CFB, FG1	25.91	FPC, FG4	3.471	FPC, FG3	5.956
8	OFM, FG2	15.28	FG2, FG3	18.18	FG1, FG3	41.74	CFB, FG1	23.83
9	CFB, OFB	24.13	FPC, CFC	13.85	CFC, FG2	23.53	OFM, FG2	20.42
10	CFC, FPC	12.31	FPB, OFM	13.76	FPB, OFM	13.23	CFC, FPB	15.26
11	CFM, FPB	15.26	OFB, CFM	16.78	OFB, CFM	25.87	CFM, OFB	26.92

Table C.2: Summary table of classification error rates for Case 2: Polarisation + Topographic.

Level	2007		2008		2009		2007	
	Class	Error Rate (%)						
1	LND, WTR	2.279	LND, WTR	1.519	LND, WTR	1.381	LND, WTR	1.036
2	VEG, NVG	0.565	VEG, NVG	0.484	VEG, NVG	0.323	VEG, NVG	0.242
3	FOR, NFR	13.82	FOR, NFR	13.23	FOR, NFR	13.91	FOR, NFR	13.74
4	MGF, FG7	0	MGF, FG7	0	MGF, FG7	0	MGF, FG7	0
5	CFB, FG6	19.11	CFB, FG6	21.34	CFB, FG6	9.181	CFB, FG6	18.73
6	CFM, FG4	11.86	CFM, FG4	7.442	OFC, FG5	10.34	OFC, FG4	17.44
7	FPB, FG3	7.181	CFB, FG2	16.76	FG2, FG4	13.88	CFB, FG2	9.574
8	FPC, FG2	9.394	FG1, FG3	13.03	FG1, FG3	11.58	CFM, FG1	11.74
9	CFB, OFB	13.64	OFM, OFC	10	FPB, OFB	10.14	OFC, OFM	10
10	OFC, OFM	13.5	FPB, OFB	9.783	FPC, CFC	10	OFB, FPB	10.87
11	CFC, FPC	10	CFC, FPC	10	CFM, OFM	11.06	CFC, FPC	9.231

Table C.3: Summary table of classification error rates for Case 3: Polarisation + Texture.

Level	2007		2008		2009		2007	
	Class	Error Rate (%)						
1	LND, WTR	2.141	LND, WTR	0.967	LND, WTR	1.174	LND, WTR	1.036
2	VEG, NVG	2.018	VEG, NVG	1.937	VEG, NVG	1.856	VEG, NVG	1.372
3	FOR, NFR	20.9	FOR, NFR	21.33	FOR, NFR	19.03	FOR, NFR	21.42
4	MGF, FG7	0.702	MGF, FG7	1.404	MGF, FG7	1.053	MGF, FG7	1.053
5	FPC, FG6	2.233	FPC, FG6	2.233	FPC, FG6	1.613	FPC, FG6	2.357
6	FG4, FG5	26.42	FG4, FG5	17.14	OFC, FG5	4.768	OFC, FG5	6.186
7	CFM, FG3	6.513	OFB, FG3	20.87	CFC, FG4	31.41	OFB, FG4	28.3
8	CFC, FG2	7.667	CFC, OFC	7.96	CFB, FG1	10.39	FPB, FG3	7.969
9	CFB, FG1	9.043	CFM, FG1	7.958	FPB, FG2	10	CFC, FG2	9.365
10	OFC, OFM	9	CFB, OFB	10.14	CFC, CFM	10.5	CFM, OFM	8.543
11	FPB, OFB	8.696	FPB, OFM	9.524	OFB, OFM	7.018	CFB, OFB	8.741

Table C.4: Summary table of classification error rates for Case 4: Polarisation + Topographic + Texture.

Level	2007		2008		2009		2007	
	Class	Error Rate (%)						
1	LND, WTR	1.796	LND, WTR	0.829	LND, WTR	1.174	LND, WTR	1.036
2	VEG, NVG	0.565	VEG, NVG	0.646	VEG, NVG	0.323	VEG, NVG	0.242
3	FOR, NFR	13.23	FOR, NFR	12.46	FOR, NFR	13.91	FOR, NFR	13.74
4	MGF, FG7	0	MGF, FG7	0	MGF, FG7	0	MGF, FG7	0
5	FPC, FG6	1.737	FPC, FG6	1.861	FPC, FG6	2.233	FPC, FG6	1.365
6	FG4, FG5	21.52	FG4, FG5	16.37	FG4, FG5	18.69	OFC, FG5	6.186
7	CFM, FG3	3.361	FG1, FG3	16.17	OFC, FG2	7.641	FG2, FG4	20.3
8	CFC, FG1	6.333	CFC, OFC	7.96	CFB, FG3	6.947	FPB, FG3	2.314
9	CFB, FG2	9.043	CFM, FG2	4.844	FPB, FG1	4.533	CFC, FG1	7.358
10	FPB, OFB	7.971	FPB, OFM	6.349	CFC, CFM	7	CFB, OFB	9.441
11	OFC, OFM	9	CFB, OFB	11.19	OFB, OFM	6.667	CFM, OFM	8.543

C.3 Notes on implementation of the methodology

The author has provided the following notes as an informal advice for researchers, practitioners, and students who intend to implement, replicate, or build on the methods used in this study for their own research work.

A basic assumption is that the reader is familiar with fundamental remote sensing concepts and in implementing digital image processing techniques. A key understanding of the following topics will be required, particularly on synthetic aperture radar, forest structure, land cover classification and monitoring, and statistical approaches such as hierarchical clustering and decision tree classification. Users are referred to the main bibliographical references that were cited in the text for further reading on these specific topics.

Concerning software tools, the main applications used in the study were the following: ENVI, QGIS, R, eCognition. ENVI and QGIS are standard image processing and spatial analysis platforms, respectively, which are assumed to basic tools that users are familiar with. Learning additional software tools, particularly R and eCognition, are required to be able to implement the methods used in this study. For new users, the learning curves required may be moderate to high for R software, particularly in scripting; and low to moderate for eCognition, particularly on ruleset development, and could translate to 1-2 months of intensive learning of the softwares. Users are also encouraged to register for an account in the Trimble eCognition Community ([link](#)) where tutorials, sample rulesets, and support from community of practitioners may be accessed. The R scripts used in this study are provided in Appendix 7 for reference.

Regarding hardware, all image pre-processing tasks were successfully carried out

on Windows XP operating system with 8GB RAM running within a VirtualBox virtual machine that was executed on MacBook Pro machine with Mac OSX v.10.11 operating system, 64-bit 2.5 GHz Intel Core i7 processor, and 16GB 1600 MHz DDR RAM. Statistical and spatial analyses were successfully implemented on native Mac OSX platform. For image segmentation, high computing power was required to complete the task of generating segments over a large regional image mosaic. Hence, an HP Elitebook 8770w workstation was used with 64-bit Windows 7 operating system and 32GB RAM specifications, which finished the image segmentation tasks for one image mosaic within 4-5 hours.

C.4 R scripts developed from this study

C.4.1. Sample script for generating boxplots.

```
# This R script generates box-whisker plots using ggplot2 package.
# The aim of generating the boxplots is to visualise the distribution of
# values for
# each land cover type per year for each feature predictor variable.
#
# Script By:      Jose Don T. De Alba
# Date Created: 09 May 2015
# Last Modified: 04 Mar 2016

# Set working directory
# Note: Need to change the path name
setwd("/Users/dondealba/Image_Processing/Thesis/R/1_boxplots/")

# Load the required R libraries
library(ggplot2)

# [1] Image Objects by Chessboard Segmentation

#### LOAD DATA

# Read data, define variables, and store data in variables
# Note: Need to change filenames
ALL2 <- read.csv(file="PALSARwDEM-Stats-Selected-Join-Date-2010.csv",
                 header=TRUE, sep=",")
sardem2 <- subset(ALL2, subset=(LCCode=="FCFB" | LCCode=="FCFC" | LCCode
                                == "FCFM" | LCCode=="FFPB" | LCCode=="FFPC" | LCCode=="FOFB" | LCCode
                                == "FOFC" | LCCode=="FOFM" | LCCode=="LVMG"))

#### MEAN ASPECT
# Generate boxplots per land cover type across all years
mnASP2 <- ggplot() + geom_boxplot(aes(y = MeanAspect, x = LCCode, fill=
                                         factor(AcquisitionDate)), data=sardem2, outlier.shape = 1, outlier.size
                                         = 1)
mnASP2 <- mnASP2 + labs(title="Mean_Aspect_of_Forest_Cover_Types_in_"
                           "Northern_Luzon", x="Land_Cover_Type", y="Value_(degrees)", fill=""
                           "Year")
mnASP2 <- mnASP2 + ylim(0,360)

#### MEAN DEM
# Generate boxplots per land cover type across all years
mnDEM2 <- ggplot() + geom_boxplot(aes(y = MeandEM, x = LCCode, fill=
                                         factor(AcquisitionDate)), data=sardem2, outlier.shape = 1, outlier.size
```

```

= 1)
mnDEM2 <- mnDEM2 + labs(title="Mean DEM of Forest Cover Types in
Northern Luzon", x="Land_Cover_Type", y="Value_(m)", fill="Year")
mnDEM2 <- mnDEM2 + ylim(0,2500)

#### MEAN HH
# Generate boxplots per land cover type across all years
mnHH2 <- ggplot() + geom_boxplot(aes(y = MeanHH,x = LCCode, fill=factor(
AcquisitionDate)),data=sardem2,outlier.shape = 1,outlier.size = 1)
mnHH2 <- mnHH2 + labs(title="Mean HH of Forest Cover Types in Northern
Luzon", x="Land_Cover_Type", y="Value_(dB)", fill="Year")
mnHH2 <- mnHH2 + ylim(-12,0)

#### MEAN HV
# Generate boxplots per land cover type across all years
mnHV2 <- ggplot() + geom_boxplot(aes(y = MeanHV,x = LCCode, fill=factor(
AcquisitionDate)),data=sardem2,outlier.shape = 1,outlier.size = 1)
mnHV2 <- mnHV2 + labs(title="Mean HV of Forest Cover Types in Northern
Luzon", x="Land_Cover_Type", y="Value_(dB)", fill="Year")
mnHV2 <- mnHV2 + ylim(-18,-6)

#### MEAN HH/HV
# Generate boxplots per land cover type across all years
mnRAT2 <- ggplot() + geom_boxplot(aes(y = MeanRAT,x = LCCode, fill=
factor(AcquisitionDate)),data=sardem2,outlier.shape = 1,outlier.size
= 1)
mnRAT2 <- mnRAT2 + labs(title="Mean HH/HV of Forest Cover Types in
Northern Luzon", x="Land_Cover_Type", y="Value", fill="Year")
mnRAT2 <- mnRAT2 + ylim(1,4)

#### MEAN SLOPE
# Generate boxplots per land cover type across all years
mnSLOPE2 <- ggplot() + geom_boxplot(aes(y = MeanSlope,x = LCCode, fill=
factor(AcquisitionDate)),data=sardem2,outlier.shape = 1,outlier.size
= 1)
mnSLOPE2 <- mnSLOPE2 + labs(title="Mean Slope of Forest Cover Types in
Northern Luzon", x="Land_Cover_Type", y="Value_(degrees)", fill="Year")
mnSLOPE2 <- mnSLOPE2 + ylim(0,90)

#### BOX-WHISKER PLOTS

# Output boxplots to a PDF file
ggsave(mnASP2, file="Boxplot-2010-Mean-Aspect-chess.pdf", width=7,
height=5.5)
ggsave(mnDEM2, file="Boxplot-2010-Mean-DEM-chess.pdf", width=7, height
=5.5)

```

```
ggsave(mnHH2, file="Boxplot-2010-Mean-HH-chess.pdf", width=7, height  
=5.5)  
ggsave(mnHV2, file="Boxplot-2010-Mean-HV-chess.pdf", width=7, height  
=5.5)  
ggsave(mnRAT2, file="Boxplot-2010-Mean-RAT-chess.pdf", width=7, height  
=5.5)  
ggsave(mnSLOPE2, file="Boxplot-2010-Mean-Slope-chess.pdf", width=7,  
height=5.5)
```

C.4.2. Sample script for hierarchical clustering.

```

# This R script implements hierarchical clustering using pvclust package
# and compare
# dendrograms using dendextend package. The aim of the clustering is to
# find natural
# clusters of forest types based on predictor variables or features
# derived from
# 5x5 square image object statistics generated by eCognition. The
# purpose of comparing
# dendrograms is to assess consistencies in clustering across years and
# variables by
# evaluating similarities of dendrograms based on available metrics.
#
# Script By:      Jose Don T. De Alba
# Date Created: 22 Jan 2016
# Last Modified: 03 Mar 2016

# Load the required R libraries
library(pvclust)
library(dendextend)
library(magrittr)
library(corrplot)
library(ggplot2)

# CASE 1: POLARISATION

# Set working directory
setwd("/Users/dondealban/Image_Processing/Thesis/R/3_clustering_2016/")

# Read data, define variables, and store data in variables
raw2007 <- as.matrix(read.csv(file="PALSARwDEM-NLuzon-Clustering-2007-
CBS.csv", header=TRUE, sep=",", row.names = 1))
raw2008 <- as.matrix(read.csv(file="PALSARwDEM-NLuzon-Clustering-2008-
CBS.csv", header=TRUE, sep=",", row.names = 1))
raw2009 <- as.matrix(read.csv(file="PALSARwDEM-NLuzon-Clustering-2009-
CBS.csv", header=TRUE, sep=",", row.names = 1))
raw2010 <- as.matrix(read.csv(file="PALSARwDEM-NLuzon-Clustering-2010-
CBS.csv", header=TRUE, sep=",", row.names = 1))

# Normalise variables by rescaling for compatibility of resulting
# clusters
nlz2007 <- scale(raw2007)
nlz2008 <- scale(raw2008)
nlz2009 <- scale(raw2009)
nlz2010 <- scale(raw2010)

# Subset data by selecting variables

```

```

nlz2007c1 <- nlz2007[,c(40:42,48:50)]
nlz2008c1 <- nlz2008[,c(40:42,48:50)]
nlz2009c1 <- nlz2009[,c(40:42,48:50)]
nlz2010c1 <- nlz2010[,c(40:42,48:50)]


# Transpose matrix elements
tnlz2007c1 = t(nlz2007c1)
tnlz2008c1 = t(nlz2008c1)
tnlz2009c1 = t(nlz2009c1)
tnlz2010c1 = t(nlz2010c1)


# Hierarchical clustering using pvclust package
clust.nlz2007c1 <- pvclust(tnlz2007c1, method.hclust="average", method.
    dist="euclidean", nboot=1000)
clust.nlz2008c1 <- pvclust(tnlz2008c1, method.hclust="average", method.
    dist="euclidean", nboot=1000)
clust.nlz2009c1 <- pvclust(tnlz2009c1, method.hclust="average", method.
    dist="euclidean", nboot=1000)
clust.nlz2010c1 <- pvclust(tnlz2010c1, method.hclust="average", method.
    dist="euclidean", nboot=1000)

# Plot dendograms and save as PDF file
pdf("output-dendrogram-2007-case1.pdf", width=7, height=5.5)
clust.nlz2007c1 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
    "Dendrogram: 2007_Case_1")
clust.nlz2007c1 %>% text
clust.nlz2007c1 %>% pvrect(alpha=0.95)
dev.off()
pdf("output-dendrogram-2008-case1.pdf", width=7, height=5.5)
clust.nlz2008c1 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
    "Dendrogram: 2008_Case_1")
clust.nlz2008c1 %>% text
clust.nlz2008c1 %>% pvrect(alpha=0.95)
dev.off()
pdf("output-dendrogram-2009-case1.pdf", width=7, height=5.5)
clust.nlz2009c1 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
    "Dendrogram: 2009_Case_1")
clust.nlz2009c1 %>% text
clust.nlz2009c1 %>% pvrect(alpha=0.95)
dev.off()
pdf("output-dendrogram-2010-case1.pdf", width=7, height=5.5)
clust.nlz2010c1 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
    "Dendrogram: 2010_Case_1")
clust.nlz2010c1 %>% text
clust.nlz2010c1 %>% pvrect(alpha=0.95)
dev.off()

# Convert pvclust object variables to dendextend object variables

```

```

dend2007c1 <- as.dendrogram(clust.nlz2007c1)
dend2008c1 <- as.dendrogram(clust.nlz2008c1)
dend2009c1 <- as.dendrogram(clust.nlz2009c1)
dend2010c1 <- as.dendrogram(clust.nlz2010c1)

# Create a dendlist object from several dendrograms
list.dend78910c1 <- dendlist(dend2007c1, dend2008c1, dend2009c1,
                               dend2010c1)

# Perform global comparison of two or more dendrograms to test if they
are equal
equal.dend78910c1 <- all.equal(list.dend78910c1)

# Generate correlation matrix between a list of trees
cormat.dend78910c1 <- cor.dendlist(list.dend78910c1)

# Compute Baker's Gamma Index
bg0708c1 <- cor_bakers_gamma(dend2007c1, dend2008c1)
bg0709c1 <- cor_bakers_gamma(dend2007c1, dend2009c1)
bg0710c1 <- cor_bakers_gamma(dend2007c1, dend2010c1)
bg0809c1 <- cor_bakers_gamma(dend2008c1, dend2009c1)
bg0810c1 <- cor_bakers_gamma(dend2008c1, dend2010c1)
bg0910c1 <- cor_bakers_gamma(dend2009c1, dend2010c1)

# Compute Cophenetic Correlation coefficient
cr0708c1 <- cor_cophenetic(dend2007c1, dend2008c1)
cr0709c1 <- cor_cophenetic(dend2007c1, dend2009c1)
cr0710c1 <- cor_cophenetic(dend2007c1, dend2010c1)
cr0809c1 <- cor_cophenetic(dend2008c1, dend2009c1)
cr0810c1 <- cor_cophenetic(dend2008c1, dend2010c1)
cr0910c1 <- cor_cophenetic(dend2009c1, dend2010c1)

# Compute Common Nodes
cn0708c1 <- cor_common_nodes(dend2007c1, dend2008c1)
cn0709c1 <- cor_common_nodes(dend2007c1, dend2009c1)
cn0710c1 <- cor_common_nodes(dend2007c1, dend2010c1)
cn0809c1 <- cor_common_nodes(dend2008c1, dend2009c1)
cn0810c1 <- cor_common_nodes(dend2008c1, dend2010c1)
cn0910c1 <- cor_common_nodes(dend2009c1, dend2010c1)

# Generate summary of dendrogram comparisons and save as TXT file
sink("output-summary-case1.txt", append=FALSE, split=TRUE)
print("DENDLIST_OBJECT")
print(list.dend78910c1)
print("GLOBAL_COMPARISON_OF_DENDROGRAMS")
print(equal.dend78910c1)
print("CORRELATION_MATRIX_BETWEEN_TREES")
print(cormat.dend78910c1)

```

```

print("BAKER'S_GAMMA_INDICES")
print(bg0708c1)
print(bg0709c1)
print(bg0710c1)
print(bg0809c1)
print(bg0810c1)
print(bg0910c1)
print("COPHENETIC_CORRELATION")
print(cr0708c1)
print(cr0709c1)
print(cr0710c1)
print(cr0809c1)
print(cr0810c1)
print(cr0910c1)
print("COMMON_NODES")
print(cn0708c1)
print(cn0709c1)
print(cn0710c1)
print(cn0809c1)
print(cn0810c1)
print(cn0910c1)
sink()

# CASE 2: POLARISATION + TOPOGRAPHIC

# Set working directory
setwd("/Users/dondealban/Image_Processing/Thesis/R/3_clustering_2016/")

# Subset data by selecting variables
nlz2007c2 <- nlz2007[,c(40:45,48:53)]
nlz2008c2 <- nlz2008[,c(40:45,48:53)]
nlz2009c2 <- nlz2009[,c(40:45,48:53)]
nlz2010c2 <- nlz2010[,c(40:45,48:53)]

# Transpose matrix elements
tnlz2007c2 = t(nlz2007c2)
tnlz2008c2 = t(nlz2008c2)
tnlz2009c2 = t(nlz2009c2)
tnlz2010c2 = t(nlz2010c2)

# Hierarchical clustering using pvclust package
clust.nlz2007c2 <- pvclust(tnlz2007c2, method.hclust="average", method.
  dist="euclidean", nboot=1000)
clust.nlz2008c2 <- pvclust(tnlz2008c2, method.hclust="average", method.
  dist="euclidean", nboot=1000)
clust.nlz2009c2 <- pvclust(tnlz2009c2, method.hclust="average", method.
  dist="euclidean", nboot=1000)
clust.nlz2010c2 <- pvclust(tnlz2010c2, method.hclust="average", method.

```

```

dist="euclidean", nboot=1000)

# Plot dendograms and save as PDF file
pdf("output-dendrogram-2007-case2.pdf", width=7, height=5.5)
clust.nlz2007c2 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
  "Dendrogram:_2007_Case_2")
clust.nlz2007c2 %>% text
clust.nlz2007c2 %>% pvrect(alpha=0.95)
dev.off()
pdf("output-dendrogram-2008-case2.pdf", width=7, height=5.5)
clust.nlz2008c2 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
  "Dendrogram:_2008_Case_2")
clust.nlz2008c2 %>% text
clust.nlz2008c2 %>% pvrect(alpha=0.95)
dev.off()
pdf("output-dendrogram-2009-case2.pdf", width=7, height=5.5)
clust.nlz2009c2 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
  "Dendrogram:_2009_Case_2")
clust.nlz2009c2 %>% text
clust.nlz2009c2 %>% pvrect(alpha=0.95)
dev.off()
pdf("output-dendrogram-2010-case2.pdf", width=7, height=5.5)
clust.nlz2010c2 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
  "Dendrogram:_2010_Case_2")
clust.nlz2010c2 %>% text
clust.nlz2010c2 %>% pvrect(alpha=0.95)
dev.off()

# Convert pvclust object variables to dendextend object variables
dend2007c2 <- as.dendrogram(clust.nlz2007c2)
dend2008c2 <- as.dendrogram(clust.nlz2008c2)
dend2009c2 <- as.dendrogram(clust.nlz2009c2)
dend2010c2 <- as.dendrogram(clust.nlz2010c2)

# Create a dendlist object from several dendograms
list.dend78910c2 <- dendlist(dend2007c2, dend2008c2, dend2009c2,
  dend2010c2)

# Perform global comparison of two or more dendograms to test if they
# are equal
equal.dend78910c2 <- all.equal(list.dend78910c2)

# Generate correlation matrix between a list of trees
cormat.dend78910c2 <- cor.dendlist(list.dend78910c2)

# Compute Baker's Gamma Index
bg0708c2 <- cor_bakers_gamma(dend2007c2, dend2008c2)
bg0709c2 <- cor_bakers_gamma(dend2007c2, dend2009c2)

```

```

bg0710c2 <- cor_bakers_gamma(dend2007c2, dend2010c2)
bg0809c2 <- cor_bakers_gamma(dend2008c2, dend2009c2)
bg0810c2 <- cor_bakers_gamma(dend2008c2, dend2010c2)
bg0910c2 <- cor_bakers_gamma(dend2009c2, dend2010c2)

# Compute Cophenetic Correlation coefficient
cr0708c2 <- cor_cophenetic(dend2007c2, dend2008c2)
cr0709c2 <- cor_cophenetic(dend2007c2, dend2009c2)
cr0710c2 <- cor_cophenetic(dend2007c2, dend2010c2)
cr0809c2 <- cor_cophenetic(dend2008c2, dend2009c2)
cr0810c2 <- cor_cophenetic(dend2008c2, dend2010c2)
cr0910c2 <- cor_cophenetic(dend2009c2, dend2010c2)

# Compute Common Nodes
cn0708c2 <- cor_common_nodes(dend2007c2, dend2008c2)
cn0709c2 <- cor_common_nodes(dend2007c2, dend2009c2)
cn0710c2 <- cor_common_nodes(dend2007c2, dend2010c2)
cn0809c2 <- cor_common_nodes(dend2008c2, dend2009c2)
cn0810c2 <- cor_common_nodes(dend2008c2, dend2010c2)
cn0910c2 <- cor_common_nodes(dend2009c2, dend2010c2)

# Generate summary of dendrogram comparisons and save as TXT file
sink("output-summary-case2.txt", append=FALSE, split=TRUE)
print("DENDLIST_OBJECT")
print(list.dend78910c2)
print("GLOBAL_COMPARISON_OF_DENDROGRAMS")
print(equal.dend78910c2)
print("CORRELATION_MATRIX_BETWEEN_TREES")
print(cormat.dend78910c2)
print("BAKER'S_GAMMA_INDICES")
print(bg0708c2)
print(bg0709c2)
print(bg0710c2)
print(bg0809c2)
print(bg0810c2)
print(bg0910c2)
print("COPHENETIC_CORRELATION")
print(cr0708c2)
print(cr0709c2)
print(cr0710c2)
print(cr0809c2)
print(cr0810c2)
print(cr0910c2)
print("COMMON_NODES")
print(cn0708c2)
print(cn0709c2)
print(cn0710c2)
print(cn0809c2)

```

```

print(cn0810c2)
print(cn0910c2)
sink()

# CASE 3: POLARISATION + TEXTURE

# Set working directory
setwd("/Users/dondealban/Image_Processing/Thesis/R/3_clustering_2016/")

# Subset data by selecting variables
nlz2007c3 <- nlz2007[,c
(7:9,11:13,15:17,19:21,23:25,27:29,31:33,35:37,40:42,48:50)]
nlz2008c3 <- nlz2008[,c
(7:9,11:13,15:17,19:21,23:25,27:29,31:33,35:37,40:42,48:50)]
nlz2009c3 <- nlz2009[,c
(7:9,11:13,15:17,19:21,23:25,27:29,31:33,35:37,40:42,48:50)]
nlz2010c3 <- nlz2010[,c
(7:9,11:13,15:17,19:21,23:25,27:29,31:33,35:37,40:42,48:50)]

# Transpose matrix elements
tnlz2007c3 = t(nlz2007c3)
tnlz2008c3 = t(nlz2008c3)
tnlz2009c3 = t(nlz2009c3)
tnlz2010c3 = t(nlz2010c3)

# Hierarchical clustering using pvclust package
clust.nlz2007c3 <- pvclust(tnlz2007c3, method.hclust="average", method.
dist="euclidean", nboot=1000)
clust.nlz2008c3 <- pvclust(tnlz2008c3, method.hclust="average", method.
dist="euclidean", nboot=1000)
clust.nlz2009c3 <- pvclust(tnlz2009c3, method.hclust="average", method.
dist="euclidean", nboot=1000)
clust.nlz2010c3 <- pvclust(tnlz2010c3, method.hclust="average", method.
dist="euclidean", nboot=1000)

# Plot dendograms and save as PDF file
pdf("output-dendrogram-2007-case3.pdf", width=7, height=5.5)
clust.nlz2007c3 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
Dendrogram:_2007_Case_3")
clust.nlz2007c3 %>% text
clust.nlz2007c3 %>% pvrect(alpha=0.95)
dev.off()
pdf("output-dendrogram-2008-case3.pdf", width=7, height=5.5)
clust.nlz2008c3 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
Dendrogram:_2008_Case_3")
clust.nlz2008c3 %>% text
clust.nlz2008c3 %>% pvrect(alpha=0.95)
dev.off()

```

```

pdf("output-dendrogram-2009-case3.pdf", width=7, height=5.5)
clust.nlz2009c3 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
  "Dendrogram: 2009_Case_3")
clust.nlz2009c3 %>% text
clust.nlz2009c3 %>% pvrect(alpha=0.95)
dev.off()
pdf("output-dendrogram-2010-case3.pdf", width=7, height=5.5)
clust.nlz2010c3 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
  "Dendrogram: 2010_Case_3")
clust.nlz2010c3 %>% text
clust.nlz2010c3 %>% pvrect(alpha=0.95)
dev.off()

# Convert pvclust object variables to dendextend object variables
dend2007c3 <- as.dendrogram(clust.nlz2007c3)
dend2008c3 <- as.dendrogram(clust.nlz2008c3)
dend2009c3 <- as.dendrogram(clust.nlz2009c3)
dend2010c3 <- as.dendrogram(clust.nlz2010c3)

# Create a dendlist object from several dendrograms
list.dend78910c3 <- dendlist(dend2007c3, dend2008c3, dend2009c3,
  dend2010c3)

# Perform global comparison of two or more dendrograms to test if they
# are equal
equal.dend78910c3 <- all.equal(list.dend78910c3)

# Generate correlation matrix between a list of trees
cormat.dend78910c3 <- cor.dendlist(list.dend78910c3)

# Compute Baker's Gamma Index
bg0708c3 <- cor_bakers_gamma(dend2007c3, dend2008c3)
bg0709c3 <- cor_bakers_gamma(dend2007c3, dend2009c3)
bg0710c3 <- cor_bakers_gamma(dend2007c3, dend2010c3)
bg0809c3 <- cor_bakers_gamma(dend2008c3, dend2009c3)
bg0810c3 <- cor_bakers_gamma(dend2008c3, dend2010c3)
bg0910c3 <- cor_bakers_gamma(dend2009c3, dend2010c3)

# Compute Cophenetic Correlation coefficient
cr0708c3 <- cor_cophenetic(dend2007c3, dend2008c3)
cr0709c3 <- cor_cophenetic(dend2007c3, dend2009c3)
cr0710c3 <- cor_cophenetic(dend2007c3, dend2010c3)
cr0809c3 <- cor_cophenetic(dend2008c3, dend2009c3)
cr0810c3 <- cor_cophenetic(dend2008c3, dend2010c3)
cr0910c3 <- cor_cophenetic(dend2009c3, dend2010c3)

# Compute Common Nodes
cn0708c3 <- cor_common_nodes(dend2007c3, dend2008c3)

```

```

cn0709c3 <- cor_common_nodes(dend2007c3, dend2009c3)
cn0710c3 <- cor_common_nodes(dend2007c3, dend2010c3)
cn0809c3 <- cor_common_nodes(dend2008c3, dend2009c3)
cn0810c3 <- cor_common_nodes(dend2008c3, dend2010c3)
cn0910c3 <- cor_common_nodes(dend2009c3, dend2010c3)

# Generate summary of dendrogram comparisons and save as TXT file
sink("output-summary-case3.txt", append=FALSE, split=TRUE)
print("DENDLIST_OBJECT")
print(list.dend78910c3)
print("GLOBAL_COMPARISON_OF_DENDROGRAMS")
print(equal.dend78910c3)
print("CORRELATION_MATRIX_BETWEEN_TREES")
print(cormat.dend78910c3)
print("BAKER'S_GAMMA_INDICES")
print(bg0708c3)
print(bg0709c3)
print(bg0710c3)
print(bg0809c3)
print(bg0810c3)
print(bg0910c3)
print("COPHENETIC_CORRELATION")
print(cr0708c3)
print(cr0709c3)
print(cr0710c3)
print(cr0809c3)
print(cr0810c3)
print(cr0910c3)
print("COMMON_NODES")
print(cn0708c3)
print(cn0709c3)
print(cn0710c3)
print(cn0809c3)
print(cn0810c3)
print(cn0910c3)
sink()

# CASE 4: POLARISATION + TOPOGRAPHIC + TEXTURE

# Set working directory
setwd("/Users/dondealban/Image_Processing/Thesis/R/3_clustering_2016/")

# Subset data by selecting variables
nlz2007c4 <- nlz2007[, c
(7:9, 11:13, 15:17, 19:21, 23:25, 27:29, 31:33, 35:37, 40:45, 48:53)]
nlz2008c4 <- nlz2008[, c
(7:9, 11:13, 15:17, 19:21, 23:25, 27:29, 31:33, 35:37, 40:45, 48:53)]
nlz2009c4 <- nlz2009[, c

```

```

(7:9 ,11:13 ,15:17 ,19:21 ,23:25 ,27:29 ,31:33 ,35:37 ,40:45 ,48:53) ]
nlz2010c4 <- nlz2010 [,c
(7:9 ,11:13 ,15:17 ,19:21 ,23:25 ,27:29 ,31:33 ,35:37 ,40:45 ,48:53) ]

# Transpose matrix elements
tnlz2007c4 = t(nlz2007c4)
tnlz2008c4 = t(nlz2008c4)
tnlz2009c4 = t(nlz2009c4)
tnlz2010c4 = t(nlz2010c4)

# Hierarchical clustering using pvclust package
clust.nlz2007c4 <- pvclust(tnlz2007c4, method.hclust="average", method.
    dist="euclidean", nboot=1000)
clust.nlz2008c4 <- pvclust(tnlz2008c4, method.hclust="average", method.
    dist="euclidean", nboot=1000)
clust.nlz2009c4 <- pvclust(tnlz2009c4, method.hclust="average", method.
    dist="euclidean", nboot=1000)
clust.nlz2010c4 <- pvclust(tnlz2010c4, method.hclust="average", method.
    dist="euclidean", nboot=1000)

# Plot dendograms and save as PDF file
pdf("output-dendrogram-2007-case4.pdf", width=7, height=5.5)
clust.nlz2007c4 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
    "Dendrogram:_2007_Case_4")
clust.nlz2007c4 %>% text
clust.nlz2007c4 %>% pvrect(alpha=0.95)
dev.off()
pdf("output-dendrogram-2008-case4.pdf", width=7, height=5.5)
clust.nlz2008c4 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
    "Dendrogram:_2008_Case_4")
clust.nlz2008c4 %>% text
clust.nlz2008c4 %>% pvrect(alpha=0.95)
dev.off()
pdf("output-dendrogram-2009-case4.pdf", width=7, height=5.5)
clust.nlz2009c4 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
    "Dendrogram:_2009_Case_4")
clust.nlz2009c4 %>% text
clust.nlz2009c4 %>% pvrect(alpha=0.95)
dev.off()
pdf("output-dendrogram-2010-case4.pdf", width=7, height=5.5)
clust.nlz2010c4 %>% as.dendrogram %>% hang.dendrogram %>% plot(main =
    "Dendrogram:_2010_Case_4")
clust.nlz2010c4 %>% text
clust.nlz2010c4 %>% pvrect(alpha=0.95)
dev.off()

# Convert pvclust object variables to dendextend object variables
dend2007c4 <- as.dendrogram(clust.nlz2007c4)

```

```

dend2008c4 <- as.dendrogram(clust.nlz2008c4)
dend2009c4 <- as.dendrogram(clust.nlz2009c4)
dend2010c4 <- as.dendrogram(clust.nlz2010c4)

# Create a dendlist object from several dendrograms
list.dend78910c4 <- dendlist(dend2007c4, dend2008c4, dend2009c4,
                                dend2010c4)

# Perform global comparison of two or more dendrograms to test if they
are equal
equal.dend78910c4 <- all.equal(list.dend78910c4)

# Generate correlation matrix between a list of trees
cormat.dend78910c4 <- cor.dendlist(list.dend78910c4)

# Compute Baker's Gamma Index
bg0708c4 <- cor_bakers_gamma(dend2007c4, dend2008c4)
bg0709c4 <- cor_bakers_gamma(dend2007c4, dend2009c4)
bg0710c4 <- cor_bakers_gamma(dend2007c4, dend2010c4)
bg0809c4 <- cor_bakers_gamma(dend2008c4, dend2009c4)
bg0810c4 <- cor_bakers_gamma(dend2008c4, dend2010c4)
bg0910c4 <- cor_bakers_gamma(dend2009c4, dend2010c4)

# Compute Cophenetic Correlation coefficient
cr0708c4 <- cor_cophenetic(dend2007c4, dend2008c4)
cr0709c4 <- cor_cophenetic(dend2007c4, dend2009c4)
cr0710c4 <- cor_cophenetic(dend2007c4, dend2010c4)
cr0809c4 <- cor_cophenetic(dend2008c4, dend2009c4)
cr0810c4 <- cor_cophenetic(dend2008c4, dend2010c4)
cr0910c4 <- cor_cophenetic(dend2009c4, dend2010c4)

# Compute Common Nodes
cn0708c4 <- cor_common_nodes(dend2007c4, dend2008c4)
cn0709c4 <- cor_common_nodes(dend2007c4, dend2009c4)
cn0710c4 <- cor_common_nodes(dend2007c4, dend2010c4)
cn0809c4 <- cor_common_nodes(dend2008c4, dend2009c4)
cn0810c4 <- cor_common_nodes(dend2008c4, dend2010c4)
cn0910c4 <- cor_common_nodes(dend2009c4, dend2010c4)

# Generate summary of dendrogram comparisons and save as TXT file
sink("output-summary-case4.txt", append=FALSE, split=TRUE)
print("DENDLIST_OBJECT")
print(list.dend78910c4)
print("GLOBAL_COMPARISON_OF_DENDROGRAMS")
print(equal.dend78910c4)
print("CORRELATION_MATRIX_BETWEEN_TREES")
print(cormat.dend78910c4)
print("BAKER'S_GAMMA_INDICES")

```

```
print(bg0708c4)
print(bg0709c4)
print(bg0710c4)
print(bg0809c4)
print(bg0810c4)
print(bg0910c4)
print("COPHENETIC_CORRELATION")
print(cr0708c4)
print(cr0709c4)
print(cr0710c4)
print(cr0809c4)
print(cr0810c4)
print(cr0910c4)
print("COMMON_NODES")
print(cn0708c4)
print(cn0709c4)
print(cn0710c4)
print(cn0809c4)
print(cn0810c4)
print(cn0910c4)
sink()
```

C.4.3. Sample script for decision tree classification.

```

# This R script implements decision tree classification using tree
# package.
# The decision tree classification aims to grow a tree by binary
# recursive partitioning
# using the predictor variables or features derived from image object
# statistics
# generated by eCognition (at scale 10) and then choosing splits by land
# /forest cover types.
#
# Script By:      Jose Don T. De Alban
# Date Created:  09 May 2015
# Last Modified: 06 Mar 2016

# Set working directory
# Note: Need to change the path name as necessary
setwd("/Users/dondealban/Image_Processing/Thesis/R/4_tree_classification
       /2016.03.06/")

# Load the required R libraries
library(tree)

# Read data, define variables, and store data in variables
dat2007 <- read.csv(file="PALSARwDEM-NLuzon-Classification -2007.csv",
                     header=TRUE, sep=",")
dat2008 <- read.csv(file="PALSARwDEM-NLuzon-Classification -2008.csv",
                     header=TRUE, sep=",")
dat2009 <- read.csv(file="PALSARwDEM-NLuzon-Classification -2009.csv",
                     header=TRUE, sep=",")
dat2010 <- read.csv(file="PALSARwDEM-NLuzon-Classification -2010.csv",
                     header=TRUE, sep=",")

# CASE 4: Polarisation + Topographic + Texture

# Insert new standard variables for classification levels 1,2,3 for
# multi-year data

# Level1: land - water

# 2007
dat2007$LC1[dat2007$LCCode=="WNOW" | dat2007$LCCode=="WNRB" | dat2007$-
            LCCode=="WNSS"] <- "WIR"
dat2007$LC1[dat2007$LCCode!="WNOW" & dat2007$LCCode!="WNRB" & dat2007$-
            LCCode!="WNSS"] <- "LND"
# 2008
dat2008$LC1[dat2008$LCCode=="WNOW" | dat2008$LCCode=="WNRB" | dat2008$-
            LCCode=="WNSS"] <- "WIR"

```

```

dat2008$LC1[ dat2008$LCCode!="WNOW" & dat2008$LCCode!="WNRB" & dat2008$  

LCCode!="WNSS" ] <- "LND"  

# 2009  

dat2009$LC1[ dat2009$LCCode=="WNOW" | dat2009$LCCode=="WNRB" | dat2009$  

LCCode=="WNSS" ] <- "WIR"  

dat2009$LC1[ dat2009$LCCode!="WNOW" & dat2009$LCCode!="WNRB" & dat2009$  

LCCode!="WNSS" ] <- "LND"  

# 2010  

dat2010$LC1[ dat2010$LCCode=="WNOW" | dat2010$LCCode=="WNRB" | dat2010$  

LCCode=="WNSS" ] <- "WIR"  

dat2010$LC1[ dat2010$LCCode!="WNOW" & dat2010$LCCode!="WNRB" & dat2010$  

LCCode!="WNSS" ] <- "LND"

# Level2: vegetation - non-vegetation

# 2007  

dat2007$LC2[ dat2007$LC1=="WIR" ] <- "WIR"  

dat2007$LC2[ dat2007$LC1!="WIR" & dat2007$LCCode=="LNUB" ] <- "NVG"  

dat2007$LC2[ dat2007$LC1!="WIR" & dat2007$LCCode!="LNUB" ] <- "VEG"  

# 2008  

dat2008$LC2[ dat2008$LC1=="WIR" ] <- "WIR"  

dat2008$LC2[ dat2008$LC1!="WIR" & dat2008$LCCode=="LNUB" ] <- "NVG"  

dat2008$LC2[ dat2008$LC1!="WIR" & dat2008$LCCode!="LNUB" ] <- "VEG"  

# 2009  

dat2009$LC2[ dat2009$LC1=="WIR" ] <- "WIR"  

dat2009$LC2[ dat2009$LC1!="WIR" & dat2009$LCCode=="LNUB" ] <- "NVG"  

dat2009$LC2[ dat2009$LC1!="WIR" & dat2009$LCCode!="LNUB" ] <- "VEG"  

# 2010  

dat2010$LC2[ dat2010$LC1=="WIR" ] <- "WIR"  

dat2010$LC2[ dat2010$LC1!="WIR" & dat2010$LCCode=="LNUB" ] <- "NVG"  

dat2010$LC2[ dat2010$LC1!="WIR" & dat2010$LCCode!="LNUB" ] <- "VEG"

# Level3: forest - non-forest

# 2007  

dat2007$LC3 <- "OOO"  

dat2007$LC3[ dat2007$LC2=="WIR" ] <- "WIR"  

dat2007$LC3[ dat2007$LC2=="NVG" ] <- "NVG"  

dat2007$LC3[ dat2007$LCCode=="LVGR" | dat2007$LCCode=="LVSH" | dat2007$  

LCCode=="LWVG" ] <- "NFR"  

dat2007$LC3[ dat2007$LC3=="OOO" ] <- "FOR"  

# 2008  

dat2008$LC3 <- "OOO"  

dat2008$LC3[ dat2008$LC2=="WIR" ] <- "WIR"  

dat2008$LC3[ dat2008$LC2=="NVG" ] <- "NVG"  

dat2008$LC3[ dat2008$LCCode=="LVGR" | dat2008$LCCode=="LVSH" | dat2008$  

LCCode=="LWVG" ] <- "NFR"  

dat2008$LC3[ dat2008$LC3=="OOO" ] <- "FOR"

```

```

# 2009
dat2009$LC3 <- "OOO"
dat2009$LC3[ dat2009$LC2=="WIR" ] <- "WIR"
dat2009$LC3[ dat2009$LC2=="NVG" ] <- "NVG"
dat2009$LC3[ dat2009$LCCode=="LVGR" | dat2009$LCCode=="LVSH" | dat2009$LCCode=="LWVG" ] <- "NFR"
dat2009$LC3[ dat2009$LC3=="OOO" ] <- "FOR"
# 2010
dat2010$LC3 <- "OOO"
dat2010$LC3[ dat2010$LC2=="WIR" ] <- "WIR"
dat2010$LC3[ dat2010$LC2=="NVG" ] <- "NVG"
dat2010$LC3[ dat2010$LCCode=="LVGR" | dat2010$LCCode=="LVSH" | dat2010$LCCode=="LWVG" ] <- "NFR"
dat2010$LC3[ dat2010$LC3=="OOO" ] <- "FOR"

# Insert new variables for remaining classification levels based on clustering dendograms

# Level4: Node 8 Split

# 2007
dat2007$LC4 <- "OOO"
dat2007$LC4[ dat2007$LC3=="WIR" ] <- "WIR"
dat2007$LC4[ dat2007$LC3=="NVG" ] <- "NVG"
dat2007$LC4[ dat2007$LC3=="NFR" ] <- "NFR"
dat2007$LC4[ dat2007$LCCode=="LVMG" ] <- "MGF"
dat2007$LC4[ dat2007$LC4=="OOO" ] <- "FG7"
# 2008
dat2008$LC4 <- "OOO"
dat2008$LC4[ dat2008$LC3=="WIR" ] <- "WIR"
dat2008$LC4[ dat2008$LC3=="NVG" ] <- "NVG"
dat2008$LC4[ dat2008$LC3=="NFR" ] <- "NFR"
dat2008$LC4[ dat2008$LCCode=="LVMG" ] <- "MGF"
dat2008$LC4[ dat2008$LC4=="OOO" ] <- "FG7"
# 2009
dat2009$LC4 <- "OOO"
dat2009$LC4[ dat2009$LC3=="WIR" ] <- "WIR"
dat2009$LC4[ dat2009$LC3=="NVG" ] <- "NVG"
dat2009$LC4[ dat2009$LC3=="NFR" ] <- "NFR"
dat2009$LC4[ dat2009$LCCode=="LVMG" ] <- "MGF"
dat2009$LC4[ dat2009$LC4=="OOO" ] <- "FG7"
# 2010
dat2010$LC4 <- "OOO"
dat2010$LC4[ dat2010$LC3=="WIR" ] <- "WIR"
dat2010$LC4[ dat2010$LC3=="NVG" ] <- "NVG"
dat2010$LC4[ dat2010$LC3=="NFR" ] <- "NFR"
dat2010$LC4[ dat2010$LCCode=="LVMG" ] <- "MGF"
dat2010$LC4[ dat2010$LC4=="OOO" ] <- "FG7"

```

```

# Level5: Node 7 Split

# 2007
dat2007$LC5 <- "OOO"
dat2007$LC5[ dat2007$LC4=="WIR" ] <- "WIR"
dat2007$LC5[ dat2007$LC4=="NVG" ] <- "NVG"
dat2007$LC5[ dat2007$LC4=="NFR" ] <- "NFR"
dat2007$LC5[ dat2007$LC4=="MGF" ] <- "MGF"
dat2007$LC5[ dat2007$LCCode=="FFPC" ] <- "FPC"
dat2007$LC5[ dat2007$LC5=="OOO" ] <- "FG6"

# 2008
dat2008$LC5 <- "OOO"
dat2008$LC5[ dat2008$LC4=="WIR" ] <- "WIR"
dat2008$LC5[ dat2008$LC4=="NVG" ] <- "NVG"
dat2008$LC5[ dat2008$LC4=="NFR" ] <- "NFR"
dat2008$LC5[ dat2008$LC4=="MGF" ] <- "MGF"
dat2008$LC5[ dat2008$LCCode=="FFPC" ] <- "FPC"
dat2008$LC5[ dat2008$LC5=="OOO" ] <- "FG6"

# 2009
dat2009$LC5 <- "OOO"
dat2009$LC5[ dat2009$LC4=="WIR" ] <- "WIR"
dat2009$LC5[ dat2009$LC4=="NVG" ] <- "NVG"
dat2009$LC5[ dat2009$LC4=="NFR" ] <- "NFR"
dat2009$LC5[ dat2009$LC4=="MGF" ] <- "MGF"
dat2009$LC5[ dat2009$LCCode=="FFPC" ] <- "FPC"
dat2009$LC5[ dat2009$LC5=="OOO" ] <- "FG6"

# 2010
dat2010$LC5 <- "OOO"
dat2010$LC5[ dat2010$LC4=="WIR" ] <- "WIR"
dat2010$LC5[ dat2010$LC4=="NVG" ] <- "NVG"
dat2010$LC5[ dat2010$LC4=="NFR" ] <- "NFR"
dat2010$LC5[ dat2010$LC4=="MGF" ] <- "MGF"
dat2010$LC5[ dat2010$LCCode=="FFPC" ] <- "FPC"
dat2010$LC5[ dat2010$LC5=="OOO" ] <- "FG6"

# Level6: Node 6 Split

# 2007
dat2007$LC6 <- "OOO"
dat2007$LC6[ dat2007$LC5=="WIR" ] <- "WIR"
dat2007$LC6[ dat2007$LC5=="NVG" ] <- "NVG"
dat2007$LC6[ dat2007$LC5=="NFR" ] <- "NFR"
dat2007$LC6[ dat2007$LC5=="MGF" ] <- "MGF"
dat2007$LC6[ dat2007$LC5=="FPC" ] <- "FPC"
dat2007$LC6[ dat2007$LCCode=="FCFC" | dat2007$LCCode=="FOFC" | dat2007$LCCode=="FOFM" ] <- "FG4"
dat2007$LC6[ dat2007$LC6=="OOO" ] <- "FG5"

```

```

# 2008
dat2008$LC6 <- "OOO"
dat2008$LC6[ dat2008$LC5=="WIR" ] <- "WIR"
dat2008$LC6[ dat2008$LC5=="NVG" ] <- "NVG"
dat2008$LC6[ dat2008$LC5=="NFR" ] <- "NFR"
dat2008$LC6[ dat2008$LC5=="MGF" ] <- "MGF"
dat2008$LC6[ dat2008$LC5=="FPC" ] <- "FPC"
dat2008$LC6[ dat2008$LCCode=="FCFC" | dat2008$LCCode=="FOFC" ] <- "FG4"
dat2008$LC6[ dat2008$LC6=="OOO" ] <- "FG5"
# 2009
dat2009$LC6 <- "OOO"
dat2009$LC6[ dat2009$LC5=="WIR" ] <- "WIR"
dat2009$LC6[ dat2009$LC5=="NVG" ] <- "NVG"
dat2009$LC6[ dat2009$LC5=="NFR" ] <- "NFR"
dat2009$LC6[ dat2009$LC5=="MGF" ] <- "MGF"
dat2009$LC6[ dat2009$LC5=="FPC" ] <- "FPC"
dat2009$LC6[ dat2009$LCCode=="FOFC" | dat2009$LCCode=="FCFC" | dat2009$LCCode=="FCFM" ] <- "FG5"
dat2009$LC6[ dat2009$LC6=="OOO" ] <- "FG4"
# 2010
dat2010$LC6 <- "OOO"
dat2010$LC6[ dat2010$LC5=="WIR" ] <- "WIR"
dat2010$LC6[ dat2010$LC5=="NVG" ] <- "NVG"
dat2010$LC6[ dat2010$LC5=="NFR" ] <- "NFR"
dat2010$LC6[ dat2010$LC5=="MGF" ] <- "MGF"
dat2010$LC6[ dat2010$LC5=="FPC" ] <- "FPC"
dat2010$LC6[ dat2010$LCCode=="FOFC" ] <- "OFC"
dat2010$LC6[ dat2010$LC6=="OOO" ] <- "FG5"

# Level7: Node 5 Split

# 2007
dat2007$LC7 <- "OOO"
dat2007$LC7[ dat2007$LC6=="WIR" ] <- "WIR"
dat2007$LC7[ dat2007$LC6=="NVG" ] <- "NVG"
dat2007$LC7[ dat2007$LC6=="NFR" ] <- "NFR"
dat2007$LC7[ dat2007$LC6=="MGF" ] <- "MGF"
dat2007$LC7[ dat2007$LC6=="FPC" ] <- "FPC"
dat2007$LC7[ dat2007$LC6=="FG4" ] <- "FG4"
dat2007$LC7[ dat2007$LCCode=="FCFM" ] <- "CFM"
dat2007$LC7[ dat2007$LC7=="OOO" ] <- "FG3"
# 2008
dat2008$LC7 <- "OOO"
dat2008$LC7[ dat2008$LC6=="WIR" ] <- "WIR"
dat2008$LC7[ dat2008$LC6=="NVG" ] <- "NVG"
dat2008$LC7[ dat2008$LC6=="NFR" ] <- "NFR"
dat2008$LC7[ dat2008$LC6=="MGF" ] <- "MGF"
dat2008$LC7[ dat2008$LC6=="FPC" ] <- "FPC"

```

```

dat2008$LC7[ dat2008$LC6=="FG4" ] <- "FG4"
dat2008$LC7[ dat2008$LCCode=="FCFB" | dat2008$LCCode=="FOFB" ] <- "FG1"
dat2008$LC7[ dat2008$LC7=="OOO" ] <- "FG3"
# 2009
dat2009$LC7 <- "OOO"
dat2009$LC7[ dat2009$LC6=="WIR" ] <- "WIR"
dat2009$LC7[ dat2009$LC6=="NVG" ] <- "NVG"
dat2009$LC7[ dat2009$LC6=="NFR" ] <- "NFR"
dat2009$LC7[ dat2009$LC6=="MGF" ] <- "MGF"
dat2009$LC7[ dat2009$LC6=="FPC" ] <- "FPC"
dat2009$LC7[ dat2009$LC6=="FG4" ] <- "FG4"
dat2009$LC7[ dat2009$LCCode=="FOFC" ] <- "OFC"
dat2009$LC7[ dat2009$LC7=="OOO" ] <- "FG2"
# 2010
dat2010$LC7 <- "OOO"
dat2010$LC7[ dat2010$LC6=="WIR" ] <- "WIR"
dat2010$LC7[ dat2010$LC6=="NVG" ] <- "NVG"
dat2010$LC7[ dat2010$LC6=="NFR" ] <- "NFR"
dat2010$LC7[ dat2010$LC6=="MGF" ] <- "MGF"
dat2010$LC7[ dat2010$LC6=="FPC" ] <- "FPC"
dat2010$LC7[ dat2010$LC6=="OFC" ] <- "OFC"
dat2010$LC7[ dat2010$LCCode=="FOFB" | dat2010$LCCode=="FCFB" ] <- "FG2"
dat2010$LC7[ dat2010$LC7=="OOO" ] <- "FG4"

# Level8: Node 4 Split

# 2007
dat2007$LC8 <- "OOO"
dat2007$LC8[ dat2007$LC7=="WIR" ] <- "WIR"
dat2007$LC8[ dat2007$LC7=="NVG" ] <- "NVG"
dat2007$LC8[ dat2007$LC7=="NFR" ] <- "NFR"
dat2007$LC8[ dat2007$LC7=="MGF" ] <- "MGF"
dat2007$LC8[ dat2007$LC7=="FPC" ] <- "FPC"
dat2007$LC8[ dat2007$LC7=="CFM" ] <- "CFM"
dat2007$LC8[ dat2007$LC7=="FG3" ] <- "FG3"
dat2007$LC8[ dat2007$LCCode=="FCFC" ] <- "CFC"
dat2007$LC8[ dat2007$LC8=="OOO" ] <- "FG1"
# 2008
dat2008$LC8 <- "OOO"
dat2008$LC8[ dat2008$LC7=="WIR" ] <- "WIR"
dat2008$LC8[ dat2008$LC7=="NVG" ] <- "NVG"
dat2008$LC8[ dat2008$LC7=="NFR" ] <- "NFR"
dat2008$LC8[ dat2008$LC7=="MGF" ] <- "MGF"
dat2008$LC8[ dat2008$LC7=="FPC" ] <- "FPC"
dat2008$LC8[ dat2008$LC7=="FG1" ] <- "FG1"
dat2008$LC8[ dat2008$LC7=="FG3" ] <- "FG3"
dat2008$LC8[ dat2008$LCCode=="FCFC" ] <- "CFC"
dat2008$LC8[ dat2008$LCCode=="FOFC" ] <- "OFC"

```

```

# 2009
dat2009$LC8 <- "OOO"
dat2009$LC8[ dat2009$LC7=="WIR" ] <- "WIR"
dat2009$LC8[ dat2009$LC7=="NVG" ] <- "NVG"
dat2009$LC8[ dat2009$LC7=="NFR" ] <- "NFR"
dat2009$LC8[ dat2009$LC7=="MGF" ] <- "MGF"
dat2009$LC8[ dat2009$LC7=="FPC" ] <- "FPC"
dat2009$LC8[ dat2009$LC7=="OFC" ] <- "OFC"
dat2009$LC8[ dat2009$LC7=="FG2" ] <- "FG2"
dat2009$LC8[ dat2009$LCCode=="FCFB" ] <- "CFB"
dat2009$LC8[ dat2009$LC8=="OOO" ] <- "FG3"

# 2010
dat2010$LC8 <- "OOO"
dat2010$LC8[ dat2010$LC7=="WIR" ] <- "WIR"
dat2010$LC8[ dat2010$LC7=="NVG" ] <- "NVG"
dat2010$LC8[ dat2010$LC7=="NFR" ] <- "NFR"
dat2010$LC8[ dat2010$LC7=="MGF" ] <- "MGF"
dat2010$LC8[ dat2010$LC7=="FPC" ] <- "FPC"
dat2010$LC8[ dat2010$LC7=="OFC" ] <- "OFC"
dat2010$LC8[ dat2010$LC7=="FG2" ] <- "FG2"
dat2010$LC8[ dat2010$LCCode=="FFPB" ] <- "FPB"
dat2010$LC8[ dat2010$LC8=="OOO" ] <- "FG3"

```

Level9: Node 3 Split

```

# 2007
dat2007$LC9 <- "OOO"
dat2007$LC9[ dat2007$LC8=="WIR" ] <- "WIR"
dat2007$LC9[ dat2007$LC8=="NVG" ] <- "NVG"
dat2007$LC9[ dat2007$LC8=="NFR" ] <- "NFR"
dat2007$LC9[ dat2007$LC8=="MGF" ] <- "MGF"
dat2007$LC9[ dat2007$LC8=="FPC" ] <- "FPC"
dat2007$LC9[ dat2007$LC8=="CFM" ] <- "CFM"
dat2007$LC9[ dat2007$LC8=="CFC" ] <- "CFC"
dat2007$LC9[ dat2007$LC8=="FG1" ] <- "FG1"
dat2007$LC9[ dat2007$LCCode=="FCFB" ] <- "CFB"
dat2007$LC9[ dat2007$LC9=="OOO" ] <- "FG2"

# 2008
dat2008$LC9 <- "OOO"
dat2008$LC9[ dat2008$LC8=="WIR" ] <- "WIR"
dat2008$LC9[ dat2008$LC8=="NVG" ] <- "NVG"
dat2008$LC9[ dat2008$LC8=="NFR" ] <- "NFR"
dat2008$LC9[ dat2008$LC8=="MGF" ] <- "MGF"
dat2008$LC9[ dat2008$LC8=="FPC" ] <- "FPC"
dat2008$LC9[ dat2008$LC8=="CFC" ] <- "CFC"
dat2008$LC9[ dat2008$LC8=="OFC" ] <- "OFC"
dat2008$LC9[ dat2008$LC8=="FG1" ] <- "FG1"
dat2008$LC9[ dat2008$LCCode=="FCFM" ] <- "CFM"

```

```

dat2008$LC9[ dat2008$LC9=="OOO" ] <- "FG2"
# 2009
dat2009$LC9 <- "OOO"
dat2009$LC9[ dat2009$LC8=="WIR" ] <- "WIR"
dat2009$LC9[ dat2009$LC8=="NVG" ] <- "NVG"
dat2009$LC9[ dat2009$LC8=="NFR" ] <- "NFR"
dat2009$LC9[ dat2009$LC8=="MGF" ] <- "MGF"
dat2009$LC9[ dat2009$LC8=="FPC" ] <- "FPC"
dat2009$LC9[ dat2009$LC8=="OFC" ] <- "OFC"
dat2009$LC9[ dat2009$LC8=="CFB" ] <- "CFB"
dat2009$LC9[ dat2009$LC8=="FG2" ] <- "FG2"
dat2009$LC9[ dat2009$LCCode=="FFPB" ] <- "FPB"
dat2009$LC9[ dat2009$LC9=="OOO" ] <- "FG1"
# 2010
dat2010$LC9 <- "OOO"
dat2010$LC9[ dat2010$LC8=="WIR" ] <- "WIR"
dat2010$LC9[ dat2010$LC8=="NVG" ] <- "NVG"
dat2010$LC9[ dat2010$LC8=="NFR" ] <- "NFR"
dat2010$LC9[ dat2010$LC8=="MGF" ] <- "MGF"
dat2010$LC9[ dat2010$LC8=="FPC" ] <- "FPC"
dat2010$LC9[ dat2010$LC8=="OFC" ] <- "OFC"
dat2010$LC9[ dat2010$LC8=="FPB" ] <- "FPB"
dat2010$LC9[ dat2010$LC8=="FG2" ] <- "FG2"
dat2010$LC9[ dat2010$LCCode=="FCFC" ] <- "CFC"
dat2010$LC9[ dat2010$LC9=="OOO" ] <- "FG1"

# Level10: Node 2 Split

# 2007
dat2007$LC10 <- "OOO"
dat2007$LC10[ dat2007$LC9=="WIR" ] <- "WIR"
dat2007$LC10[ dat2007$LC9=="NVG" ] <- "NVG"
dat2007$LC10[ dat2007$LC9=="NFR" ] <- "NFR"
dat2007$LC10[ dat2007$LC9=="MGF" ] <- "MGF"
dat2007$LC10[ dat2007$LC9=="FPC" ] <- "FPC"
dat2007$LC10[ dat2007$LC9=="CFM" ] <- "CFM"
dat2007$LC10[ dat2007$LC9=="CFC" ] <- "CFC"
dat2007$LC10[ dat2007$LC9=="CFB" ] <- "CFB"
dat2007$LC10[ dat2007$LC9=="FG1" ] <- "FG1"
dat2007$LC10[ dat2007$LCCode=="FOFB" ] <- "OFB"
dat2007$LC10[ dat2007$LCCode=="FFPB" ] <- "FPB"
# 2008
dat2008$LC10 <- "OOO"
dat2008$LC10[ dat2008$LC9=="WIR" ] <- "WIR"
dat2008$LC10[ dat2008$LC9=="NVG" ] <- "NVG"
dat2008$LC10[ dat2008$LC9=="NFR" ] <- "NFR"
dat2008$LC10[ dat2008$LC9=="MGF" ] <- "MGF"
dat2008$LC10[ dat2008$LC9=="FPC" ] <- "FPC"

```

```

dat2008$LC10[ dat2008$LC9=="CFC" ] <- "CFC"
dat2008$LC10[ dat2008$LC9=="OFC" ] <- "OFC"
dat2008$LC10[ dat2008$LC9=="CFM" ] <- "CFM"
dat2008$LC10[ dat2008$LC9=="FG1" ] <- "FG1"
dat2008$LC10[ dat2008$LCCode=="FOFM" ] <- "OFM"
dat2008$LC10[ dat2008$LCCode=="FFPB" ] <- "FPB"
# 2009
dat2009$LC10 <- "OOO"
dat2009$LC10[ dat2009$LC9=="WIR" ] <- "WIR"
dat2009$LC10[ dat2009$LC9=="NVG" ] <- "NVG"
dat2009$LC10[ dat2009$LC9=="NFR" ] <- "NFR"
dat2009$LC10[ dat2009$LC9=="MGF" ] <- "MGF"
dat2009$LC10[ dat2009$LC9=="FPC" ] <- "FPC"
dat2009$LC10[ dat2009$LC9=="OFC" ] <- "OFC"
dat2009$LC10[ dat2009$LC9=="CFB" ] <- "CFB"
dat2009$LC10[ dat2009$LC9=="FPB" ] <- "FPB"
dat2009$LC10[ dat2009$LC9=="FG1" ] <- "FG1"
dat2009$LC10[ dat2009$LCCode=="FCFC" ] <- "CFC"
dat2009$LC10[ dat2009$LCCode=="FCFM" ] <- "CFM"
# 2010
dat2010$LC10 <- "OOO"
dat2010$LC10[ dat2010$LC9=="WIR" ] <- "WIR"
dat2010$LC10[ dat2010$LC9=="NVG" ] <- "NVG"
dat2010$LC10[ dat2010$LC9=="NFR" ] <- "NFR"
dat2010$LC10[ dat2010$LC9=="MGF" ] <- "MGF"
dat2010$LC10[ dat2010$LC9=="FPC" ] <- "FPC"
dat2010$LC10[ dat2010$LC9=="OFC" ] <- "OFC"
dat2010$LC10[ dat2010$LC9=="FPB" ] <- "FPB"
dat2010$LC10[ dat2010$LC9=="CFC" ] <- "CFC"
dat2010$LC10[ dat2010$LC9=="FG1" ] <- "FG1"
dat2010$LC10[ dat2010$LCCode=="FOFB" ] <- "OFB"
dat2010$LC10[ dat2010$LCCode=="FCFB" ] <- "CFB"

# Level11: Node 1 Split

# 2007
dat2007$LC11 <- "OOO"
dat2007$LC11[ dat2007$LC10=="WIR" ] <- "WIR"
dat2007$LC11[ dat2007$LC10=="NVG" ] <- "NVG"
dat2007$LC11[ dat2007$LC10=="NFR" ] <- "NFR"
dat2007$LC11[ dat2007$LC10=="MGF" ] <- "MGF"
dat2007$LC11[ dat2007$LC10=="FPC" ] <- "FPC"
dat2007$LC11[ dat2007$LC10=="CFM" ] <- "CFM"
dat2007$LC11[ dat2007$LC10=="CFC" ] <- "CFC"
dat2007$LC11[ dat2007$LC10=="CFB" ] <- "CFB"
dat2007$LC11[ dat2007$LC10=="FPB" ] <- "FPB"
dat2007$LC11[ dat2007$LC10=="OFB" ] <- "OFB"
dat2007$LC11[ dat2007$LCCode=="FOFC" ] <- "OFC"

```

```

dat2007$LC11[ dat2007$LCCode=="FOFM" ] <- "OFM"
# 2008
dat2008$LC11 <- "OOO"
dat2008$LC11[ dat2008$LC10=="WIR" ] <- "WIR"
dat2008$LC11[ dat2008$LC10=="NVG" ] <- "NVG"
dat2008$LC11[ dat2008$LC10=="NFR" ] <- "NFR"
dat2008$LC11[ dat2008$LC10=="MGF" ] <- "MGF"
dat2008$LC11[ dat2008$LC10=="FPC" ] <- "FPC"
dat2008$LC11[ dat2008$LC10=="CFC" ] <- "CFC"
dat2008$LC11[ dat2008$LC10=="OFC" ] <- "OFC"
dat2008$LC11[ dat2008$LC10=="CFM" ] <- "CFM"
dat2008$LC11[ dat2008$LC10=="FPB" ] <- "FPB"
dat2008$LC11[ dat2008$LC10=="OFM" ] <- "OFM"
dat2008$LC11[ dat2008$LCCode=="FOFB" ] <- "OFB"
dat2008$LC11[ dat2008$LCCode=="FCFB" ] <- "CFB"
# 2009
dat2009$LC11 <- "OOO"
dat2009$LC11[ dat2009$LC10=="WIR" ] <- "WIR"
dat2009$LC11[ dat2009$LC10=="NVG" ] <- "NVG"
dat2009$LC11[ dat2009$LC10=="NFR" ] <- "NFR"
dat2009$LC11[ dat2009$LC10=="MGF" ] <- "MGF"
dat2009$LC11[ dat2009$LC10=="FPC" ] <- "FPC"
dat2009$LC11[ dat2009$LC10=="OFC" ] <- "OFC"
dat2009$LC11[ dat2009$LC10=="CFB" ] <- "CFB"
dat2009$LC11[ dat2009$LC10=="FPB" ] <- "FPB"
dat2009$LC11[ dat2009$LC10=="CFC" ] <- "CFC"
dat2009$LC11[ dat2009$LC10=="CFM" ] <- "CFM"
dat2009$LC11[ dat2009$LCCode=="FOFB" ] <- "OFB"
dat2009$LC11[ dat2009$LCCode=="FOFM" ] <- "OFM"
# 2010
dat2010$LC11 <- "OOO"
dat2010$LC11[ dat2010$LC10=="WIR" ] <- "WIR"
dat2010$LC11[ dat2010$LC10=="NVG" ] <- "NVG"
dat2010$LC11[ dat2010$LC10=="NFR" ] <- "NFR"
dat2010$LC11[ dat2010$LC10=="MGF" ] <- "MGF"
dat2010$LC11[ dat2010$LC10=="FPC" ] <- "FPC"
dat2010$LC11[ dat2010$LC10=="OFC" ] <- "OFC"
dat2010$LC11[ dat2010$LC10=="FPB" ] <- "FPB"
dat2010$LC11[ dat2010$LC10=="CFC" ] <- "CFC"
dat2010$LC11[ dat2010$LC10=="CFB" ] <- "CFB"
dat2010$LC11[ dat2010$LC10=="OFB" ] <- "OFB"
dat2010$LC11[ dat2010$LCCode=="FCFM" ] <- "CFM"
dat2010$LC11[ dat2010$LCCode=="FOFM" ] <- "OFM"

# Select observations per classification level and store selected data
# in variables

# 2007

```

```

dat2007.lc1 <- dat2007
dat2007.lc2 <- subset(dat2007, subset=(LC1=="LND"))
dat2007.lc3 <- subset(dat2007, subset=(LC2=="VEG"))
dat2007.lc4 <- subset(dat2007, subset=(LC3=="FOR"))
dat2007.lc5 <- subset(dat2007, subset=(LC4=="FG7"))
dat2007.lc6 <- subset(dat2007, subset=(LC5=="FG6"))
dat2007.lc7 <- subset(dat2007, subset=(LC6=="FG5"))
dat2007.lc8 <- subset(dat2007, subset=(LC7=="FG4"))
dat2007.lc9 <- subset(dat2007, subset=(LC8=="FG3"))
dat2007.lc10 <- subset(dat2007, subset=(LC9=="FG2"))
dat2007.lc11 <- subset(dat2007, subset=(LC10=="FG1"))

# 2008
dat2008.lc1 <- dat2008
dat2008.lc2 <- subset(dat2008, subset=(LC1=="LND"))
dat2008.lc3 <- subset(dat2008, subset=(LC2=="VEG"))
dat2008.lc4 <- subset(dat2008, subset=(LC3=="FOR"))
dat2008.lc5 <- subset(dat2008, subset=(LC4=="FG7"))
dat2008.lc6 <- subset(dat2008, subset=(LC5=="FG6"))
dat2008.lc7 <- subset(dat2008, subset=(LC6=="FG5"))
dat2008.lc8 <- subset(dat2008, subset=(LC7=="FG4"))
dat2008.lc9 <- subset(dat2008, subset=(LC8=="FG3"))
dat2008.lc10 <- subset(dat2008, subset=(LC9=="FG2"))
dat2008.lc11 <- subset(dat2008, subset=(LC10=="FG1"))

# 2009
dat2009.lc1 <- dat2009
dat2009.lc2 <- subset(dat2009, subset=(LC1=="LND"))
dat2009.lc3 <- subset(dat2009, subset=(LC2=="VEG"))
dat2009.lc4 <- subset(dat2009, subset=(LC3=="FOR"))
dat2009.lc5 <- subset(dat2009, subset=(LC4=="FG7"))
dat2009.lc6 <- subset(dat2009, subset=(LC5=="FG6"))
dat2009.lc7 <- subset(dat2009, subset=(LC6=="FG5"))
dat2009.lc8 <- subset(dat2009, subset=(LC7=="FG4"))
dat2009.lc9 <- subset(dat2009, subset=(LC8=="FG3"))
dat2009.lc10 <- subset(dat2009, subset=(LC9=="FG2"))
dat2009.lc11 <- subset(dat2009, subset=(LC10=="FG1"))

# 2010
dat2010.lc1 <- dat2010
dat2010.lc2 <- subset(dat2010, subset=(LC1=="LND"))
dat2010.lc3 <- subset(dat2010, subset=(LC2=="VEG"))
dat2010.lc4 <- subset(dat2010, subset=(LC3=="FOR"))
dat2010.lc5 <- subset(dat2010, subset=(LC4=="FG7"))
dat2010.lc6 <- subset(dat2010, subset=(LC5=="FG6"))
dat2010.lc7 <- subset(dat2010, subset=(LC6=="FG5"))
dat2010.lc8 <- subset(dat2010, subset=(LC7=="FG4"))
dat2010.lc9 <- subset(dat2010, subset=(LC8=="FG3"))
dat2010.lc10 <- subset(dat2010, subset=(LC9=="FG2"))
dat2010.lc11 <- subset(dat2010, subset=(LC10=="FG1"))

```

```

# Execute tree function using specified classification level and
# predictor variables

# 2007
tree2007.lc1 <- tree(factor(LC1) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2007.lc1 )
tree2007.lc2 <- tree(factor(LC2) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2007.lc2 )
tree2007.lc3 <- tree(factor(LC3) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2007.lc3 )
tree2007.lc4 <- tree(factor(LC4) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2007.lc4 )
tree2007.lc5 <- tree(factor(LC5) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2007.lc5 )
tree2007.lc6 <- tree(factor(LC6) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2007.lc6 )
tree2007.lc7 <- tree(factor(LC7) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +

```

```

SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2007.1c7)
tree2007.1c8 <- tree(factor(LC8) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2007.1c8)
tree2007.1c9 <- tree(factor(LC9) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2007.1c9)
tree2007.1c10 <- tree(factor(LC10) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2007.1c10
)
tree2007.1c11 <- tree(factor(LC11) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2007.1c11
)
# 2008
tree2008.1c1 <- tree(factor(LC1) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2008.1c1)
tree2008.1c2 <- tree(factor(LC2) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +

```



```

GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDHH + GLCMSDHV + GLCMSDRAT, dat2008.1c9 )
tree2008.1c10 <- tree(factor(LC10) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDHH + GLCMSDHV + GLCMSDRAT, dat2008.1c10
)
tree2008.1c11 <- tree(factor(LC11) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDHH + GLCMSDHV + GLCMSDRAT, dat2008.1c11
)
# 2009
tree2009.1c1 <- tree(factor(LC1) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDHH + GLCMSDHV + GLCMSDRAT, dat2009.1c1 )
tree2009.1c2 <- tree(factor(LC2) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDHH + GLCMSDHV + GLCMSDRAT, dat2009.1c2 )
tree2009.1c3 <- tree(factor(LC3) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDHH + GLCMSDHV + GLCMSDRAT, dat2009.1c3 )
tree2009.1c4 <- tree(factor(LC4) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDHH + GLCMSDHV + GLCMSDRAT, dat2009.1c4 )

```

```

tree2009.lc5 <- tree(factor(LC5) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2009.lc5)
tree2009.lc6 <- tree(factor(LC6) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2009.lc6)
tree2009.lc7 <- tree(factor(LC7) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2009.lc7)
tree2009.lc8 <- tree(factor(LC8) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2009.lc8)
tree2009.lc9 <- tree(factor(LC9) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2009.lc9)
tree2009.lc10 <- tree(factor(LC10) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2009.lc10
)
tree2009.lc11 <- tree(factor(LC11) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +

```

```

GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2009.lc11
)
# 2010
tree2010.lc1 <- tree(factor(LC1) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2010.lc1)
tree2010.lc2 <- tree(factor(LC2) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2010.lc2)
tree2010.lc3 <- tree(factor(LC3) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2010.lc3)
tree2010.lc4 <- tree(factor(LC4) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2010.lc4)
tree2010.lc5 <- tree(factor(LC5) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2010.lc5)
tree2010.lc6 <- tree(factor(LC6) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2010.lc6)
tree2010.lc7 <- tree(factor(LC7) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +

```

```

SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2010.1c7)
tree2010.lc8 <- tree(factor(LC8) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2010.1c8)
tree2010.lc9 <- tree(factor(LC9) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2010.1c9)
tree2010.lc10 <- tree(factor(LC10) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2010.1c10
)
tree2010.lc11 <- tree(factor(LC11) ~ MeanHH + MeanHV + MeanRAT + SDHH +
SDHV + SDRAT + MeanAspect + MeanDEM + MeanSlope + SDAspect + SDDEM +
SDSlope + GLCMAngHH + GLCMAngHV + GLCMAngRAT + GLCMConHH +
GLCMConHV + GLCMConRAT + GLCMCorHH + GLCMCorHV + GLCMCorRAT +
GLCMDisHH + GLCMDisHV + GLCMDisRAT + GLCMEntHH + GLCMEntHV +
GLCMEntRAT + GLCMHomHH + GLCMHomHV + GLCMHomRAT + GLCMMnHH +
GLCMMnHV + GLCMMnRAT + GLCMSDH + GLCMSDHV + GLCMSDRAT, dat2010.1c11
)

# Perform k-fold cross-validation for each classification level of multi-
# year data

# 2007
cv2007.lc1 <- cv.tree(tree2007.lc1)
cv2007.lc2 <- cv.tree(tree2007.lc2)
cv2007.lc3 <- cv.tree(tree2007.lc3)
cv2007.lc4 <- cv.tree(tree2007.lc4)
cv2007.lc5 <- cv.tree(tree2007.lc5)
cv2007.lc6 <- cv.tree(tree2007.lc6)
cv2007.lc7 <- cv.tree(tree2007.lc7)
cv2007.lc8 <- cv.tree(tree2007.lc8)

```

```

cv2007.lc9  <- cv.tree(tree2007.lc9)
cv2007.lc10 <- cv.tree(tree2007.lc10)
cv2007.lc11 <- cv.tree(tree2007.lc11)
# 2008
cv2008.lc1  <- cv.tree(tree2008.lc1)
cv2008.lc2  <- cv.tree(tree2008.lc2)
cv2008.lc3  <- cv.tree(tree2008.lc3)
cv2008.lc4  <- cv.tree(tree2008.lc4)
cv2008.lc5  <- cv.tree(tree2008.lc5)
cv2008.lc6  <- cv.tree(tree2008.lc6)
cv2008.lc7  <- cv.tree(tree2008.lc7)
cv2008.lc8  <- cv.tree(tree2008.lc8)
cv2008.lc9  <- cv.tree(tree2008.lc9)
cv2008.lc10 <- cv.tree(tree2008.lc10)
cv2008.lc11 <- cv.tree(tree2008.lc11)
# 2009
cv2009.lc1  <- cv.tree(tree2009.lc1)
cv2009.lc2  <- cv.tree(tree2009.lc2)
cv2009.lc3  <- cv.tree(tree2009.lc3)
cv2009.lc4  <- cv.tree(tree2009.lc4)
cv2009.lc5  <- cv.tree(tree2009.lc5)
cv2009.lc6  <- cv.tree(tree2009.lc6)
cv2009.lc7  <- cv.tree(tree2009.lc7)
cv2009.lc8  <- cv.tree(tree2009.lc8)
cv2009.lc9  <- cv.tree(tree2009.lc9)
cv2009.lc10 <- cv.tree(tree2009.lc10)
cv2009.lc11 <- cv.tree(tree2009.lc11)
# 2010
cv2010.lc1  <- cv.tree(tree2010.lc1)
cv2010.lc2  <- cv.tree(tree2010.lc2)
cv2010.lc3  <- cv.tree(tree2010.lc3)
cv2010.lc4  <- cv.tree(tree2010.lc4)
cv2010.lc5  <- cv.tree(tree2010.lc5)
cv2010.lc6  <- cv.tree(tree2010.lc6)
cv2010.lc7  <- cv.tree(tree2010.lc7)
cv2010.lc8  <- cv.tree(tree2010.lc8)
cv2010.lc9  <- cv.tree(tree2010.lc9)
cv2010.lc10 <- cv.tree(tree2010.lc10)
cv2010.lc11 <- cv.tree(tree2010.lc11)

# Extract smallest/best deviance from k-fold cross-validation results of
# multi-year data

# 2007
best2007.lc1 <- 1 + cv2007.lc1$size [which(cv2007.lc1$dev==min(cv2007.
lc1$dev))]
best2007.lc2 <- 1 + cv2007.lc2$size [which(cv2007.lc2$dev==min(cv2007.
lc2$dev))]
```

```

best2007.lc3 <- 1 + cv2007.lc3$size [which(cv2007.lc3$dev=min(cv2007.
lc3$dev))]
best2007.lc4 <- 1 + cv2007.lc4$size [which(cv2007.lc4$dev=min(cv2007.
lc4$dev))]
best2007.lc5 <- 1 + cv2007.lc5$size [which(cv2007.lc5$dev=min(cv2007.
lc5$dev))]
best2007.lc6 <- 1 + cv2007.lc6$size [which(cv2007.lc6$dev=min(cv2007.
lc6$dev))]
best2007.lc7 <- 1 + cv2007.lc7$size [which(cv2007.lc7$dev=min(cv2007.
lc7$dev))]
best2007.lc8 <- 1 + cv2007.lc8$size [which(cv2007.lc8$dev=min(cv2007.
lc8$dev))]
best2007.lc9 <- 1 + cv2007.lc9$size [which(cv2007.lc9$dev=min(cv2007.
lc9$dev))]
best2007.lc10 <- 1 + cv2007.lc10$size [which(cv2007.lc10$dev=min(cv2007.
lc10$dev))]
best2007.lc11 <- 1 + cv2007.lc11$size [which(cv2007.lc11$dev=min(cv2007.
lc11$dev))]
# 2008
best2008.lc1 <- 1 + cv2008.lc1$size [which(cv2008.lc1$dev=min(cv2008.
lc1$dev))]
best2008.lc2 <- 1 + cv2008.lc2$size [which(cv2008.lc2$dev=min(cv2008.
lc2$dev))]
best2008.lc3 <- 1 + cv2008.lc3$size [which(cv2008.lc3$dev=min(cv2008.
lc3$dev))]
best2008.lc4 <- 1 + cv2008.lc4$size [which(cv2008.lc4$dev=min(cv2008.
lc4$dev))]
best2008.lc5 <- 1 + cv2008.lc5$size [which(cv2008.lc5$dev=min(cv2008.
lc5$dev))]
best2008.lc6 <- 1 + cv2008.lc6$size [which(cv2008.lc6$dev=min(cv2008.
lc6$dev))]
best2008.lc7 <- 1 + cv2008.lc7$size [which(cv2008.lc7$dev=min(cv2008.
lc7$dev))]
best2008.lc8 <- 1 + cv2008.lc8$size [which(cv2008.lc8$dev=min(cv2008.
lc8$dev))]
best2008.lc9 <- 1 + cv2008.lc9$size [which(cv2008.lc9$dev=min(cv2008.
lc9$dev))]
best2008.lc10 <- 1 + cv2008.lc10$size [which(cv2008.lc10$dev=min(cv2008.
lc10$dev))]
best2008.lc11 <- 1 + cv2008.lc11$size [which(cv2008.lc11$dev=min(cv2008.
lc11$dev))]
# 2009
best2009.lc1 <- 1 + cv2009.lc1$size [which(cv2009.lc1$dev=min(cv2009.
lc1$dev))]
best2009.lc2 <- 1 + cv2009.lc2$size [which(cv2009.lc2$dev=min(cv2009.
lc2$dev))]
best2009.lc3 <- 1 + cv2009.lc3$size [which(cv2009.lc3$dev=min(cv2009.
lc3$dev))]

```

```

best2009.lc4 <- 1 + cv2009.lc4$size [which(cv2009.lc4$dev=min(cv2009.
  lc4$dev))]
best2009.lc5 <- 1 + cv2009.lc5$size [which(cv2009.lc5$dev=min(cv2009.
  lc5$dev))]
best2009.lc6 <- 1 + cv2009.lc6$size [which(cv2009.lc6$dev=min(cv2009.
  lc6$dev))]
best2009.lc7 <- 1 + cv2009.lc7$size [which(cv2009.lc7$dev=min(cv2009.
  lc7$dev))]
best2009.lc8 <- 1 + cv2009.lc8$size [which(cv2009.lc8$dev=min(cv2009.
  lc8$dev))]
best2009.lc9 <- 1 + cv2009.lc9$size [which(cv2009.lc9$dev=min(cv2009.
  lc9$dev))]
best2009.lc10 <- 1 + cv2009.lc10$size [which(cv2009.lc10$dev=min(cv2009.
  lc10$dev))]
best2009.lc11 <- 1 + cv2009.lc11$size [which(cv2009.lc11$dev=min(cv2009.
  lc11$dev))]
# 2010
best2010.lc1 <- 1 + cv2010.lc1$size [which(cv2010.lc1$dev=min(cv2010.
  lc1$dev))]
best2010.lc2 <- 1 + cv2010.lc2$size [which(cv2010.lc2$dev=min(cv2010.
  lc2$dev))]
best2010.lc3 <- 1 + cv2010.lc3$size [which(cv2010.lc3$dev=min(cv2010.
  lc3$dev))]
best2010.lc4 <- 1 + cv2010.lc4$size [which(cv2010.lc4$dev=min(cv2010.
  lc4$dev))]
best2010.lc5 <- 1 + cv2010.lc5$size [which(cv2010.lc5$dev=min(cv2010.
  lc5$dev))]
best2010.lc6 <- 1 + cv2010.lc6$size [which(cv2010.lc6$dev=min(cv2010.
  lc6$dev))]
best2010.lc7 <- 1 + cv2010.lc7$size [which(cv2010.lc7$dev=min(cv2010.
  lc7$dev))]
best2010.lc8 <- 1 + cv2010.lc8$size [which(cv2010.lc8$dev=min(cv2010.
  lc8$dev))]
best2010.lc9 <- 1 + cv2010.lc9$size [which(cv2010.lc9$dev=min(cv2010.
  lc9$dev))]
best2010.lc10 <- 1 + cv2010.lc10$size [which(cv2010.lc10$dev=min(cv2010.
  lc10$dev))]
best2010.lc11 <- 1 + cv2010.lc11$size [which(cv2010.lc11$dev=min(cv2010.
  lc11$dev))]

# Prune the tree model using smallest/best deviance from cross-
# validation results
# (Note: the number of leafs will be no more than the smallest/best
# deviance)

# 2007
prune2007.lc1 <- prune.misclass(tree2007.lc1, best=best2007.lc1)
prune2007.lc2 <- prune.misclass(tree2007.lc2, best=best2007.lc2)

```

```

prune2007.lc3 <- prune.misclass(tree2007.lc3, best=best2007.lc3)
prune2007.lc4 <- prune.misclass(tree2007.lc4, best=best2007.lc4)
prune2007.lc5 <- prune.misclass(tree2007.lc5, best=best2007.lc5)
prune2007.lc6 <- prune.misclass(tree2007.lc6, best=best2007.lc6)
prune2007.lc7 <- prune.misclass(tree2007.lc7, best=best2007.lc7)
prune2007.lc8 <- prune.misclass(tree2007.lc8, best=best2007.lc8)
prune2007.lc9 <- prune.misclass(tree2007.lc9, best=best2007.lc9)
prune2007.lc10 <- prune.misclass(tree2007.lc10, best=best2007.lc10)
prune2007.lc11 <- prune.misclass(tree2007.lc11, best=best2007.lc11)
# 2008
prune2008.lc1 <- prune.misclass(tree2008.lc1, best=best2008.lc1)
prune2008.lc2 <- prune.misclass(tree2008.lc2, best=best2008.lc2)
prune2008.lc3 <- prune.misclass(tree2008.lc3, best=best2008.lc3)
prune2008.lc4 <- prune.misclass(tree2008.lc4, best=best2008.lc4)
prune2008.lc5 <- prune.misclass(tree2008.lc5, best=best2008.lc5)
prune2008.lc6 <- prune.misclass(tree2008.lc6, best=best2008.lc6)
prune2008.lc7 <- prune.misclass(tree2008.lc7, best=best2008.lc7)
prune2008.lc8 <- prune.misclass(tree2008.lc8, best=best2008.lc8)
prune2008.lc9 <- prune.misclass(tree2008.lc9, best=best2008.lc9)
prune2008.lc10 <- prune.misclass(tree2008.lc10, best=best2008.lc10)
prune2008.lc11 <- prune.misclass(tree2008.lc11, best=best2008.lc11)
# 2009
prune2009.lc1 <- prune.misclass(tree2009.lc1, best=best2009.lc1)
prune2009.lc2 <- prune.misclass(tree2009.lc2, best=best2009.lc2)
prune2009.lc3 <- prune.misclass(tree2009.lc3, best=best2009.lc3)
prune2009.lc4 <- prune.misclass(tree2009.lc4, best=best2009.lc4)
prune2009.lc5 <- prune.misclass(tree2009.lc5, best=best2009.lc5)
prune2009.lc6 <- prune.misclass(tree2009.lc6, best=best2009.lc6)
prune2009.lc7 <- prune.misclass(tree2009.lc7, best=best2009.lc7)
prune2009.lc8 <- prune.misclass(tree2009.lc8, best=best2009.lc8)
prune2009.lc9 <- prune.misclass(tree2009.lc9, best=best2009.lc9)
prune2009.lc10 <- prune.misclass(tree2009.lc10, best=best2009.lc10)
prune2009.lc11 <- prune.misclass(tree2009.lc11, best=best2009.lc11)
# 2010
prune2010.lc1 <- prune.misclass(tree2010.lc1, best=best2010.lc1)
prune2010.lc2 <- prune.misclass(tree2010.lc2, best=best2010.lc2)
prune2010.lc3 <- prune.misclass(tree2010.lc3, best=best2010.lc3)
prune2010.lc4 <- prune.misclass(tree2010.lc4, best=best2010.lc4)
prune2010.lc5 <- prune.misclass(tree2010.lc5, best=best2010.lc5)
prune2010.lc6 <- prune.misclass(tree2010.lc6, best=best2010.lc6)
prune2010.lc7 <- prune.misclass(tree2010.lc7, best=best2010.lc7)
prune2010.lc8 <- prune.misclass(tree2010.lc8, best=best2010.lc8)
prune2010.lc9 <- prune.misclass(tree2010.lc9, best=best2010.lc9)
prune2010.lc10 <- prune.misclass(tree2010.lc10, best=best2010.lc10)
prune2010.lc11 <- prune.misclass(tree2010.lc11, best=best2010.lc11)

# Plot cross-validation graph and save as PDF file

```

```

# 2007
pdf("output-2007-cv-lc1.pdf", width=7, height=5.5)
plot(cv.tree(tree2007.lc1))
dev.off()
pdf("output-2007-cv-lc2.pdf", width=7, height=5.5)
plot(cv.tree(tree2007.lc2))
dev.off()
pdf("output-2007-cv-lc3.pdf", width=7, height=5.5)
plot(cv.tree(tree2007.lc3))
dev.off()
pdf("output-2007-cv-lc4.pdf", width=7, height=5.5)
plot(cv.tree(tree2007.lc4))
dev.off()
pdf("output-2007-cv-lc5.pdf", width=7, height=5.5)
plot(cv.tree(tree2007.lc5))
dev.off()
pdf("output-2007-cv-lc6.pdf", width=7, height=5.5)
plot(cv.tree(tree2007.lc6))
dev.off()
pdf("output-2007-cv-lc7.pdf", width=7, height=5.5)
plot(cv.tree(tree2007.lc7))
dev.off()
pdf("output-2007-cv-lc8.pdf", width=7, height=5.5)
plot(cv.tree(tree2007.lc8))
dev.off()
pdf("output-2007-cv-lc9.pdf", width=7, height=5.5)
plot(cv.tree(tree2007.lc9))
dev.off()
pdf("output-2007-cv-lc10.pdf", width=7, height=5.5)
plot(cv.tree(tree2007.lc10))
dev.off()
pdf("output-2007-cv-lc11.pdf", width=7, height=5.5)
plot(cv.tree(tree2007.lc11))
dev.off()
# 2008
pdf("output-2008-cv-lc1.pdf", width=7, height=5.5)
plot(cv.tree(tree2008.lc1))
dev.off()
pdf("output-2008-cv-lc2.pdf", width=7, height=5.5)
plot(cv.tree(tree2008.lc2))
dev.off()
pdf("output-2008-cv-lc3.pdf", width=7, height=5.5)
plot(cv.tree(tree2008.lc3))
dev.off()
pdf("output-2008-cv-lc4.pdf", width=7, height=5.5)
plot(cv.tree(tree2008.lc4))
dev.off()
pdf("output-2008-cv-lc5.pdf", width=7, height=5.5)

```

```

plot(cv.tree(tree2008.lc5))
dev.off()
pdf("output-2008-cv-lc6.pdf", width=7, height=5.5)
plot(cv.tree(tree2008.lc6))
dev.off()
pdf("output-2008-cv-lc7.pdf", width=7, height=5.5)
plot(cv.tree(tree2008.lc7))
dev.off()
pdf("output-2008-cv-lc8.pdf", width=7, height=5.5)
plot(cv.tree(tree2008.lc8))
dev.off()
pdf("output-2008-cv-lc9.pdf", width=7, height=5.5)
plot(cv.tree(tree2008.lc9))
dev.off()
pdf("output-2008-cv-lc10.pdf", width=7, height=5.5)
plot(cv.tree(tree2008.lc10))
dev.off()
pdf("output-2008-cv-lc11.pdf", width=7, height=5.5)
plot(cv.tree(tree2008.lc11))
dev.off()
# 2009
pdf("output-2009-cv-lc1.pdf", width=7, height=5.5)
plot(cv.tree(tree2009.lc1))
dev.off()
pdf("output-2009-cv-lc2.pdf", width=7, height=5.5)
plot(cv.tree(tree2009.lc2))
dev.off()
pdf("output-2009-cv-lc3.pdf", width=7, height=5.5)
plot(cv.tree(tree2009.lc3))
dev.off()
pdf("output-2009-cv-lc4.pdf", width=7, height=5.5)
plot(cv.tree(tree2009.lc4))
dev.off()
pdf("output-2009-cv-lc5.pdf", width=7, height=5.5)
plot(cv.tree(tree2009.lc5))
dev.off()
pdf("output-2009-cv-lc6.pdf", width=7, height=5.5)
plot(cv.tree(tree2009.lc6))
dev.off()
pdf("output-2009-cv-lc7.pdf", width=7, height=5.5)
plot(cv.tree(tree2009.lc7))
dev.off()
pdf("output-2009-cv-lc8.pdf", width=7, height=5.5)
plot(cv.tree(tree2009.lc8))
dev.off()
pdf("output-2009-cv-lc9.pdf", width=7, height=5.5)
plot(cv.tree(tree2009.lc9))
dev.off()

```

```

pdf("output-2009-cv-lc10.pdf", width=7, height=5.5)
plot(cv.tree(tree2009.lc10))
dev.off()
pdf("output-2009-cv-lc11.pdf", width=7, height=5.5)
plot(cv.tree(tree2009.lc11))
dev.off()
# 2010
pdf("output-2010-cv-lc1.pdf", width=7, height=5.5)
plot(cv.tree(tree2010.lc1))
dev.off()
pdf("output-2010-cv-lc2.pdf", width=7, height=5.5)
plot(cv.tree(tree2010.lc2))
dev.off()
pdf("output-2010-cv-lc3.pdf", width=7, height=5.5)
plot(cv.tree(tree2010.lc3))
dev.off()
pdf("output-2010-cv-lc4.pdf", width=7, height=5.5)
plot(cv.tree(tree2010.lc4))
dev.off()
pdf("output-2010-cv-lc5.pdf", width=7, height=5.5)
plot(cv.tree(tree2010.lc5))
dev.off()
pdf("output-2010-cv-lc6.pdf", width=7, height=5.5)
plot(cv.tree(tree2010.lc6))
dev.off()
pdf("output-2010-cv-lc7.pdf", width=7, height=5.5)
plot(cv.tree(tree2010.lc7))
dev.off()
pdf("output-2010-cv-lc8.pdf", width=7, height=5.5)
plot(cv.tree(tree2010.lc8))
dev.off()
pdf("output-2010-cv-lc9.pdf", width=7, height=5.5)
plot(cv.tree(tree2010.lc9))
dev.off()
pdf("output-2010-cv-lc10.pdf", width=7, height=5.5)
plot(cv.tree(tree2010.lc10))
dev.off()
pdf("output-2010-cv-lc11.pdf", width=7, height=5.5)
plot(cv.tree(tree2010.lc11))
dev.off()

# Generate tree summary and misclassification statistics (overfitted)
# and save as TXT file

# 2007
sink("output-2007-tree-overfit-summary-lc1.txt", append=FALSE, split=
TRUE)
print(tree2007.lc1)

```

```

print(summary(tree2007.lc1))
print(misclass.tree(tree2007.lc1, detail=TRUE))
sink()
sink("output-2007-tree-overfit-summary-lc2.txt", append=FALSE, split=
    TRUE)
print(tree2007.lc2)
print(summary(tree2007.lc2))
print(misclass.tree(tree2007.lc2, detail=TRUE))
sink()
sink("output-2007-tree-overfit-summary-lc3.txt", append=FALSE, split=
    TRUE)
print(tree2007.lc3)
print(summary(tree2007.lc3))
print(misclass.tree(tree2007.lc3, detail=TRUE))
sink()
sink("output-2007-tree-overfit-summary-lc4.txt", append=FALSE, split=
    TRUE)
print(tree2007.lc4)
print(summary(tree2007.lc4))
print(misclass.tree(tree2007.lc4, detail=TRUE))
sink()
sink("output-2007-tree-overfit-summary-lc5.txt", append=FALSE, split=
    TRUE)
print(tree2007.lc5)
print(summary(tree2007.lc5))
print(misclass.tree(tree2007.lc5, detail=TRUE))
sink()
sink("output-2007-tree-overfit-summary-lc6.txt", append=FALSE, split=
    TRUE)
print(tree2007.lc6)
print(summary(tree2007.lc6))
print(misclass.tree(tree2007.lc6, detail=TRUE))
sink()
sink("output-2007-tree-overfit-summary-lc7.txt", append=FALSE, split=
    TRUE)
print(tree2007.lc7)
print(summary(tree2007.lc7))
print(misclass.tree(tree2007.lc7, detail=TRUE))
sink()
sink("output-2007-tree-overfit-summary-lc8.txt", append=FALSE, split=
    TRUE)
print(tree2007.lc8)
print(summary(tree2007.lc8))
print(misclass.tree(tree2007.lc8, detail=TRUE))
sink()
sink("output-2007-tree-overfit-summary-lc9.txt", append=FALSE, split=
    TRUE)
print(tree2007.lc9)

```

```

print(summary(tree2007.lc9))
print(misclass.tree(tree2007.lc9, detail=TRUE))
sink()
sink("output-2007-tree-overfit-summary-lc10.txt", append=FALSE, split=
    TRUE)
print(tree2007.lc10)
print(summary(tree2007.lc10))
print(misclass.tree(tree2007.lc10, detail=TRUE))
sink()
sink("output-2007-tree-overfit-summary-lc11.txt", append=FALSE, split=
    TRUE)
print(tree2007.lc11)
print(summary(tree2007.lc11))
print(misclass.tree(tree2007.lc11, detail=TRUE))
sink()
# 2008
sink("output-2008-tree-overfit-summary-lc1.txt", append=FALSE, split=
    TRUE)
print(tree2008.lc1)
print(summary(tree2008.lc1))
print(misclass.tree(tree2008.lc1, detail=TRUE))
sink()
sink("output-2008-tree-overfit-summary-lc2.txt", append=FALSE, split=
    TRUE)
print(tree2008.lc2)
print(summary(tree2008.lc2))
print(misclass.tree(tree2008.lc2, detail=TRUE))
sink()
sink("output-2008-tree-overfit-summary-lc3.txt", append=FALSE, split=
    TRUE)
print(tree2008.lc3)
print(summary(tree2008.lc3))
print(misclass.tree(tree2008.lc3, detail=TRUE))
sink()
sink("output-2008-tree-overfit-summary-lc4.txt", append=FALSE, split=
    TRUE)
print(tree2008.lc4)
print(summary(tree2008.lc4))
print(misclass.tree(tree2008.lc4, detail=TRUE))
sink()
sink("output-2008-tree-overfit-summary-lc5.txt", append=FALSE, split=
    TRUE)
print(tree2008.lc5)
print(summary(tree2008.lc5))
print(misclass.tree(tree2008.lc5, detail=TRUE))
sink()
sink("output-2008-tree-overfit-summary-lc6.txt", append=FALSE, split=
    TRUE)

```

```

print(tree2008.lc6)
print(summary(tree2008.lc6))
print(misclass.tree(tree2008.lc6, detail=TRUE))
sink()
sink("output-2008-tree-overfit-summary-lc7.txt", append=FALSE, split=
    TRUE)
print(tree2008.lc7)
print(summary(tree2008.lc7))
print(misclass.tree(tree2008.lc7, detail=TRUE))
sink()
sink("output-2008-tree-overfit-summary-lc8.txt", append=FALSE, split=
    TRUE)
print(tree2008.lc8)
print(summary(tree2008.lc8))
print(misclass.tree(tree2008.lc8, detail=TRUE))
sink()
sink("output-2008-tree-overfit-summary-lc9.txt", append=FALSE, split=
    TRUE)
print(tree2008.lc9)
print(summary(tree2008.lc9))
print(misclass.tree(tree2008.lc9, detail=TRUE))
sink()
sink("output-2008-tree-overfit-summary-lc10.txt", append=FALSE, split=
    TRUE)
print(tree2008.lc10)
print(summary(tree2008.lc10))
print(misclass.tree(tree2008.lc10, detail=TRUE))
sink()
sink("output-2008-tree-overfit-summary-lc11.txt", append=FALSE, split=
    TRUE)
print(tree2008.lc11)
print(summary(tree2008.lc11))
print(misclass.tree(tree2008.lc11, detail=TRUE))
sink()
# 2009
sink("output-2009-tree-overfit-summary-lc1.txt", append=FALSE, split=
    TRUE)
print(tree2009.lc1)
print(summary(tree2009.lc1))
print(misclass.tree(tree2009.lc1, detail=TRUE))
sink()
sink("output-2009-tree-overfit-summary-lc2.txt", append=FALSE, split=
    TRUE)
print(tree2009.lc2)
print(summary(tree2009.lc2))
print(misclass.tree(tree2009.lc2, detail=TRUE))
sink()
sink("output-2009-tree-overfit-summary-lc3.txt", append=FALSE, split=
    TRUE)

```

```

TRUE)
print(tree2009.lc3)
print(summary(tree2009.lc3))
print(misclass.tree(tree2009.lc3, detail=TRUE))
sink()
sink("output-2009-tree-overfit-summary-lc4.txt", append=FALSE, split=
TRUE)
print(tree2009.lc4)
print(summary(tree2009.lc4))
print(misclass.tree(tree2009.lc4, detail=TRUE))
sink()
sink("output-2009-tree-overfit-summary-lc5.txt", append=FALSE, split=
TRUE)
print(tree2009.lc5)
print(summary(tree2009.lc5))
print(misclass.tree(tree2009.lc5, detail=TRUE))
sink()
sink("output-2009-tree-overfit-summary-lc6.txt", append=FALSE, split=
TRUE)
print(tree2009.lc6)
print(summary(tree2009.lc6))
print(misclass.tree(tree2009.lc6, detail=TRUE))
sink()
sink("output-2009-tree-overfit-summary-lc7.txt", append=FALSE, split=
TRUE)
print(tree2009.lc7)
print(summary(tree2009.lc7))
print(misclass.tree(tree2009.lc7, detail=TRUE))
sink()
sink("output-2009-tree-overfit-summary-lc8.txt", append=FALSE, split=
TRUE)
print(tree2009.lc8)
print(summary(tree2009.lc8))
print(misclass.tree(tree2009.lc8, detail=TRUE))
sink()
sink("output-2009-tree-overfit-summary-lc9.txt", append=FALSE, split=
TRUE)
print(tree2009.lc9)
print(summary(tree2009.lc9))
print(misclass.tree(tree2009.lc9, detail=TRUE))
sink()
sink("output-2009-tree-overfit-summary-lc10.txt", append=FALSE, split=
TRUE)
print(tree2009.lc10)
print(summary(tree2009.lc10))
print(misclass.tree(tree2009.lc10, detail=TRUE))
sink()
sink("output-2009-tree-overfit-summary-lc11.txt", append=FALSE, split=

```

```

TRUE)
print(tree2009.lc11)
print(summary(tree2009.lc11))
print(misclass.tree(tree2009.lc11, detail=TRUE))
sink()
# 2010
sink("output-2010-tree-overfit-summary-lc1.txt", append=FALSE, split=
TRUE)
print(tree2010.lc1)
print(summary(tree2010.lc1))
print(misclass.tree(tree2010.lc1, detail=TRUE))
sink()
sink("output-2010-tree-overfit-summary-lc2.txt", append=FALSE, split=
TRUE)
print(tree2010.lc2)
print(summary(tree2010.lc2))
print(misclass.tree(tree2010.lc2, detail=TRUE))
sink()
sink("output-2010-tree-overfit-summary-lc3.txt", append=FALSE, split=
TRUE)
print(tree2010.lc3)
print(summary(tree2010.lc3))
print(misclass.tree(tree2010.lc3, detail=TRUE))
sink()
sink("output-2010-tree-overfit-summary-lc4.txt", append=FALSE, split=
TRUE)
print(tree2010.lc4)
print(summary(tree2010.lc4))
print(misclass.tree(tree2010.lc4, detail=TRUE))
sink()
sink("output-2010-tree-overfit-summary-lc5.txt", append=FALSE, split=
TRUE)
print(tree2010.lc5)
print(summary(tree2010.lc5))
print(misclass.tree(tree2010.lc5, detail=TRUE))
sink()
sink("output-2010-tree-overfit-summary-lc6.txt", append=FALSE, split=
TRUE)
print(tree2010.lc6)
print(summary(tree2010.lc6))
print(misclass.tree(tree2010.lc6, detail=TRUE))
sink()
sink("output-2010-tree-overfit-summary-lc7.txt", append=FALSE, split=
TRUE)
print(tree2010.lc7)
print(summary(tree2010.lc7))
print(misclass.tree(tree2010.lc7, detail=TRUE))
sink()

```

```

sink("output-2010-tree-overfit-summary-lc8.txt", append=FALSE, split=
TRUE)
print(tree2010.lc8)
print(summary(tree2010.lc8))
print(misclass.tree(tree2010.lc8, detail=TRUE))
sink()
sink("output-2010-tree-overfit-summary-lc9.txt", append=FALSE, split=
TRUE)
print(tree2010.lc9)
print(summary(tree2010.lc9))
print(misclass.tree(tree2010.lc9, detail=TRUE))
sink()
sink("output-2010-tree-overfit-summary-lc10.txt", append=FALSE, split=
TRUE)
print(tree2010.lc10)
print(summary(tree2010.lc10))
print(misclass.tree(tree2010.lc10, detail=TRUE))
sink()
sink("output-2010-tree-overfit-summary-lc11.txt", append=FALSE, split=
TRUE)
print(tree2010.lc11)
print(summary(tree2010.lc11))
print(misclass.tree(tree2010.lc11, detail=TRUE))
sink()

# Generate tree summary and misclassification statistics (pruned) and
# save as TXT file

# 2007
sink("output-2007-tree-pruned-summary-lc1.txt", append=FALSE, split=TRUE
)
print(prune2007.lc1)
print(summary(prune2007.lc1))
print(misclass.tree(prune2007.lc1, detail=TRUE))
sink()
sink("output-2007-tree-pruned-summary-lc2.txt", append=FALSE, split=TRUE
)
print(prune2007.lc2)
print(summary(prune2007.lc2))
print(misclass.tree(prune2007.lc2, detail=TRUE))
sink()
sink("output-2007-tree-pruned-summary-lc3.txt", append=FALSE, split=TRUE
)
print(prune2007.lc3)
print(summary(prune2007.lc3))
print(misclass.tree(prune2007.lc3, detail=TRUE))
sink()
sink("output-2007-tree-pruned-summary-lc4.txt", append=FALSE, split=TRUE
)

```

```

)
print(prune2007.lc4)
print(summary(prune2007.lc4))
print(misclass.tree(prune2007.lc4, detail=TRUE))
sink()
sink("output-2007-tree-pruned-summary-lc5.txt", append=FALSE, split=TRUE
)
print(prune2007.lc5)
print(summary(prune2007.lc5))
print(misclass.tree(prune2007.lc5, detail=TRUE))
sink()
sink("output-2007-tree-pruned-summary-lc6.txt", append=FALSE, split=TRUE
)
print(prune2007.lc6)
print(summary(prune2007.lc6))
print(misclass.tree(prune2007.lc6, detail=TRUE))
sink()
sink("output-2007-tree-pruned-summary-lc7.txt", append=FALSE, split=TRUE
)
print(prune2007.lc7)
print(summary(prune2007.lc7))
print(misclass.tree(prune2007.lc7, detail=TRUE))
sink()
sink("output-2007-tree-pruned-summary-lc8.txt", append=FALSE, split=TRUE
)
print(prune2007.lc8)
print(summary(prune2007.lc8))
print(misclass.tree(prune2007.lc8, detail=TRUE))
sink()
sink("output-2007-tree-pruned-summary-lc9.txt", append=FALSE, split=TRUE
)
print(prune2007.lc9)
print(summary(prune2007.lc9))
print(misclass.tree(prune2007.lc9, detail=TRUE))
sink()
sink("output-2007-tree-pruned-summary-lc10.txt", append=FALSE, split=
TRUE)
print(prune2007.lc10)
print(summary(prune2007.lc10))
print(misclass.tree(prune2007.lc10, detail=TRUE))
sink()
sink("output-2007-tree-pruned-summary-lc11.txt", append=FALSE, split=
TRUE)
print(prune2007.lc11)
print(summary(prune2007.lc11))
print(misclass.tree(prune2007.lc11, detail=TRUE))
sink()
# 2008

```

```

sink( "output-2008-tree-pruned-summary-lc1.txt" , append=FALSE, split=TRUE
      )
print(prune2008.lc1)
print(summary(prune2008.lc1))
print(misclass.tree(prune2008.lc1, detail=TRUE))
sink()
sink( "output-2008-tree-pruned-summary-lc2.txt" , append=FALSE, split=TRUE
      )
print(prune2008.lc2)
print(summary(prune2008.lc2))
print(misclass.tree(prune2008.lc2, detail=TRUE))
sink()
sink( "output-2008-tree-pruned-summary-lc3.txt" , append=FALSE, split=TRUE
      )
print(prune2008.lc3)
print(summary(prune2008.lc3))
print(misclass.tree(prune2008.lc3, detail=TRUE))
sink()
sink( "output-2008-tree-pruned-summary-lc4.txt" , append=FALSE, split=TRUE
      )
print(prune2008.lc4)
print(summary(prune2008.lc4))
print(misclass.tree(prune2008.lc4, detail=TRUE))
sink()
sink( "output-2008-tree-pruned-summary-lc5.txt" , append=FALSE, split=TRUE
      )
print(prune2008.lc5)
print(summary(prune2008.lc5))
print(misclass.tree(prune2008.lc5, detail=TRUE))
sink()
sink( "output-2008-tree-pruned-summary-lc6.txt" , append=FALSE, split=TRUE
      )
print(prune2008.lc6)
print(summary(prune2008.lc6))
print(misclass.tree(prune2008.lc6, detail=TRUE))
sink()
sink( "output-2008-tree-pruned-summary-lc7.txt" , append=FALSE, split=TRUE
      )
print(prune2008.lc7)
print(summary(prune2008.lc7))
print(misclass.tree(prune2008.lc7, detail=TRUE))
sink()
sink( "output-2008-tree-pruned-summary-lc8.txt" , append=FALSE, split=TRUE
      )
print(prune2008.lc8)
print(summary(prune2008.lc8))
print(misclass.tree(prune2008.lc8, detail=TRUE))
sink()

```

```

sink( "output-2008-tree-pruned-summary-lc9.txt" , append=FALSE, split=TRUE
      )
print(prune2008.lc9)
print(summary(prune2008.lc9))
print(misclass.tree(prune2008.lc9, detail=TRUE))
sink()
sink( "output-2008-tree-pruned-summary-lc10.txt" , append=FALSE, split=
      TRUE)
print(prune2008.lc10)
print(summary(prune2008.lc10))
print(misclass.tree(prune2008.lc10, detail=TRUE))
sink()
sink( "output-2008-tree-pruned-summary-lc11.txt" , append=FALSE, split=
      TRUE)
print(prune2008.lc11)
print(summary(prune2008.lc11))
print(misclass.tree(prune2008.lc11, detail=TRUE))
sink()
# 2009
sink( "output-2009-tree-pruned-summary-lc1.txt" , append=FALSE, split=TRUE
      )
print(prune2009.lc1)
print(summary(prune2009.lc1))
print(misclass.tree(prune2009.lc1, detail=TRUE))
sink()
sink( "output-2009-tree-pruned-summary-lc2.txt" , append=FALSE, split=TRUE
      )
print(prune2009.lc2)
print(summary(prune2009.lc2))
print(misclass.tree(prune2009.lc2, detail=TRUE))
sink()
sink( "output-2009-tree-pruned-summary-lc3.txt" , append=FALSE, split=TRUE
      )
print(prune2009.lc3)
print(summary(prune2009.lc3))
print(misclass.tree(prune2009.lc3, detail=TRUE))
sink()
sink( "output-2009-tree-pruned-summary-lc4.txt" , append=FALSE, split=TRUE
      )
print(prune2009.lc4)
print(summary(prune2009.lc4))
print(misclass.tree(prune2009.lc4, detail=TRUE))
sink()
sink( "output-2009-tree-pruned-summary-lc5.txt" , append=FALSE, split=TRUE
      )
print(prune2009.lc5)
print(summary(prune2009.lc5))
print(misclass.tree(prune2009.lc5, detail=TRUE))

```

```

sink()
sink( "output-2009-tree-pruned-summary-lc6.txt" , append=FALSE, split=TRUE
      )
print(prune2009.lc6)
print(summary(prune2009.lc6))
print(misclass.tree(prune2009.lc6, detail=TRUE) )
sink()
sink( "output-2009-tree-pruned-summary-lc7.txt" , append=FALSE, split=TRUE
      )
print(prune2009.lc7)
print(summary(prune2009.lc7))
print(misclass.tree(prune2009.lc7, detail=TRUE) )
sink()
sink( "output-2009-tree-pruned-summary-lc8.txt" , append=FALSE, split=TRUE
      )
print(prune2009.lc8)
print(summary(prune2009.lc8))
print(misclass.tree(prune2009.lc8, detail=TRUE) )
sink()
sink( "output-2009-tree-pruned-summary-lc9.txt" , append=FALSE, split=TRUE
      )
print(prune2009.lc9)
print(summary(prune2009.lc9))
print(misclass.tree(prune2009.lc9, detail=TRUE) )
sink()
sink( "output-2009-tree-pruned-summary-lc10.txt" , append=FALSE, split=
      TRUE)
print(prune2009.lc10)
print(summary(prune2009.lc10))
print(misclass.tree(prune2009.lc10, detail=TRUE) )
sink()
sink( "output-2009-tree-pruned-summary-lc11.txt" , append=FALSE, split=
      TRUE)
print(prune2009.lc11)
print(summary(prune2009.lc11))
print(misclass.tree(prune2009.lc11, detail=TRUE) )
sink()
# 2010
sink( "output-2010-tree-pruned-summary-lc1.txt" , append=FALSE, split=TRUE
      )
print(prune2010.lc1)
print(summary(prune2010.lc1))
print(misclass.tree(prune2010.lc1, detail=TRUE) )
sink()
sink( "output-2010-tree-pruned-summary-lc2.txt" , append=FALSE, split=TRUE
      )
print(prune2010.lc2)
print(summary(prune2010.lc2))

```

```

print(misclass.tree(prune2010.lc2, detail=TRUE))
sink()
sink("output-2010-tree-pruned-summary-lc3.txt", append=FALSE, split=TRUE)
)
print(prune2010.lc3)
print(summary(prune2010.lc3))
print(misclass.tree(prune2010.lc3, detail=TRUE))
sink()
sink("output-2010-tree-pruned-summary-lc4.txt", append=FALSE, split=TRUE)
)
print(prune2010.lc4)
print(summary(prune2010.lc4))
print(misclass.tree(prune2010.lc4, detail=TRUE))
sink()
sink("output-2010-tree-pruned-summary-lc5.txt", append=FALSE, split=TRUE)
)
print(prune2010.lc5)
print(summary(prune2010.lc5))
print(misclass.tree(prune2010.lc5, detail=TRUE))
sink()
sink("output-2010-tree-pruned-summary-lc6.txt", append=FALSE, split=TRUE)
)
print(prune2010.lc6)
print(summary(prune2010.lc6))
print(misclass.tree(prune2010.lc6, detail=TRUE))
sink()
sink("output-2010-tree-pruned-summary-lc7.txt", append=FALSE, split=TRUE)
)
print(prune2010.lc7)
print(summary(prune2010.lc7))
print(misclass.tree(prune2010.lc7, detail=TRUE))
sink()
sink("output-2010-tree-pruned-summary-lc8.txt", append=FALSE, split=TRUE)
)
print(prune2010.lc8)
print(summary(prune2010.lc8))
print(misclass.tree(prune2010.lc8, detail=TRUE))
sink()
sink("output-2010-tree-pruned-summary-lc9.txt", append=FALSE, split=TRUE)
)
print(prune2010.lc9)
print(summary(prune2010.lc9))
print(misclass.tree(prune2010.lc9, detail=TRUE))
sink()
sink("output-2010-tree-pruned-summary-lc10.txt", append=FALSE, split=
TRUE)
print(prune2010.lc10)
print(summary(prune2010.lc10))

```

```

print(misclass.tree(prune2010.lc10, detail=TRUE))
sink()
sink("output-2010-tree-pruned-summary-lc11.txt", append=FALSE, split=
    TRUE)
print(prune2010.lc11)
print(summary(prune2010.lc11))
print(misclass.tree(prune2010.lc11, detail=TRUE))
sink()

# Plot tree object/dendrogram (overfitted) and save as PDF file

# 2007
pdf("output-2007-overfit-dendrogram-lc1.pdf", width=7, height=5.5)
plot(tree2007.lc1)
text(tree2007.lc1, cex=0.50)
dev.off()
pdf("output-2007-overfit-dendrogram-lc2.pdf", width=7, height=5.5)
plot(tree2007.lc2)
text(tree2007.lc2, cex=0.50)
dev.off()
pdf("output-2007-overfit-dendrogram-lc3.pdf", width=7, height=5.5)
plot(tree2007.lc3)
text(tree2007.lc3, cex=0.50)
dev.off()
pdf("output-2007-overfit-dendrogram-lc4.pdf", width=7, height=5.5)
plot(tree2007.lc4)
text(tree2007.lc4, cex=0.50)
dev.off()
pdf("output-2007-overfit-dendrogram-lc5.pdf", width=7, height=5.5)
plot(tree2007.lc5)
text(tree2007.lc5, cex=0.50)
dev.off()
pdf("output-2007-overfit-dendrogram-lc6.pdf", width=7, height=5.5)
plot(tree2007.lc6)
text(tree2007.lc6, cex=0.50)
dev.off()
pdf("output-2007-overfit-dendrogram-lc7.pdf", width=7, height=5.5)
plot(tree2007.lc7)
text(tree2007.lc7, cex=0.50)
dev.off()
pdf("output-2007-overfit-dendrogram-lc8.pdf", width=7, height=5.5)
plot(tree2007.lc8)
text(tree2007.lc8, cex=0.50)
dev.off()
pdf("output-2007-overfit-dendrogram-lc9.pdf", width=7, height=5.5)
plot(tree2007.lc9)
text(tree2007.lc9, cex=0.50)
dev.off()

```

```

pdf("output-2007-overfit-dendrogram-lc10.pdf", width=7, height=5.5)
plot(tree2007.lc10)
text(tree2007.lc10, cex=0.50)
dev.off()
pdf("output-2007-overfit-dendrogram-lc11.pdf", width=7, height=5.5)
plot(tree2007.lc11)
text(tree2007.lc11, cex=0.50)
dev.off()
# 2008
pdf("output-2008-overfit-dendrogram-lc1.pdf", width=7, height=5.5)
plot(tree2008.lc1)
text(tree2008.lc1, cex=0.50)
dev.off()
pdf("output-2008-overfit-dendrogram-lc2.pdf", width=7, height=5.5)
plot(tree2008.lc2)
text(tree2008.lc2, cex=0.50)
dev.off()
pdf("output-2008-overfit-dendrogram-lc3.pdf", width=7, height=5.5)
plot(tree2008.lc3)
text(tree2008.lc3, cex=0.50)
dev.off()
pdf("output-2008-overfit-dendrogram-lc4.pdf", width=7, height=5.5)
plot(tree2008.lc4)
text(tree2008.lc4, cex=0.50)
dev.off()
pdf("output-2008-overfit-dendrogram-lc5.pdf", width=7, height=5.5)
plot(tree2008.lc5)
text(tree2008.lc5, cex=0.50)
dev.off()
pdf("output-2008-overfit-dendrogram-lc6.pdf", width=7, height=5.5)
plot(tree2008.lc6)
text(tree2008.lc6, cex=0.50)
dev.off()
pdf("output-2008-overfit-dendrogram-lc7.pdf", width=7, height=5.5)
plot(tree2008.lc7)
text(tree2008.lc7, cex=0.50)
dev.off()
pdf("output-2008-overfit-dendrogram-lc8.pdf", width=7, height=5.5)
plot(tree2008.lc8)
text(tree2008.lc8, cex=0.50)
dev.off()
pdf("output-2008-overfit-dendrogram-lc9.pdf", width=7, height=5.5)
plot(tree2008.lc9)
text(tree2008.lc9, cex=0.50)
dev.off()
pdf("output-2008-overfit-dendrogram-lc10.pdf", width=7, height=5.5)
plot(tree2008.lc10)
text(tree2008.lc10, cex=0.50)

```

```

dev.off()
pdf("output-2008-overfit-dendrogram-lc11.pdf", width=7, height=5.5)
plot(tree2008.lc11)
text(tree2008.lc11, cex=0.50)
dev.off()
# 2009
pdf("output-2009-overfit-dendrogram-lc1.pdf", width=7, height=5.5)
plot(tree2009.lc1)
text(tree2009.lc1, cex=0.50)
dev.off()
pdf("output-2009-overfit-dendrogram-lc2.pdf", width=7, height=5.5)
plot(tree2009.lc2)
text(tree2009.lc2, cex=0.50)
dev.off()
pdf("output-2009-overfit-dendrogram-lc3.pdf", width=7, height=5.5)
plot(tree2009.lc3)
text(tree2009.lc3, cex=0.50)
dev.off()
pdf("output-2009-overfit-dendrogram-lc4.pdf", width=7, height=5.5)
plot(tree2009.lc4)
text(tree2009.lc4, cex=0.50)
dev.off()
pdf("output-2009-overfit-dendrogram-lc5.pdf", width=7, height=5.5)
plot(tree2009.lc5)
text(tree2009.lc5, cex=0.50)
dev.off()
pdf("output-2009-overfit-dendrogram-lc6.pdf", width=7, height=5.5)
plot(tree2009.lc6)
text(tree2009.lc6, cex=0.50)
dev.off()
pdf("output-2009-overfit-dendrogram-lc7.pdf", width=7, height=5.5)
plot(tree2009.lc7)
text(tree2009.lc7, cex=0.50)
dev.off()
pdf("output-2009-overfit-dendrogram-lc8.pdf", width=7, height=5.5)
plot(tree2009.lc8)
text(tree2009.lc8, cex=0.50)
dev.off()
pdf("output-2009-overfit-dendrogram-lc9.pdf", width=7, height=5.5)
plot(tree2009.lc9)
text(tree2009.lc9, cex=0.50)
dev.off()
pdf("output-2009-overfit-dendrogram-lc10.pdf", width=7, height=5.5)
plot(tree2009.lc10)
text(tree2009.lc10, cex=0.50)
dev.off()
pdf("output-2009-overfit-dendrogram-lc11.pdf", width=7, height=5.5)
plot(tree2009.lc11)

```

```

text(tree2009.lc11, cex=0.50)
dev.off()
# 2010
pdf("output-2010-overfit-dendrogram-lc1.pdf", width=7, height=5.5)
plot(tree2010.lc1)
text(tree2010.lc1, cex=0.50)
dev.off()
pdf("output-2010-overfit-dendrogram-lc2.pdf", width=7, height=5.5)
plot(tree2010.lc2)
text(tree2010.lc2, cex=0.50)
dev.off()
pdf("output-2010-overfit-dendrogram-lc3.pdf", width=7, height=5.5)
plot(tree2010.lc3)
text(tree2010.lc3, cex=0.50)
dev.off()
pdf("output-2010-overfit-dendrogram-lc4.pdf", width=7, height=5.5)
plot(tree2010.lc4)
text(tree2010.lc4, cex=0.50)
dev.off()
pdf("output-2010-overfit-dendrogram-lc5.pdf", width=7, height=5.5)
plot(tree2010.lc5)
text(tree2010.lc5, cex=0.50)
dev.off()
pdf("output-2010-overfit-dendrogram-lc6.pdf", width=7, height=5.5)
plot(tree2010.lc6)
text(tree2010.lc6, cex=0.50)
dev.off()
pdf("output-2010-overfit-dendrogram-lc7.pdf", width=7, height=5.5)
plot(tree2010.lc7)
text(tree2010.lc7, cex=0.50)
dev.off()
pdf("output-2010-overfit-dendrogram-lc8.pdf", width=7, height=5.5)
plot(tree2010.lc8)
text(tree2010.lc8, cex=0.50)
dev.off()
pdf("output-2010-overfit-dendrogram-lc9.pdf", width=7, height=5.5)
plot(tree2010.lc9)
text(tree2010.lc9, cex=0.50)
dev.off()
pdf("output-2010-overfit-dendrogram-lc10.pdf", width=7, height=5.5)
plot(tree2010.lc10)
text(tree2010.lc10, cex=0.50)
dev.off()
pdf("output-2010-overfit-dendrogram-lc11.pdf", width=7, height=5.5)
plot(tree2010.lc11)
text(tree2010.lc11, cex=0.50)
dev.off()

```

```

# Plot tree object (pruned) and save as PDF file

# 2007
pdf("output-2007-pruned-dendrogram-lc1.pdf", width=7, height=5.5)
plot(prune2007.lc1)
text(prune2007.lc1, cex=0.50)
dev.off()
pdf("output-2007-pruned-dendrogram-lc2.pdf", width=7, height=5.5)
plot(prune2007.lc2)
text(prune2007.lc2, cex=0.50)
dev.off()
pdf("output-2007-pruned-dendrogram-lc3.pdf", width=7, height=5.5)
plot(prune2007.lc3)
text(prune2007.lc3, cex=0.50)
dev.off()
pdf("output-2007-pruned-dendrogram-lc4.pdf", width=7, height=5.5)
plot(prune2007.lc4)
text(prune2007.lc4, cex=0.50)
dev.off()
pdf("output-2007-pruned-dendrogram-lc5.pdf", width=7, height=5.5)
plot(prune2007.lc5)
text(prune2007.lc5, cex=0.50)
dev.off()
pdf("output-2007-pruned-dendrogram-lc6.pdf", width=7, height=5.5)
plot(prune2007.lc6)
text(prune2007.lc6, cex=0.50)
dev.off()
pdf("output-2007-pruned-dendrogram-lc7.pdf", width=7, height=5.5)
plot(prune2007.lc7)
text(prune2007.lc7, cex=0.50)
dev.off()
pdf("output-2007-pruned-dendrogram-lc8.pdf", width=7, height=5.5)
plot(prune2007.lc8)
text(prune2007.lc8, cex=0.50)
dev.off()
pdf("output-2007-pruned-dendrogram-lc9.pdf", width=7, height=5.5)
plot(prune2007.lc9)
text(prune2007.lc9, cex=0.50)
dev.off()
pdf("output-2007-pruned-dendrogram-lc10.pdf", width=7, height=5.5)
plot(prune2007.lc10)
text(prune2007.lc10, cex=0.50)
dev.off()
pdf("output-2007-pruned-dendrogram-lc11.pdf", width=7, height=5.5)
plot(prune2007.lc11)
text(prune2007.lc11, cex=0.50)
dev.off()
# 2008

```

```

pdf("output-2008-pruned-dendrogram-lc1.pdf", width=7, height=5.5)
plot(prune2008.lc1)
text(prune2008.lc1, cex=0.50)
dev.off()
pdf("output-2008-pruned-dendrogram-lc2.pdf", width=7, height=5.5)
plot(prune2008.lc2)
text(prune2008.lc2, cex=0.50)
dev.off()
pdf("output-2008-pruned-dendrogram-lc3.pdf", width=7, height=5.5)
plot(prune2008.lc3)
text(prune2008.lc3, cex=0.50)
dev.off()
pdf("output-2008-pruned-dendrogram-lc4.pdf", width=7, height=5.5)
plot(prune2008.lc4)
text(prune2008.lc4, cex=0.50)
dev.off()
pdf("output-2008-pruned-dendrogram-lc5.pdf", width=7, height=5.5)
plot(prune2008.lc5)
text(prune2008.lc5, cex=0.50)
dev.off()
pdf("output-2008-pruned-dendrogram-lc6.pdf", width=7, height=5.5)
plot(prune2008.lc6)
text(prune2008.lc6, cex=0.50)
dev.off()
pdf("output-2008-pruned-dendrogram-lc7.pdf", width=7, height=5.5)
plot(prune2008.lc7)
text(prune2008.lc7, cex=0.50)
dev.off()
pdf("output-2008-pruned-dendrogram-lc8.pdf", width=7, height=5.5)
plot(prune2008.lc8)
text(prune2008.lc8, cex=0.50)
dev.off()
pdf("output-2008-pruned-dendrogram-lc9.pdf", width=7, height=5.5)
plot(prune2008.lc9)
text(prune2008.lc9, cex=0.50)
dev.off()
pdf("output-2008-pruned-dendrogram-lc10.pdf", width=7, height=5.5)
plot(prune2008.lc10)
text(prune2008.lc10, cex=0.50)
dev.off()
pdf("output-2008-pruned-dendrogram-lc11.pdf", width=7, height=5.5)
plot(prune2008.lc11)
text(prune2008.lc11, cex=0.50)
dev.off()
# 2009
pdf("output-2009-pruned-dendrogram-lc1.pdf", width=7, height=5.5)
plot(prune2009.lc1)
text(prune2009.lc1, cex=0.50)

```

```

dev.off()
pdf("output-2009-pruned-dendrogram-lc2.pdf", width=7, height=5.5)
plot(prune2009.lc2)
text(prune2009.lc2, cex=0.50)
dev.off()
pdf("output-2009-pruned-dendrogram-lc3.pdf", width=7, height=5.5)
plot(prune2009.lc3)
text(prune2009.lc3, cex=0.50)
dev.off()
pdf("output-2009-pruned-dendrogram-lc4.pdf", width=7, height=5.5)
plot(prune2009.lc4)
text(prune2009.lc4, cex=0.50)
dev.off()
pdf("output-2009-pruned-dendrogram-lc5.pdf", width=7, height=5.5)
plot(prune2009.lc5)
text(prune2009.lc5, cex=0.50)
dev.off()
pdf("output-2009-pruned-dendrogram-lc6.pdf", width=7, height=5.5)
plot(prune2009.lc6)
text(prune2009.lc6, cex=0.50)
dev.off()
pdf("output-2009-pruned-dendrogram-lc7.pdf", width=7, height=5.5)
plot(prune2009.lc7)
text(prune2009.lc7, cex=0.50)
dev.off()
pdf("output-2009-pruned-dendrogram-lc8.pdf", width=7, height=5.5)
plot(prune2009.lc8)
text(prune2009.lc8, cex=0.50)
dev.off()
pdf("output-2009-pruned-dendrogram-lc9.pdf", width=7, height=5.5)
plot(prune2009.lc9)
text(prune2009.lc9, cex=0.50)
dev.off()
pdf("output-2009-pruned-dendrogram-lc10.pdf", width=7, height=5.5)
plot(prune2009.lc10)
text(prune2009.lc10, cex=0.50)
dev.off()
pdf("output-2009-pruned-dendrogram-lc11.pdf", width=7, height=5.5)
plot(prune2009.lc11)
text(prune2009.lc11, cex=0.50)
dev.off()
# 2010
pdf("output-2010-pruned-dendrogram-lc1.pdf", width=7, height=5.5)
plot(prune2010.lc1)
text(prune2010.lc1, cex=0.50)
dev.off()
pdf("output-2010-pruned-dendrogram-lc2.pdf", width=7, height=5.5)
plot(prune2010.lc2)

```

```

text(prune2010.lc2, cex=0.50)
dev.off()
pdf("output-2010-pruned-dendrogram-lc3.pdf", width=7, height=5.5)
plot(prune2010.lc3)
text(prune2010.lc3, cex=0.50)
dev.off()
pdf("output-2010-pruned-dendrogram-lc4.pdf", width=7, height=5.5)
plot(prune2010.lc4)
text(prune2010.lc4, cex=0.50)
dev.off()
pdf("output-2010-pruned-dendrogram-lc5.pdf", width=7, height=5.5)
plot(prune2010.lc5)
text(prune2010.lc5, cex=0.50)
dev.off()
pdf("output-2010-pruned-dendrogram-lc6.pdf", width=7, height=5.5)
plot(prune2010.lc6)
text(prune2010.lc6, cex=0.50)
dev.off()
pdf("output-2010-pruned-dendrogram-lc7.pdf", width=7, height=5.5)
plot(prune2010.lc7)
text(prune2010.lc7, cex=0.50)
dev.off()
pdf("output-2010-pruned-dendrogram-lc8.pdf", width=7, height=5.5)
plot(prune2010.lc8)
text(prune2010.lc8, cex=0.50)
dev.off()
pdf("output-2010-pruned-dendrogram-lc9.pdf", width=7, height=5.5)
plot(prune2010.lc9)
text(prune2010.lc9, cex=0.50)
dev.off()
pdf("output-2010-pruned-dendrogram-lc10.pdf", width=7, height=5.5)
plot(prune2010.lc10)
text(prune2010.lc10, cex=0.50)
dev.off()
pdf("output-2010-pruned-dendrogram-lc11.pdf", width=7, height=5.5)
plot(prune2010.lc11)
text(prune2010.lc11, cex=0.50)
dev.off()

```