



Forest Extent Monitoring in the Congo Basin using
L-Band Synthetic Aperture Radar (SAR)

Thesis submitted for the degree of
Doctor of Philosophy
In Geographical Information Systems
at the University of Leicester

James E.M. Wheeler MSc.
University of Leicester
School of Geography, Geology and the Environment
Centre for Landscape and Climate Research

Submitted: June, 2018

Abstract

Forest Extent Monitoring in the Congo Basin using L-Band Synthetic Aperture Radar (SAR)

James E.M. Wheeler

This thesis presents improved forest extent monitoring using L-band Synthetic Aperture Radar (SAR) data freely provided by the Japanese Aerospace Exploration Agency (JAXA) over the Congo Basin, which contains the second largest area of rainforest in the world. Forest loss in the region, estimated to be up to 0.63 billion Mg in the period 1990-2005, is predominantly driven by bush fuel collection, at a characteristically small scale. Single medium resolution SAR scene (75 km x 75 km), wide area (550 km x 550 km) and full Central African (2000 km x 3300 km) forest extent classifications are generated, and inform best practice for operational annual forest cover production from L-band SAR data. Improvements in one or more of overall accuracy, consistency, scope and replicability are observed compared with existing wide area forest cover and forest/non-forest products in Central Africa, using robust statistical methods to quantify errors in reported class areas. Seasonally inundated forest, a regional obstacle to previous SAR forest cover classifications, is identified by the range of co-polarised SAR data and tested using a novel metric incorporating a normalised cumulative rainfall value aggregated by sub-basin catchment area and SAR polarimetric analysis, which is itself compared with coarse resolution Soil Moisture and Vegetation Water Content Metrics from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E). A quantitative analysis of multiple SAR resolutions supports the continued and future use of 30 m and higher resolution L-band SAR data to map forest cover in the region.

Acknowledgements

My first, and deepest, thanks goes to my supervisors, Dr. Kevin Tansey and Prof. Heiko Balzter, for their invaluable direction, support and encouragement throughout my studies.

I am secondly grateful for the many invaluable suggestions and comments of my external and internal examiners, Dr. Pete Bunting and Dr. Jörg Kaduk; they have guided me towards a vastly improved thesis, and through my revisions I have learned a great deal that will assist me in my career as a remote sensing scientist.

A special thank you goes to Dr. Virginia Nicolás Perea, project manager of the GMES Initial Operations Network for Earth Observation Training (GIONET) through which I was employed and enrolled at Leicester, who succeeded in shepherding 14 unruly men and women through three years of enjoyable European educational mayhem with only an occasional victim to the dreaded *paternoster*.

I am grateful to the Department of Geography at Leicester and the Centre for Landscape and Climate Research including all support and academic staff past and present who have provided friendly advice, professional assistance, and tolerance of my consistent forgetfulness regarding the timely submission of university travel insurance forms and risk assessment waivers. A special mention goes to Dr. Nick Tate who has given me timely pressure towards my goals in his former capacity as Director of Postgraduate Studies, as well as more informal encouragement since then.

My GIONET travels supported secondments to the German Aerospace Agency in Oberpfaffenhofen near Munich, and Gamma Remote Sensing A.G. in Gumligen, Switzerland, and I am grateful for the assistance and support I received during the months I spent there, in particular to Dr. Maurizio Santoro who assisted in the inspirational seeds of Chapter 6, and gave valuable insight into the complexities of SAR data.

My research would have halted very early on were it not for the continuous release of ever improving ALOS-PALSAR mosaic data from the Japanese Aerospace Exploration Agency and the Kyoto and Carbon Initiative, whose 21st Science Team meeting in Kyoto I had the fortune to attend and present my work. It is important to stress that any criticism of the JAXA Forest Non-Forest product in this work is respectfully made with a view to informing and assisting future improvements, and the feeling towards the K&C Initiative Science Team is one of overwhelming gratefulness for making this product and the mosaics, integral to my research and that of my colleagues, freely available and so user friendly at a global scale.

I have sincerely enjoyed my time in Leicester, and I am socially and academically indebted to my University of Leicester PhD peers, in Geography and the CLCR including, but not limited to, those I have played alongside and against in the Wednesday lunchtime staff and postgraduate intramural football league and all who came to support us. I will fondly remember Wednesday research methods seminars, coffee/crossword collectives (Alex, Valentin, Pedro, Marc, Tom and Ciaran), and of course Thirsty Thursdays (Hannah, Claire and Jen).

I have developed a penchant for swing dance during my time in Leicester, and David, Phil, Nat, James, Joel, Liam, Alice, Oz and Kate will see me Lindy-hop back from time to time beyond my studies. In addition, I couldn't have asked for nicer friends and housemates than Mariana, Mariangela, and Marycarmen.

I will never forget the GIONET ESRs I had the pleasure of travelling around Europe with for three years, especially Pedro (and Jelly and Nico) in Leicester, and I look forward to a reunion down the line at a suitably distant remote sensing conference with Linda, Bernard, Martyna, Shailesh, Penelope, Christoff, Stephanie, Matthew, Jessica, Maxim, Dimitris and, last but not least, C-band.

For introducing me to the concept of Remote Sensing in 2009, and SAR shortly afterwards, I remain thankful to all those I studied alongside, under and near in University College Cork and the Coastal and Marine Research Centre during my MSc; Fiona, Ned, Darius, Helen, Brian, Brian, Diego, Ali and Masha especially.

My friends and ever expanding family in Ireland deserve a considerable mention due to my protracted absence, and I would like to dedicate this thesis to my nephew and two nieces who I will in the not too distant future bore senseless at every opportunity with tales of SAR, GIONET, and Leicester.

Funding for this PhD research came from the GIONET Marie Curie Programme Initial Training Network, funded by the European Commission, Grant Agreement PITN-GA-2010-264509. I express my appreciation for this opportunity, and the many summer schools, secondments, training workshops and numerous conferences that it enabled me to attend, as well as the many people who worked tirelessly behind and in front of the scenes to make it a success.

Table of Contents

Abstract.....	i
Acknowledgements	ii
Table of Contents.....	v
List of Figures	x
List of Tables	xiv
Chapter 1: Introduction	1
1.1 Tropical Forest Loss.....	1
1.2 The Congo Basin.....	6
1.3 Forest Definition.....	6
1.4 Summary and Aims	8
Chapter 2: Background	9
2.1 Remote Sensing of Tropical Forests.....	9
2.2 Remote Sensing of Central African Forests.....	10
2.2.1 Central African SAR Forest Mapping	14
2.2.2 Forest Definitions Used.....	19
2.3 SAR Remote Sensing	20
2.3.1 SAR Basics and Signal Processing.....	20
2.3.2 SAR Scattering Mechanisms	22
2.3.3 SAR Frequency and Wavelength.....	23
2.3.4 SAR Polarisation	24
2.3.5 SAR Interferometry and Coherence	24
2.3.6 Polarimetric SAR Interferometry (PolInSAR)	26
2.3.7 Sensitivity of SAR to Moisture	27
2.3.8 SAR Spatial Resolution vs Speckle.....	30

2.4	Classification Methods	33
2.4.1	Thresholding Classifier	34
2.4.2	Support Vector Machine Classifier	34
2.4.3	Random Forests Classifier	35
Chapter 3: Study Sites		38
3.1	Description of Study Sites	39
Chapter 4: New Map of Forest Cover in the Congo Basin Derived from L-band SAR Mosaics and Existing Optical-based Forest Tree Cover Data.....		43
4.1	Introduction	43
4.1.1	Aim and Objectives	43
4.2	Datasets.....	44
4.2.1	ALOS PALSAR.....	44
4.2.2	Landsat Tree Cover Continuous Fields (LTC) and Moderate Resolution Imaging Spectroradiometer - Vegetation Continuous Fields (MODIS-VCF)	47
4.2.3	JAXA K&C Forest / Non-Forest.....	49
4.2.4	ESA-GlobCover 2009	50
4.3	Methods	53
4.3.1	Data Pre-Processing.....	53
4.3.2	Classification	54
4.3.3	Accuracy Assessment.....	56
4.3.4	Software Used.....	59
4.4	Results	59
4.4.1	Visual Assessment.....	59
4.4.2	Error Matrices	62
4.4.3	GlobCover class comparisons	64
4.5	Discussion.....	69

4.6	Conclusions	71
Chapter 5: Creating Annual Forest/Non-Forest Maps in a region of Central Africa		
	from 2007-10 using SAR Mosaic data	74
5.1	Introduction	74
5.1.1	Aims and Objectives.....	74
5.2	Datasets.....	75
5.2.1	ALOS PALSAR.....	75
5.2.2	JAXA K&C Forest / Non-Forest.....	76
5.3	Methods	76
5.3.1	Data and Pre-processing.....	78
5.3.2	Generating Training and Validation Data	80
5.3.3	Classification	81
5.3.4	Change Analysis	84
5.3.5	Accuracy Assessment.....	84
5.3.6	Software Used.....	86
5.4	Results	86
5.4.1	Error Matrices	86
5.4.2	Forest Area Change.....	88
5.4.3	Class Stability of Pixels	91
5.4.4	Classification Differences.....	91
5.5	Discussion.....	97
5.6	Conclusions	99
Chapter 6: Multi-scale Analysis of Forest and Inundated Forest from Small Area		
	SAR Time Series Data	100
6.1	Introduction / Aim and Objectives.....	100
6.2	Datasets.....	101

6.2.1	ALOS PALSAR L1.1	101
6.2.2	RFE 2.0.....	102
6.2.3	HYDRO-1K	102
6.2.4	AMSR-E.....	103
6.2.5	SRTM	103
6.3	Methods	103
6.3.1	ALOS PALSAR Pre-processing.....	104
6.3.2	Generation and Selection of Derived Texture and Difference Layers 106	
6.3.3	Generation of Reference Data	109
6.3.4	Class Separability	111
6.3.5	Classification Procedures	112
6.3.6	Post-Classification	113
6.3.7	Scaling Thematic Classification	113
6.3.8	Validation	113
6.3.9	Basin Excess Rainfall Maps / Normalised Cumulative Rainfall	113
6.3.10	Use of AMSR-E SM and VW	116
6.3.11	Software Used.....	116
6.4	Results	117
6.4.1	Class Separability	117
6.4.2	RF and SVM Classification Comparison	120
6.4.3	Texture Information.....	124
6.4.4	Classification at Different Scales.....	124
6.4.5	Comparison to Thematic Maps Resampling	132
6.4.6	Class Area Differences	136
6.4.7	NCR Results	139

6.4.8	AMSR-E Results	145
6.5	Discussion.....	147
6.5.1	Reference Dataset	148
6.5.2	Image Statistics and Multi-temporal Speckle Filtering.....	148
6.5.3	Scale Effects	149
6.5.4	NCR Effects.....	151
6.6	Conclusions	152
Chapter 7: Discussion and Conclusions		155
7.1	Discussion in the Context of Literature.....	155
7.2	Conclusions in the Context of Research Questions	156
7.2.1	Question 1: Ideal SAR Imagery and Methods for Tropical Forest Mapping.....	156
7.2.2	Question 2: HH Range and NCR	158
7.2.3	Question 3: Scale effects	158
7.3	Future Direction of SAR Remote Sensing in the Congo Basin.....	159
Appendix A – Chapter 4 additional figures		161
A1	GlobCover 2009 distribution comparisons.....	161
Appendix B – Chapter 6 additional figures		163
B1	Haralick GLCM Texture Class Histograms	163
B2	Filtered and Unfiltered Scatterplot comparisons	178
B3	Inter Class Transform Divergence in HH and HV channels.....	180
B4	Scale differences in forest classification.....	183
B5	Confusion Matrices from RF classifications.....	186
Bibliography		190

List of Figures

Figure 2.1: Graphic showing how a support vector machine would choose a separating hyperplane for two classes of points in 2D.	35
Figure 3.1: Study sites and relative location with corresponding GlobCover 2009 dataset	42
Figure 4.1: Samples (different scales) of the K&C Initiative ALOS-PALSAR 50 m 2009 HH mosaic	46
Figure 4.2: K&C initiative ALOS-PALSAR 50m mosaic Version 1 availability over Central and West Africa.....	47
Figure 4.3: Mosaic of Landsat Tree Cover Continuous Fields with MODIS VCF gap fill, covering Central Africa.	48
Figure 4.4: K&C Initiative FNF map 2009 for Central African region	50
Figure 4.5: GlobCover 2009 with full GlobCover LCCS class legend	51
Figure 4.6: GlobCover 2009 Aggregated Classes	52
Figure 4.7: ALOS PALSAR HV mosaic (sigma zero) from 2009.....	53
Figure 4.8: Comparison of cumulative area of HV σ^0 values with aggregated GlobCover forest cover classes.....	55
Figure 4.9: Decision tree classification schematic of SAR (2009 PALSAR HV sigma zero values) and optical (LTC) data.....	56
Figure 4.10: Distribution of validation reference points, with reference class legend.	57
Figure 4.11: Screen grabs from Google Earth of reference dataset generation.	58
Figure 4.12: Full Forest/Non-Forest optical/SAR classification comparison	60
Figure 4.13: Forest/non-forest optical/SAR classification comparison, with eight selected areas of interest	61
Figure 4.14: Globcover Aggregated Class Legend	64
Figure 4.15: Comparison of FNF classes in terms of area distribution of aggregated GlobCover classes	65
Figure 4.16: Comparison of FNF classes in terms of percent of the total area distribution of aggregated GlobCover classes	66

Figure 4.17: Reference data distribution of GlobCover 2009 aggregated classes for differing LTC and SAR threshold results	68
Figure 4.18: Cumulative LTC area comparison with areas of aggregated GlobCover 2009 classes	69
Figure 5.1: 2007 K&C Initiative products in the case study area.....	76
Figure 5.2: Processing steps for 2007-10 forest change map based on support vector machine classification.....	77
Figure 5.3: K&C Initiative ALOS-PALSAR mosaic (Version 2) tiles over Central and West Africa.....	79
Figure 5.4: Distribution of training and reference/validation plots in the study site	81
Figure 5.5: SVM Kernel function classification comparison	83
Figure 5.6: Reported forest area in km ² by both classifications over the timeseries (2007-2010).....	89
Figure 5.7: Graph of error adjusted forest area by year	90
Figure 5.8 Graph of error adjusted non-forest area by year	90
Figure 5.9: Comparison of forest gain/loss derived from SVM classification and JAXA FNF on the study site	93
Figure 5.10: Comparison of forest gain/loss derived from SVM classification and JAXA FNF on a subset of the study site.....	94
Figure 5.11: Comparison of forest gain/loss derived from SVM classification and JAXA FNF on another subset of the study site.....	95
Figure 5.12: Comparison of forest gain/loss derived from SVM classification and JAXA FNF on a further subset of the study site	96
Figure 6.1: Flow chart of ALOS-PALSAR processing chain	105
Figure 6.2: Histograms of the eight Haralick GLCM texture measures from Training and Validation sample areas.....	108
Figure 6.3: ALOS-PALSAR HH σ^0 eight-scene range image used to classify flooded forest areas	109
Figure 6.4: Spatial distribution of reference polygons	111
Figure 6.5: Study area scene extent, HYDRO1K basin catchment areas (right images), and example NCR values.....	115

Figure 6.6: Graph of mean NCR values from the ALOS-1 lifetime.....	116
Figure 6.7: Filtered ALOS-PALSAR HV/HH classified scatterplot comparison	118
Figure 6.8: Filtered/Unfiltered ALOS-PALSAR HV/HH classified scatterplot comparison	119
Figure 6.9: Inter class transformed divergence at 15 m in HH	120
Figure 6.10: Random Forests parameter search results.....	121
Figure 6.11: Processing speed comparison between SVM and RF.....	122
Figure 6.12: RF vs SVM classification	123
Figure 6.13: Overall accuracy assessment comparison between SVM and RF classifiers.....	123
Figure 6.14: Loss of accuracy from removing mean and variance texture layers at HH and HV polarisations.....	124
Figure 6.15: Comparison of classifications (top) at 15, 30, 45 and 60m resolutions with RGB composite.....	126
Figure 6.16: Comparison of classifications (top); flooded forest detail with RGB composite (below)	127
Figure 6.17: Comparison of classifications (top); non-forest detail with RGB composite (below)	128
Figure 6.18: Comparison of accuracy assessments for RF classifications at multiple scales.....	130
Figure 6.19: Flooded forest class confusion: forest and non-forest comparison..	131
Figure 6.20: Class comparison between pre- (multilooking) and post-classification (thematic resampling) scale adjustments; full scene extent.....	133
Figure 6.21: Class comparison between pre- (multilooking) and post-classification (thematic resampling) scale adjustments; flooded forest detail	134
Figure 6.22: Class comparison between pre- (multilooking) and post-classification (thematic resampling) scale adjustments; non-forest detail	135
Figure 6.23: Comparison between class areas in km ² reported by thematic map resampling and pre-classification multilooking.....	136
Figure 6.24: Comparison of reported class areas at different scales	137
Figure 6.25: Area range and standard deviation at all spatial resolutions.....	138
Figure 6.26: Comparison of NCR values with ALOS PALSAR HH σ^0 backscatter	141

Figure 6.27: Comparison of NCR values with RF classification.....	142
Figure 6.28: Comparison of mean NCR values with mean ALOS-PALSAR HH backscatter at multiple resolutions	143
Figure 6.29: Comparison of RF classified scenes with the corresponding K&C Initiative Forest/Non-Forest product at multiple resolutions.....	144
Figure 6.30: AMSR-E Soil Moisture [gcm^{-3}] and Vegetation Water Content [kgm^{-3}] compared with the NCR [%] in the study area	146
Figure 6.31: Comparison of AMSR-E VW (left) and SM (right) with NCR (top) and RFE 2.0 10 dekadal sum	147
Figure 6.32: Comparison of 8-scene coefficient of variation of class area with 8 scene mean class accuracy at multiple resolutions	150
Figure 6.33: Relationship between mean ALOS-PALSAR HH σ^0 backscatter and mean NCR values	152
Figure B2.0.1: Classified scatterplots from filtered ALOS-PALSAR HV/HH time series from 2007-10 at decreasing spatial resolutions from left to right.....	178
Figure B2.0.2: Classified scatterplots from unfiltered ALOS-PALSAR HV/HH time series from 2007-10 at decreasing spatial resolutions from left to right.....	179

List of Tables

Table 1.1: Comparison of methods for measuring tropical forest carbon stocks. Adapted and updated from Gibbs et al. (2007).....	5
Table 2.1: Major wide area remote sensing projects relevant to Central African forest mapping, with methods and reported accuracies	11
Table 2.2: Selected SAR tropical wetland and forest extent studies, from the 1990s to present, including methods, forest definition and reported accuracy	17
Table 2.3: Historic, current, and future spaceborne SAR satellites/sensors	23
Table 2.4: Random Forests Parameters.....	37
Table 4.1: Error Matrices for SAR FNF, LTC FNF, and JAXA K&C FNF products, including Mixed reference class	63
Table 4.2: Error Matrices for SAR FNF, LTC FNF, and JAXA K&C FNF products, excluding Mixed reference class.....	63
Table 4.3: Error Matrix of FNF from combined LTC and SAR thresholds	64
Table 5.1: Error matrices for different SAR filtering options from 2010 data.....	87
Table 5.2: Confusion Matrices in sample numbers for each class	88
Table 5.3: Forest area in km ² and as percentage of total study site area	88
Table 5.4: Error adjusted areas and 95% confidence margins of error, expressed as percent of total study site area	91
Table 6.1 Description of ALOS-PALSAR scenes acquired	102
Table 6.2: Class Accuracy Statistics.....	131
Table 6.3: Class area statistics from RF classification	138

Chapter 1: Introduction

This chapter outlines the context of this research, beginning with a review of the importance of tropical forests, their uses and value, and the direct and indirect threats posed by human activity, principally through wood extraction and development, and the effects of climate change. The contribution of forest loss to climate change is also discussed. The focus then moves to tropical forests in the Congo Basin, placing it in a global context and highlighting its vulnerability and the relative lack of data from the region. The difficulties associated with various forest carbon measurements in the region are covered, and a rationale for improved forest mapping to supplement forest carbon stock measuring is argued. The benefits and limitations of the two major remote sensing methods used for wide area forest mapping, optical and radar, are outlined in the context of nine major studies of the region completed since the 1980s. Finally, a justification for this thesis's use of radar remote sensing is presented, including an explanation of its fundamental principles, and the knowledge gaps and methodological difficulties, forming the basis of the three major research questions.

1.1 Tropical Forest Loss

Deforestation and forest degradation are a major source of anthropogenic greenhouse gas emissions (Malhi & Grace, 2000). Stored carbon is released into the atmosphere when forests are burned to create land available for agriculture (FAO, 2010). Exposed and especially drained soil also releases carbon, which for certain soil types such as peat has been stored in greater volumes (as soil organic carbon, SOC) than in the living woody material (as above ground biomass, AGB) (Englhart et al., 2012). Conversion of forest to other land cover and land use types reduces the area available to store carbon and therefore reduces the ability to offset the increases of atmospheric carbon from fossil fuel emissions. Deforestation and forest degradation result in loss of biodiversity, and are specifically mentioned in the Convention for Biological Diversity's (CBD) 'Aichi

Biodiversity Targets' as part of their 2011-2020 plan (CBD, 2010). While economic growth usually accompanies forest loss, caused by increased mineral extraction, logging and clearing for agriculture and other infrastructure developments, the economic cost of forest loss can be considerable. The United Nations Food and Agriculture Organisation (UN-FAO) produces a regular global Forest Resources Assessment (FRA), which describes the current state of the world's forests based on submitted reports from individual countries, and more recently incorporates independent remote sensing based surveys. The second most recent FRA, in 2010, describes in detail the various economic uses of forests, or 'ecosystem services', divided into productive and protective categories (FAO, 2010). Productive uses include extraction of wood for fuel and industrial roundwood (wood forest products), and for food and fodder (non-wood forest products (NWFPs)). Wood forest product extraction is relatively easy to quantify, and while it can be destructive and lead to forest degradation, sustainable management of these resources can encourage conservation and regulation of carbon loss, especially in low population density rural areas (Megevand et al., 2013). The protective function of forests relates mostly to their role in soil and water protection (including preventing desertification and for conservation of water reserves), but also to coastal defences (FAO, 2010).

The size of the contribution of forests to the global carbon cycle, both as sources and as sinks of carbon, is currently not known to an acceptable standard for climate modelling. Plants capture atmospheric CO₂ during photosynthesis and store it as biomass, which contains ~50% carbon (Gibbs et al., 2007). The net carbon captured by plants during photosynthesis, after the carbon lost through plant (autotrophic) respiration is accounted for, is known as net primary production (NPP), and is used to assess the response of land based plants to climate change (Cao & Woodward, 1998; Melillo et al., 1993). NPP increases with increasing levels of atmospheric CO₂ (CO₂ fertilisation), up to a saturation point as plants physiologically acclimatise (Cao & Woodward, 1998). As the rate of plant respiration may increase with increasing levels of atmospheric CO₂, the effect of increased photosynthetic activity may be lessened. However, the amount of carbon captured through global forest NPP is exceeded by the carbon released from

forest loss. While there is uncertainty about the exact quantity of carbon stored globally in forests, tropical forests contain a greater mass of carbon per hectare of forest and harbour greater biodiversity than temperate and boreal forests. Global carbon emissions from land use change (which principally refers to emissions from changes in areas of forest) has reportedly remained constant since 1959 at 1.5 ± 0.5 billion tonnes of carbon per year (PgC year^{-1}) (Canadell et al., 2007). However, due to reforestation and afforestation efforts in temperate and boreal biomes, the spatial distribution of land use change emissions is now almost entirely within the tropics (*ibid.*). This means that tropical forests play a greater role in both land use change emissions, and at the same time demonstrate a greater capacity for carbon sequestration than other forest biomes. A much longer growing season in tropical areas allows for faster forest regeneration during reforestation efforts, and higher volume and faster carbon sequestration. In addition, the capacity to mitigate climate change through afforestation in boreal regions may be less than previously expected, due to the decrease in surface sunlight reflectivity (albedo) associated with forest cover compared with snow/ice that it would replace (Betts, 2000). Over tropical forests, however, it is predicted that the albedo may be higher than unforested areas due to the increased cloud cover from evapotranspiration (Canadell & Raupach, 2008). This suggests that efforts to reduce climate change through reforestation and reduction of deforestation and forest degradation are better directed towards tropical regions. Monitoring the loss of carbon into the atmosphere is of particular importance in tropical forests, where the density of AGB, measured in metric tonnes of carbon per hectare, is much higher on average than in temperate or boreal forests. In addition, tropical forests tend to be less accessible, for logistical and often political reasons, causing problems for wide-scale and regular ground-based monitoring.

A key intergovernmental programme that has been created in recognition of the important role of tropical forests in climate change, and the potential benefits of their conservation, is the UN REDD programme (Reducing Emissions from Deforestation and forest Degradation in developing countries), which aims to incentivise reducing carbon emissions from forest loss in developing countries through financial means (Herold & Johns, 2007). REDD supports REDD+, which adds “conservation of forest stocks,

sustainable management of forests, and enhancement of forest carbon stocks” to the climate change mitigating actions of REDD (UNFCCC, 2010). In order to achieve the goals of REDD, a system of monitoring, reporting and verification (MRV) is necessary to assess forest area and stocks of carbon contained in AGB in participating countries. Table 1.1 outlines the benefits and limitations of several techniques used for measuring tropical forest carbon stocks, and remains largely correct since original publication by Gibbs *et al.* (2007). All methods listed ultimately rely on forest inventory data to calculate either local, regional, species-specific or biome-specific relationships between their forest measurements and the carbon contained.

Table 1.1: Comparison of methods for measuring tropical forest carbon stocks. Adapted and updated from Gibbs et al. (2007)

Method	Description	Benefits	Limitations	Uncertainty
Biome averages	Estimates of average forest carbon stocks for broad forest categories based on a variety of input data sources	<ul style="list-style-type: none"> • Immediately available at no cost • Data refinements could increase accuracy • Globally consistent 	<ul style="list-style-type: none"> • Fairly generalized • Data sources not properly sampled to describe large areas 	High
Forest inventory	Relates ground-based measurements of tree diameters or volume to forest carbon stocks using allometric relationships	<ul style="list-style-type: none"> • Generic relationships readily available • Low-tech method widely understood • Can be relatively inexpensive as field-labor is largest cost 	<ul style="list-style-type: none"> • Generic relationships not appropriate for all regions • Can be expensive and slow • Challenging to produce globally consistent results 	Low
Optical remote sensors	<ul style="list-style-type: none"> • Uses visible and infrared wavelengths to measure spectral indices and correlate to groundbased forest carbon measurements • Ex: Landsat, MODIS 	<ul style="list-style-type: none"> • Satellite data routinely collected and freely available at global scale • Globally consistent 	<ul style="list-style-type: none"> • Limited ability to develop good models for tropical forests • Spectral indices saturate at relatively low C stocks • Can be technically demanding 	High
Very high-res. airborne optical remote sensors	<ul style="list-style-type: none"> • Uses very high resolution (~10–20 cm) images to measure tree height and crown area and allometry to estimate carbon stocks • Ex: Aerial photos, 3D digital aerial imagery 	<ul style="list-style-type: none"> • Reduces time and cost of collecting forest inventory data • Reasonable accuracy • Excellent ground verification for deforestation baseline 	<ul style="list-style-type: none"> • Only covers small areas (10 000s ha) • Can be expensive and technically demanding • Few allometric relations based on crown area are available¹ 	Low to medium
Radar remote sensors	<ul style="list-style-type: none"> • Uses microwave or radar signal to measure forest vertical structure • Ex: ALOS PALSAR, ERS-1, JERS-1, Envisat, Sentinel 1A, ALOS-2 PALSAR-2, Radarsat 1,2, TerraSAR-X 	<ul style="list-style-type: none"> • Satellite data are generally free • Can be accurate for young or sparse forest • Launch of BIOMASS mission in 2020 will provide dedicated forest carbon sensor with improved signal saturation 	<ul style="list-style-type: none"> • Less accurate in complex canopies of mature forests because signal saturates • Mountainous terrain also increases errors • Can be expensive and technically demanding 	Medium
Laser remote sensors	<ul style="list-style-type: none"> • LiDAR uses laser light to estimate forest height/vertical structure • Ex: Carbon 3-D satellite system combines Vegetation canopy LiDAR (VCL) with horizontal imager 	<ul style="list-style-type: none"> • Accurately estimates full spatial variability of forest carbon stocks • Satellite based IceSAT-GLAS system provided >3 million point estimates of forest structure to assist in tropical forest carbon map² 	<ul style="list-style-type: none"> • Airplane-mounted sensors only currently operational option • Satellite system no longer operational • Requires extensive field data for calibration • Can be expensive and technically demanding 	Low to medium

¹Since 2007 there has been some work on crown allometry, e.g. (Goodman et al., 2014) ²Based on (Saatchi et al., 2011)

1.2 The Congo Basin

There are several recent estimates for the volume of carbon stored in Central African tropical forests (Gaston et al., 1998; Gibbs et al., 2007; Baccini et al., 2008; Saatchi et al., 2011), but there is often a reliance on allometric models that are limited to forest inventory information with very limited samples from the more accessible periphery of the Congo Basin's forests or even fail to include any Central African inventory data and instead use global biome averages (Kearsley et al., 2013).

For reasons such as political instability, numerous civil wars and lack of infrastructure, forest loss is not as severe in Africa compared with the two other major tropical forested continents, South America and South East Asia (Justice et al., 2001; Mayaux & Achard, 2010). Relatively small scale felling to provide biomass fuel is currently the largest driver of anthropogenic deforestation in the Congo Basin, but the potential for agricultural development and exploitation of logging and mining concessions pose a considerable threat to the region's forests (Megevand et al., 2013). In addition to direct human clearing of forests, there is a reported ongoing drying trend in the Northern Congo Basin (Zhou et al., 2014; Chambers & Roberts, 2014), which could build up to a climate change induced drought that would cause a tropical forest collapse and release large volumes of carbon (Lewis, 2006).

For the same reasons, as well as sheer scale, a ground based survey of forest extent for the entire Congo Basin would be hugely impractical, especially for the Democratic Republic of Congo (DRC), which recent estimates report contains over 60% of the Central African forest area (Mayaux et al., 2013).

1.3 Forest Definition

The definition of what constitutes a 'forest' typically refers to a threshold value of tree / canopy cover in a minimum area, with a minimum threshold for tree height and sometimes stem diameter (diameter at breast height – DBH – is a usual measurement)

for those trees included in the calculation. The values for each of these measures is changeable depending on the mapping authority, usually by nation or international monitoring agency (Lund (2015) maintains an exhaustive and regularly updated list of national and international tree and forest area definitions), with differences across the globe, often between neighbouring countries (as a more extreme example, the reported Zimbabwe definition of forest area lists thresholds of >80% canopy cover and >15 m tree height, whereas Mozambique's definition has thresholds of >25% canopy cover and >7 m tree height). This leads to very different forest area reporting, which has a further impact on reporting deforestation and forest degradation (Romijn et al., 2013). The differences often relate to the objectives of the mapping authority, whether they are for environmental purposes, commercial logging, agriculture etc. For example, the United Nations Food and Agriculture Organisation (FAO) consider land use in their definition, excluding tree plantations, whereas the definition by the United Nations Framework Convention on Climate Change (UNFCCC) does not (*ibid.*). Remote sensing measurements do not directly report on forest metrics and always rely to some extent on modelling based on the relationship between a received signal and the particular characteristic, with sources of error at each step. In addition, there are definitions for deforestation, afforestation, forest degradation and even what constitute trees that vary from authority to authority. This study uses the following UNFCCC definition of forest where applicable:

“a minimum area of land of 0.05-1.0 hectares with tree crown cover (or equivalent stocking level) of more than 10-30 per cent with trees with the potential to reach a minimum height of 2-5 metres at maturity in situ. A forest may consist either of closed forest formations where trees of various storeys and undergrowth cover a high proportion of the ground or open forest” (UNFCCC, 2001)

The disparity in reported forest area in one forest cover and areal extent study at a global scale Since there are ranges in this definition, generally the upper thresholds (1 ha minimum area; 30% tree crown cover; 5 m tree height) were used. When analysing remote sensing data in areas where there are few forest plots, it is difficult to adhere to this definition, especially for SAR, where the signal is more closely correlated to the tree density or AGB, especially when multiple wavelengths are used (Englhart et al., 2011).

In addition, the forest degradation definition is not set, as it was considered beyond the scope of the available data to measure and report more subtle changes, particularly in terms of field measurements and validation. Further information about forest definitions used in forest and wetland studies using remote sensing is detailed in section 2.3

1.4 Summary and Aims

The central objective of this research is the generation of forest extent maps, as high quality as possible, of the Congo Basin from spaceborne wide area synthetic aperture radar data acquired during the period 2007-2010. Three research questions are identified as the major challenges posed by this task, and their solutions aim to contribute to scientific knowledge in this field as well as inform the generation of a benchmark forest extent product for local and international forest research in the Congo Basin. These are:

1. What are the ideal image acquisition parameters and classification techniques for an annual forest map in the Congo Basin from spaceborne SAR data, given currently available data and regional seasonal effects on image quality?
2. To what extent, and how, can spaceborne SAR data in the Congo Basin, combined with rainfall data metrics, account for seasonally inundated forest in SAR-derived forest maps?
3. What are the effects of scale on Congo Basin forest maps generated from SAR data, and how does this impact the accuracy of maps generated from coarser resolution wide coverage SAR datasets?

Identifying areas of change in the region, and examining the applicability of the developed techniques to surrounding Central African countries, using similar data, is a key aspect of this research.

Chapter 2: Background

2.1 Remote Sensing of Tropical Forests

Ground-based forest inventories provide the most direct and accurate method of assessing tree extent, as well as other forest parameters such as species distribution and AGB, and are necessary for interpreting remote sensing data and validating results, but alone they are not practical for regular wide area surveys, particularly in dense tropical forest. The UN FAO FRAs, mentioned in Chapter 1, use a template that allows individual countries to report their forest area, most of which is produced from forest inventory surveys of differing quality (three tiers, from ‘expert estimates’ to ‘low intensity or incomplete surveys, older data’ to ‘high reliability, recent sources with national scope’) the most recent being in 2015 . There is a wide range of country report qualities, in terms of detail, how recently and frequently data has been collected, and completeness, with examples at either end of the scale including those from North Korea (FAO, 2015a) and Mexico (FAO, 2015b). The North Korean report is largely incomplete with extrapolated forest area using information from a remote sensing survey conducted in 1998, whereas the country report for Mexico includes information from multiple data sources, including an established network of forest inventories and remote sensing data, and is regularly conducted to monitor change in forest extent.

Remote sensing from airborne and especially spaceborne sensors has the potential for consistent ‘wall-to-wall’, regular surveys of forest extent. Forest extent is measurable from the two most widely used earth observation systems: passive optical (visible, near infrared and short wave infrared frequencies) and active Synthetic Aperture Radar (SAR) (microwave frequencies). SAR also has the potential to inform on aspects of forest structure, depending on the modes of acquisition used. A reliable wall-to-wall forest/non-forest map can also be used to improve methods of AGB modelling by facilitating interpolation between known or modelled values of AGB where there is spacing between samples, an example of which is AGB derived from tree-height

measurements from the Ice, Cloud and land Elevation Satellite's Geoscience Laser Altimeter System (IceSAT-GLAS) instrument (Saatchi et al., 2011).

2.2 Remote Sensing of Central African Forests

Table 2.1 details ten major forest remote sensing products with full coverage of the Congo Basin, and is followed by analysis of each based on their methods, spatial and temporal resolution, and accuracy.

Table 2.1: Major wide area remote sensing projects relevant to Central African forest mapping, with methods and reported accuracies

Project name	Sensor; sensor type	Spatial resolution	Method; reported accuracy/error	Forest Definition	Coverage	Year(s)	Organisation; reference
AVHRR Pathfinder Percent Tree Cover (PTC)	AVHRR; optical	8km	Regression tree analysis, manually interpreted Landsat training data; 11.03% standard error with training data	60% canopy cover (DeFries et al., 2000)	Pan-tropical	Mid-1980s –mid-1990s	University of Maryland (UMD), Woods Hole Research Centre; (DeFries et al., 2002)
Global Remote Sensing Survey (RSS)	Landsat; optical	30m	OBIA, supervised classification of 20km x 20km samples at lat-long degree intersections (1% of global land area); 88% accuracy for change detection	>10% tree cover (Lindquist & D’Annunzio, 2016)	Global	1990, 2000, 2005	UN FAO, Joint Research Centre (JRC); (FAO & JRC, 2012; Lindquist & D’Annunzio, 2016)
GLC2000/ TREES II; Central African Mosaic Project (CAMP)*; Global Rain Forest Mapping Project (GRFM)*	SPOT- VEGETATION; optical; ERS-1; C- band SAR; JERS-1; L-band SAR	1km (100m)	Thresholding of temporal NDVI profiles, visual interpretation, supplementary SAR mosaic datasets; global accuracy of 68.6% across all classes	>40% tree cover (Mayaux et al., 2004)	Global	2000 (1994, 1996)	Joint Research Centre (JRC), NASDA (now JAXA); (Mayaux et al., 2004; De Grandi et al., 1999; Rosenqvist et al., 2000)
Vegetation Continuous Fields (VCF)	Terra-MODIS; optical	250m, 500m, 1km	Regression tree analysis, using Landsat training data; RMSE of tree cover 9-23%	Canopy cover of trees >5 m in height (Gross et al., 2017)	Global	2000-2010 (Annual)	UMD, NASA; (Hansen et al., 2003; DiMiceli et al., 2011)

Project name	Sensor; sensor type	Spatial resolution	Method; reported accuracies/error	Forest Definition	Coverage	Year(s)	Organisation; reference
Tree Cover Continuous Fields	Landsat; optical	30m	Rescaling of MODIS VCF with Landsat mosaics; RMSE of tree cover 9.4%	Tree cover of trees >5 m in height	Pan-tropical	2000	UMD; (Sexton et al., 2013)
Global Forest Change	Landsat; optical	30m	Decision tree model, with reference to high res imagery, MODIS NDVI time series; 81.9%-87% user's accuracy (gain-loss, in tropics)	Canopy cover of trees >5 m in height (Gross et al., 2017)	Global	2000-2012	UMD, NASA, Google; (Hansen et al., 2013)
GlobCover	ENVISAT-MERIS; optical	300m	Unsupervised classification of spectrotemporal information (for vegetation classes) ; 59.9%-61.34% user's accuracy across all landcover types	>15% canopy cover (closed to open forest)	Global	2005/6, 2009	ESA; (Arino et al., 2007; Bontemps et al., 2010)
CAMP*	ERS-1; C-band SAR	200m	Single C-band SAR mosaic; e.g. Amplitude and normalised variance texture analysis and thresholding; 68% overall accuracy for swamp and lowland rain forest	30-70% canopy cover (tree cover mosaic)	Central Africa	1994	JRC; (De Grandi et al., 1999; De Grandi et al., 2000a)
GRFM*	JERS-1; L- band SAR	100m	Two season L-band mosaics; e.g. speckle filtering, bespoke unsupervised pattern recognition algorithm; 87% accuracy for FNF product	>10% canopy cover (the same as that of FAO, 2001)	Pan-tropical	1996 (x2)	NASDA (now JAXA); (Rosenqvist et al., 2000) (Mayaux et al., 2004); JRC (Sgrenzaroli et al., 2004)
ALOS Kyoto & Carbon Initiative FNF	ALOS-PALSAR 1 and 2; L-band SAR	25m, 50m, 500m	OBIA, regional thresholds for FNF; 84.86% - 91.25% overall accuracy	>10% woody cover	Global	2007-2010; 2015- 2017 (Annual)	JAXA, JRC; (Shimada et al., 2014)

**GRFM and CAMP projects describe SAR mosaicking projects – methods and accuracies refer to cited studies that used these datasets*

Pan tropical optical remote sensing approaches to forest mapping began with coarse spatial resolution ($\geq 1\text{km}$) data, combined with samples of medium (what was then considered high) resolution (30m) imagery to estimate rates of forest change. The Tropical Ecosystem Environment Monitoring by Satellites (TREES) programme (Mayaux et al., 1999), the UN FAO FRA 2000 (FAO, 2001), and the Advanced Very High Resolution Radiometer (AVHRR) Pathfinder programme (DeFries et al., 2002) all incorporated one or both of these approaches during the 1990s. The AVHRR Pathfinder derived product described in DeFries et al. (2002) used training data from a global network of 200 manually interpreted Landsat scenes, aggregated to the resolution of AVHRR (8km) to derive a percent tree cover (PTC) product. The AVHRR Pathfinder Land Data Set (Smith et al., 1997) was used, which is an atmospherically corrected AVHRR product from the AVHRR sensors on the NOAA -7, -9 and -11 satellites, containing Normalised Difference Vegetation Index (NDVI) values from red and near infrared channels and five bands of calibrated reflectances (bands 1-2) and brightness temperatures (bands 3-5), and was released as daily and 10-day composite grids.

The UN FAO also conducts regular global remote sensing surveys (RSS) which now complement the FRAs (FAO & JRC, 2012). These are based on object based image analysis (OBIA) of Landsat scenes (Lindquist & D'Annunzio, 2016), and cover pan tropical regions (1990, 2000) and the entire globe (2005). The RSSs analysed 11,000 sampled Landsat scenes, from the United States Geological Survey's (USGS) Global Land Survey dataset (GLS), which were segmented using a region growing multi-resolution algorithm, using spectral values from bands 3, 4 and 5 in the Landsat 4, 5, and 7 sensors. The years of the Landsat products used for each RSS epoch do not describe the exact acquisition year of the imagery used, but rather they are the years around which the best available images were centred.

Improvements in processing power, new sensors, and availability of data, have resulted in a steady refinement of spatial resolution in optical remote sensing studies since then, with the resolution of wall-to-wall forest extent maps in Central Africa dropping to 500m (Hansen et al., 2003), 300m (Arino et al., 2007), 250m (DiMiceli et al., 2011) and more recently 30m (Sexton et al., 2013; Hansen et al., 2013). Persistent cloud cover over tropical forests remains a barrier to frequent mapping of forest extent at finer spatial

resolutions, resulting in wide-area optical cloud free mosaics generated from data spanning several years. This can reduce the accuracy of classifications, and also hinders rapid detection of fine-scale forest change over large areas.

Without standardised definitions of forest cover or land use/land cover classes, it is difficult to compare these products in a meaningful way, and this is compounded by differences in spatial and temporal resolution, minimum mapping unit (MMU), bands used, as well as the methods used for validation and accuracy reporting. Gross et al. (2017) compared the TREES, MODIS-VGT and GFC studies in Sub-Saharan Africa and found differences in forest change at regional and local levels, particularly in medium tree cover densities and in transition zones between the Congo Basin and surroundings, which was suggested to be due to the class boundaries used in the original products. The study by Hansen et al. (2013) is an achievement in terms of Landsat scenes processed and the release of an unprecedented and useful product, although the precise algorithms have not been released, and the interim annual products from which their loss/gain maps were derived have not been released for independent testing.

2.2.1 Central African SAR Forest Mapping

There have been three major projects to map Central African forests using spaceborne SAR sensors. The Central African Mosaic Project (CAMP) used data collected in 1994 from the European Space Agency's (ESA) ERS-1 (European Remote Sensing Satellite) (De Grandi et al., 1999). The Global Rainforest Mapping project (GRFM) used two 1996 mosaics (wet and dry season acquisitions) from the Japanese Space Exploration Agency's (JAXA) Japanese Earth Resources Satellite (JERS-1) (Rosenqvist et al., 2000). Most recently, the Kyoto and Carbon Initiative (K&C) has produced four annual forest/non-forest (FNF) products from 2007-2010 using data from JAXA's Advanced Land Observing Satellite – Phased Array L-band SAR (ALOS-PALSAR) (Shimada et al., 2014), and has continued to produce annual FNF products (2015-onwards) using the currently operating ALOS-PALSAR2 sensor. The utility of a SAR system for wide-area wall-to-wall forest mapping depends on several factors, the most important being the band wavelength used, the polarisation of the sent and received signal, and consistency

of atmospheric and ground conditions (particularly surface wetness) at time of data acquisition.

Sgrenzaroli et al. (2004) compiled a list of eight 1994 forest/non-forest studies that used data from the ERS-1 sensor (the same as was used in the CAMP), with spatial resolutions ranging from 25-100m, all looking at South American sites, and with local rather than regional coverage, three using a speckle filter, two using texture analysis, with various classification techniques, and with validation from Landsat TM scenes and field campaigns. The original papers are not easily accessible, but the methods likely informed later tropical forest studies using C-band SAR.

Differences in ground and atmospheric conditions at the time of acquisition (a mosaicked image is comprised of many smaller near-square scenes, or in some cases a smaller number of long datastrips, all acquired at different times by an orbiting sensor or constellation of sensors) cause problems for automated classification, since SAR is sensitive to the dielectric constant of a target, meaning the more moisture, the stronger a reflection will be. Wet conditions in one scene and dry conditions in an adjacent scene will result in an overall backscatter difference that will need to be accounted for during classification, which is especially difficult if the wet conditions are uneven throughout the scene. The closer in acquisition time (and season) the scenes or datastrips are that comprise a mosaic, the better the chances are that atmospheric and ground conditions will be more stable.

The first two SAR forest mapping projects, CAMP and GRFM, created products at a down-sampled 100 m spatial resolution (due to limited processing power at the time (De Grandi et al., 1999; De Grandi et al., 2000b)) using a single co-polarised SAR band (VV and HH, respectively), while the K&C Initiative has released down-sampled 25 and 50 m maps generated from dual co- and cross-polarised data (HH and HV) (Shimada et al., 2014). CAMP used shorter wavelength C-band data, whereas the GRFM project and the K&C Initiative used longer wavelength L-band SAR data. The acquisition strategy of the JERS-1 sensor was such that adjacent datastrips were acquired within one day of each other. This allowed the GRFM project to use one wet season (high water) and one dry season (low water) mosaic from the same year (De Grandi et al., 2000a). CAMP used

a similarly short time window to acquire all images (44 days) (De Grandi et al., 1999). The K&C Initiative acquired ALOS-PALSAR datastrips over a longer period (June-October) for each year, with missing data occasionally replaced by data from the previous or following years (Shimada et al., 2014).

Despite the longer acquisition time for each mosaic, the dual-polarisation mode (section 2.3.4 explains why dual-polarisation offers greater forest mapping capabilities), long wavelength, higher spatial resolution, and four-year time series of the K&C Initiative mosaics make them the most promising and complete dataset to work with for forest mapping in Central Africa to date. In addition, JAXA's ALOS-2 satellite has continued production of the K&C Initiative annual mosaics after a four-year hiatus. Any techniques that improve forest analysis from the initial four years of K&C Initiative data have the added advantage of being directly applicable to future releases using ALOS-2 data. JERS-1 and ALOS-PALSAR data are also of great importance in mapping tropical forest wetlands, where the sensitivity of L-band SAR to moisture allows the separation of inundated from dry areas for convenient classification (Bwangoy et al., 2010; Betbeder et al., 2014; Ferreira-Ferreira et al., 2015; Rosenqvist et al., 2015; Thomas et al., 2015).

Table 2.2 details nine studies of forest and forested wetlands in the tropics, that use L-band SAR from either JERS-1, ALOS-PALSAR, or both, as well as one study of the ERS-1 CAMP product. There are a wide range of classification methods, and in some cases the SAR dataset is used only as an ancillary one to inform an optical data driven product. The key point of this table is to illustrate the variety of classifications that have been achieved from SAR in this area, as well as the different measures of success (accuracies range from 68% overall accuracy to 97.1%, and in some cases a product accuracy assessment was not performed). The dataset parameters and the general classification methods are described in more detail in sections 2.3 and 2.4, as well as in the papers cited in the table. The study by Bueso-Bello et al. (2017) is not discussed in further detail in this study. Although it is certainly a progressive study and one which is unprecedented in terms of scale, use of multiple parameters from a unique tandem SAR acquisition system, and effort, the methods are somewhat opaque and the TanDEM-X dataset, at time of writing, is unavailable to users and researchers.

Table 2.2: Selected SAR tropical wetland and forest extent studies, from the 1990s to present, including methods, forest definition and reported accuracy

Author(s)	Dataset; Year	Study Area; Res	Classes	Forest definition	Method; Reported accuracy
De Grandi et al., 2000a	CAMP: ERS-1; 1994	Congo Basin; 200m	Swamp, Lowland Forest	30-70% canopy cover (tree cover mosaic)	Amplitude (single polarisation) and normalised variance texture analysis and thresholding; 68% overall accuracy
Sgrenzaroli et al., 2004	GRFM: JERS-1; 1996	Selected Amazon Basin sites; 100m	FNF	>10% canopy cover (the same as that of FAO, 2001)	Edge preserving smoothing, bespoke unsupervised pattern recognition algorithm from single polarisation L-band SAR; 87% accuracy for FNF product
Walker et al., 2010	ALOS-PALSAR, Landsat TM; 2007	Brazilian Amazon; 25m	FNF (including cerrado)	Unspecified, cerrado; >10m tree height	Random forest classifier of backscatter and texture measures, with ancillary topographic data; 92.4% overall accuracy for FNF
Bwangoy et al., 2010	GRFM: JERS-1, Landsat TM/ETM+, SRTM; 1996, 1990, 2000	Congo Basin; 100m	Wetland probability	Tree diameter >20cm	Classification tree bagging, with 80.2% contributions from elevation, Landsat bands 4 and 5, JERS-1 high water, and slope; 81.11% reported overall accuracy for Wetland/Non-wetland classification
Betbeder et al., 2014	MODIS EVI, ALOS-PALSAR, ICESat-GLAS;	Cuvette Centrale, Congo Basin; 100m	Forested Wetland – 4 unsupervised classes	Unspecified, based on water regime	Speckle filtered ScansAR at HH polarisation used to estimate inundation below canopy using unsupervised k-means classifier; statistically similar to thematic map from EVI
Ferreira-Ferreira et al., 2015	ALOS-PALSAR; 2007-2010	Central Amazon Floodplain; multiple resolutions	5 classes, including low and high floodplains	Unspecified forest description	Random forest/OBIA, based on dual polarised data with average, standard deviation and lowest water level backscatter metrics; 83% overall accuracy across 5 classes
Thomas et al., 2015	JERS-1, ALOS-PALSAR; 1996, 2007, 2010	Global mangroves; 25m, 100m	Mangrove, non-mangrove	Unspecified	Lee-filtered for speckle, OBIA of single (JERS-1) or dual (ALOS-PALSAR) pol data. Analysis of maximum Haralick mean texture, temporal range of HV; 97.1% total accuracy of mangrove change across global sites

Author(s)	Dataset; Year	Study Area; Res	Classes	Forest definition	Method; Reported accuracy
Qin et al., 2016	ALOS-PALSAR Mosaics, MODIS NDVI; 2010	Monsoon Asia; 50m	FNF	>10% canopy cover; >90% forested pixels in 500x500 m sample	Decision tree thresholds from dual polarisation (HH, HV) backscatter coefficients based on signature analysis from field photography and Google Earth Imagery (GEI); 95.9% overall accuracy
Bueso-Bello et al., 2017	TanDEM-X; 2011-2016	Global; 50m	FNF	Unspecified	Fuzzy clustering classification of amplitude, interferometric phase and coherence, with biome specific cluster centres; accuracy in tropics (South Amazon) of ~78%- ~92%

2.2.2 Forest Definitions Used

An important consideration for each study is also the definition of forest used (as described in section 1.3). In the studies listed in Table 2.1 and Table 2.2 there are a wide variety of forest definitions used, including those which broadly or closely follow the FAO definition (>10% tree cover) (Lindquist & D'Annunzio, 2016; Sgrenzaroli et al., 2004; Shimada et al., 2014; Qin et al., 2015), those which have a definition that is a range of tree cover values (De Grandi et al., 2000a; Arino et al., 2007), some with very high thresholds (>40% tree cover) (DeFries et al., 2000; Mayaux et al., 2004), and some continuous tree/vegetation cover studies with a reference to just the minimum tree height requirements for inclusion in the forest class (Hansen et al., 2003; Sexton et al., 2013; Hansen et al., 2013) or the minimum tree diameter (Bwangoy et al., 2010). In several cases there is no clearly stated definition of the forest class, with either a reference to other surveys' definitions (Walker et al., 2010; Betbeder et al., 2014), or no mention at all (Ferreira-Ferreira et al., 2015, Thomas et al., 2015). In some cases, the omission or lack of clarity is due to the study not focusing on a binary forest/non-forest classification, or being itself a measure of tree cover, and therefore allowing others to determine a threshold at which to define when a target is forest. In the other cases, however, there is a lack of necessary detail in these studies for meaningful use of the results with respect to the situation on the ground. The lack of clear forest definitions also hinders replicability, in terms of building consistent training and validation datasets.

The sensitivity of SAR to vegetation structure and wetness, described in more detail in section 2.3, suggests that the use of forest definitions relating to canopy cover require the development of empirical models relating the SAR signal to field data. It may be more appropriate to use forest definitions relating to forest structure, such as average stem diameter and biomass, as well as the employment of forest classes relating to hydrological conditions, such that more consistent physical models can be built instead. For optical data, canopy cover is more directly measurable, since the sensors are sensitive to the chlorophyll content of leaves, particularly in the near and shortwave infrared domains.

2.3 SAR Remote Sensing

2.3.1 SAR Basics and Signal Processing

SAR is a side-looking active remote sensing system which transmits microwaves and receives measurements from the signal backscattered from a surface, to produce an image after appropriate processing. The longer wavelengths of microwaves allow imaging of the ground through cloud cover, and as it is an active system, images can also be gathered at night.

In brief, imaging radar systems operate by emitting a pulsed beam to the side of the platform, and measuring the intensity (size of backscatter) and time taken for the signal to return. Signals that take longer to return are further from the sensor in the range direction (to the side, perpendicular to the direction of flight), and as the platform moves forward each pulse gives information about surface objects in the azimuth direction (along its track, or flight path) (Lillesand et al., 2008).

The synthetic aperture of a SAR refers to the method of increasing the spatial resolution of the image in the azimuth (along-track) direction without increasing the physical size of the radar antenna. As the beam angle of a radar is inversely proportional to its antenna (or aperture) length, a longer antenna is required to narrow the beam and improve azimuth resolution (Oliver & Quegan 2004 pp 12-13). Since airborne and spaceborne systems have size and weight restrictions, it is impractical to construct antennae greater than 10-15m in length. To overcome this problem, a synthetic aperture is created from the movement of the platform, while a series of pulses are emitted as the sensor moves along its track. The area illuminated by the beam from each pulse overlaps with the area of preceding and consecutive pulses, with the number of overlaps for a particular point on the ground depending on the speed of the sensor's platform, the pulse repetition frequency (PRF), and the size of the area imaged by each pulse. The pulse information is then processed using a SAR algorithm, most commonly the range-Doppler algorithm (Cumming & Bennett, 1979), which compresses the signal in range and combines all signal returns for each particular point target into a single pixel, to form an image. Range resolution is increased by using shorter pulses, which is limited by the precision of time measurements by the sensor. This enables much higher

resolution imagery than would be possible from a real aperture radar at the distance of a polar orbiting sensor. It is even theoretically possible to place a SAR sensor on the surface of the moon and use the movement of the Earth's rotation relative to the moon to generate the synthetic aperture and have as high as 10m resolution imagery (although this would obviously be a much greater technical feat than the installation of a polar-orbiting satellite and therefore prohibitively expensive) (Guo et al., 2014).

A characteristic of SAR imagery that distinguishes it from optical imagery is the presence of geometric distortions due to topography, whereby up-sloping surfaces facing the sensor appear narrower than they actually are, and down-sloping surfaces facing away appear wider. This is known as foreshortening and is due to the ranged signal, whereby objects oriented more perpendicular to the sensor are therefore closer to the sensor, and appear closer together in the resulting image. In extreme cases, a slope's peak may be closer to the sensor than its base, causing the feature in the resulting image to appear to lean towards the sensor (known as layover). The returned signals of sensor-facing slopes can overlap and are therefore higher than normal, with the opposite occurring in away-facing slopes as the returned signals are more dispersed. In addition, sloped surfaces facing towards the sensor show higher signal return values as the signal is more likely to be reflected directly back towards the sensor, and those facing away from the sensor have low values. With a Digital Elevation Model (DEM), SAR images may be geocoded and geometrically corrected, although correcting for the higher and lower values associated with distortion effects alters the image statistics and results in loss of information (Woodhouse 2006 pp 299). Geometric distortion is pronounced in high relief, upland areas. While some mountainous regions are found in Gabon and eastern DRC, the Congo basin is relatively flat and unaffected by topographic distortion (De Grandi et al., 1999).

Another characteristic of SAR imagery is the presence of 'speckle', which is graininess in the image caused by "interference among the coherent echoes of the individual scatterers within a resolution cell [a resolution cell corresponds to a pixel in the resulting image]" (Woodhouse 2006 pp 289). This can be reduced in a process called multilooking, whereby the synthetic aperture for each pixel is subdivided into a number of smaller sub-apertures which are then averaged. The net result is a reduction in speckle, and

therefore a clearer image, but a corresponding reduction in spatial resolution (de Vries, 1998). For this reason, SAR imagery is almost always used at a lower spatial resolution than what is received by the sensor. It is also possible to reduce speckle with successive images of the same area, in a process known as multi-temporal speckle filtering (De Grandi & Leysen, 1997).

The backscattered signal contains both an intensity and phase component. The intensity is a measure of the strength of the returned signal, and is affected by geometric and dielectric properties (essentially the moisture content) of the surface. The strength of the returned signal for a particular pixel depends on a complex addition of numerous signal interactions with surface scatterers contained within the target area. The phase describes the phase angle of the returned echo, is also a combination of hundreds of interactions with individual scatterers within a target area and is measurable due to SAR being a coherent system with control of the phase at transmission (Canada Centre for Remote Sensing 2007). The phase from a single acquisition is of no practical use, but processing phase information from multiple acquisitions can provide information relating to elevation, surface displacement, and signal decorrelation, which can be related to vegetation growth and density, depending on the SAR wavelength. To produce a scalar intensity image from a single look complex (SLC) SAR image, the sum of the squares of the real and imaginary components of the coherent signal are computed as below, from Small & Schubert (2008):

$$Intensity = DN_{(i,j)}^2 = (SLC_{real})^2 + (SLC_{imag})^2$$

Equation 1

2.3.2 SAR Scattering Mechanisms

There are three major mechanisms by which a SAR signal is scattered from a surface. These are double bounce, volumetric/canopy and surface (or 'Bragg') scattering (Freeman & Durden, 1998). Double bounce scattering occurs when the signal is reflected from two or three orthogonal surfaces, with different dielectric constants, directly back to the sensor. This is common from man-made surfaces, such as in urban environments. Naturally occurring surfaces that cause double bounce backscattering

include vertical tree trunks, particularly those in still water as found in mangrove swamps. Volumetric or canopy scattering is produced from a “cloud of randomly oriented dipoles” (*ibid*), typically seen in leaf and branch interactions in forest canopy. Scattering from a rough surface results in Bragg scattering, with the signal being scattered evenly in all directions. Scattering from still water results in very low signal intensity, as most of the signal is reflected away from the sensor.

2.3.3 SAR Frequency and Wavelength

The wavelengths of SAR range from around 2.4 cm up to 100 cm for bands X to P, as seen in (Rosenqvist et al., 2007). Longer wavelength SARs (i.e. L- and P- band) have a greater ability to penetrate the surface and canopy cover. The signal interacts with objects at the same scale or larger than its wavelength, with smaller objects not affecting the backscatter. As a result, longer wavelength SAR signals pass through leaves and small branches in the upper canopy and offer more information about differences in larger woody material such as stems and large branches, making them more suitable for forestry applications. However, antenna size, power, and accuracy limitations have favoured the use of shorter wavelength spaceborne SAR systems.

Table 2.3: Historic, current, and future spaceborne SAR satellites/sensors, with their wavelength and frequency ranges, adapted from (Rosenqvist et al., 2007).

Wavelength [cm]		3.75	7.5	15	30	100
<i>Radar Band</i>		X	C	S	L	P
<i>Completed,</i>	1990-		ERS-1/2	Almaz	JERS-1	
<i>currently active,</i>	2000					
<i>planned future</i>	2000-	TerraSAR-	<u>ENVISAT-</u>		ALOS-	
<i>and proposed</i>	2010	X, Cosmo	<u>ASAR, SRTM,</u>		PALSAR	
<i>future* SAR</i>		Skymed	Radarsat -1/2,			
<i>spaceborne</i>	2010-	TanDEM-X	Sentinel 1	NovaSAR-S	ALOS-2,	BIOMASS
<i>sensors</i>	2020		A/B, Radarsat		SAOCOM-	
			Constellation		1A/1B,	
					TanDEM-L*	
<i>Frequency [GHz]</i>		8	4	2	1	0.3

2.3.4 SAR Polarisation

Emitted SAR signals are polarised in either horizontal (H) or vertical (V) planes, and the returned signal is similarly received in either horizontal or vertical planes. Co-polarised SAR data (VV - vertical send, vertical receive, and HH – horizontal send, horizontal receive) is generally less useful than cross-polarised (HV and VH) SAR data for forest biomass measurements; a cross-polarised sensor configuration is sensitive to the changes in polarisation produced by volume scattering elements within tree canopy (Mitchard et al., 2011).

For technical and financial reasons, early spaceborne SAR systems offered limited polarisation options. Several more recent spaceborne systems offer a fully polarimetric mode, with all four polarisation combinations (HH; VV; HV; VH). This is achieved by alternating between H and V signal pulses and recording both polarisations in the return signal. Each polarisation combination gives a separate image, with the backscatter intensity showing different properties of the scattering surface. However, spatial resolution is reduced by selecting multiple instead of single polarisation acquisition modes, as the PRF available for each polarisation is reduced. As a result, current operational spaceborne systems acquire mostly single or dual polarised imagery, with fully polarimetric acquisitions reserved for experimental, limited coverages. The technique of analysing multiple polarisations of SAR from a single acquisition is known as Polarimetry.

2.3.5 SAR Interferometry and Coherence

Multiple SAR images of an area, if acquired from roughly the same position in space, and with the same image geometry such as look angle, polarisation, wavelength and spatial resolution, can be combined to take advantage of the signal phase information contained within each complex image, in a process called SAR interferometry (InSAR). Images may be acquired simultaneously by two receiving sensors in single-pass InSAR (e.g. Shuttle Radar Topography Mission – SRTM; TanDEM-X satellites), or at different times by the same or different sensors in repeat-pass InSAR (e.g. The European Space

Agency's (ESA) Environmental Satellite – Advanced Synthetic Aperture Radar (ENVISAT-ASAR) and The Earth Resources Satellite (ERS)-1 and -2). From both techniques it is possible to derive surface elevation from the difference in phase. While the distance between the sensor's (or sensors') positions in space should be sufficiently large to provide sensitivity to signal phase differences, as this distance increases there is spatial decorrelation of the signal, up to a point (the critical baseline) beyond which the phase of each image is completely decorrelated with respect to the other (Zebker & Villasenor, 1992). The critical baseline is proportional to the SAR wavelength, so for longer wavelengths there is less need for orbital manoeuvring to keep within this limit. In repeat-pass InSAR the phase difference also contains information about surface displacement (e.g. from ground subsidence or tectonic processes) between the acquisition dates. With prior knowledge of the topography this component may be removed, allowing calculation of surface displacement by differential SAR interferometry (DInSAR). The phase correlation between two acquisitions determines the reliability of InSAR and DInSAR measurements and is known as interferometric SAR coherence (γ), commonly referred to as coherence. To calculate coherence, two coregistered (where pixels in each image correspond to the same ground target) complex images (S_1) and (S_2) are combined:

$$\gamma = \frac{\langle S_1 S_2^* \rangle}{\sqrt{\langle S_1 S_1^* \rangle \langle S_2 S_2^* \rangle}}$$

Equation 2

In Equation 2 above, the * indicates the complex conjugate of the image, and the chevron brackets show that the spatial average is calculated from a moving window of pixel sample size 'N' (Woodhouse 2006 pp 326).

As well as providing information about the ability to reliably perform InSAR or DInSAR processing, coherence can be used for change detection and land cover classification. The phase element of the SAR signal also temporally decorrelates if there are changes at the same or larger scales than the SAR wavelength. At short wavelengths, high coherence values are seen only in sparsely or non- vegetated areas. For longer wavelength sensors, lower coherence between images indicates the presence of denser vegetation, as scatterer movement between image acquisitions increases with forest

growing stock volume (Tansey et al., 2004). Due to the increase in temporal decorrelation over time, examining repeat-pass interferometric coherence over greater intervals is unsuitable for tropical forest applications, such as the ERS-1 and ERS-2 tandem mission (images were acquired one day apart). Very low coherence values may indicate land cover change such as deforestation between image acquisition dates (Takeuchi et al., 2001). Currently, there are no operational spaceborne sensors offering a revisit time that is suitable for forest change detection from coherence. While the Canadian Space Agency's Radarsat 1 and 2 are still operational, they operate at different frequencies preventing InSAR processing between their imagery. Also, while Radarsat 2 can reduce the revisit time due to its sensor having left or right looking modes, the resulting change in image geometry also prevents InSAR processing. Their 24-day revisit times in the same image geometry are not suitable for coherence analysis of densely vegetated areas. However, the planned three-satellite Radarsat Constellation (due to be launched in 2018) (Canadian Space Agency, 2015; Canadian Space Agency, 2017) may have a short enough revisit time for effective forest monitoring using coherence. The Copernicus Sentinel-1 C-band SAR mission has a short enough revisit time (12 days), which is now halved as both satellites in the constellation are operational (European Space Agency, 2012). Use of L-band or longer wavelength SAR has a higher chance of maintaining coherence over a longer revisit time, due to the relative stability of the dominant scattering elements (tree trunks and branches) (Watanabe & Shimada, 2006). Also, shorter wavelength spaceborne SAR sensors such as TanDEM-X and Cosmo/SkyMed (both X-band systems with a wavelength of around 3cm), may be suitable for coherence analysis of low biomass forested areas, such as in Boreal forests (Weydahl & Eldhuset, 2012).

2.3.6 Polarimetric SAR Interferometry (PolInSAR)

Combining the techniques of Polarimetry and InSAR, using fully-polarimetric data from multiple acquisitions, can provide information about forest structure, including tree height. This is based on separating, or decomposing, the signal into coherent scatterers in order to estimate the height at which these dominant scattering processes occur (Cloude & Papathanassiou, 1998). Since different polarisations interact more strongly

with different scatterers, such as canopy (HV) and trunk (HH), this allows calculation of upper and lower tree structure heights. This technique is not yet suitable for wide area analysis, as it is currently constrained by the limitations of both InSAR (temporal decorrelation between revisit times and lack of appropriate long wavelength spaceborne SAR sensors) and Polarimetry (too few fully polarimetric spaceborne SAR sensors and a similar lack of long wavelength SAR sensors). The problem of temporal decorrelation in the recent ALOS-PALSAR (Advanced Land Observing Satellite - Phased Array type L-band SAR) has been outlined in (Cloude et al., 2011) among other studies during its years of operation (2006-2011). However, there have been numerous studies using airborne campaigns that have demonstrated the potential of L- and P-band PolInSAR (Papathanassiou et al., 2008) in future SAR sensors dedicated to PolInSAR, such as the proposed P-band BIOMASS mission (Le Toan et al., 2011), the proposed TanDEM-L mission (Moreira et al., 2011) and improved revisit time from 46 days for ALOS to 14 days for ALOS-2 (Shimada et al., 2011).

2.3.7 Sensitivity of SAR to Moisture

Radar backscatter is sensitive to the dielectric properties of a surface target. In the context of a ground scatterer, the dielectric constant of a material is *“a fundamental property that characterises the reflection and attenuation properties of a wave interacting with that material”* (Ulaby et al., 1996). This effectively describes how a pulse of microwave energy (i.e. a SAR signal) will interact with a particular surface, in terms of how much energy will be reflected, in which direction or directions, and also how deeply the signal will penetrate the surface. Ulaby et al. (1996) demonstrate that the dielectric constant is strongly correlated with soil moisture, unless the temperature is below freezing point, and as a result soil moisture can be modelled and estimated from radar backscatter. The dielectric constant can also be used as a method of field soil moisture measurement, using a calibrated (for soil type) capacitance probe (Dean et al., 1987). There are several remote sensing derived soil moisture products that are based on spaceborne SAR or other passive or active Radar systems. These systems include the L-band Soil Moisture Active Passive (SMAP) and Soil Moisture and Ocean Salinity (SMOS), C- and X-band Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), and C-band Advanced Scatterometer (ASCAT) spaceborne

sensors. It has been shown that shorter wavelengths (i.e. X- and C- band) have a lower sensitivity to soil moisture than longer wavelengths (L-band), particularly in vegetated areas (Escorihuela & Quintana-seguí, 2016; Brocca et al., 2011). This would suggest that, for the Congo Basin, SMAP and SMOS would be more suitable. However, the availability and temporal resolution of SMOS (launched in 2009, with a revisit time of 3 days at the equator) and SMAP (launched in 2015), render them unsuitable for this study. Lower suitability in vegetated areas, shorter wavelength and coarse resolution make the AMSR-E soil moisture data difficult to work with in tropical regions. There is a product, derived from the C-band horizontal and vertical polarisation brightness ratios on AMSR-E, which is linearly related to the vegetation water content (Njoku, 1999), which was also used by Lucas et al. (2010).

In the absence of reliable soil moisture information in the Congo Basin and other heavily vegetated regions, it should be possible to use rainfall data within basins and river catchment areas to model changes in surface wetness. In the African continent, it is only possible to consistently monitor continental rainfall using spaceborne sensors due to the sparseness of rain gauging stations in many parts (Cohen Liechti et al., 2012). The most common methods involve the use of Passive Microwave (PM) and Visible and Infrared (V/IR) data from sensors on geostationary orbiting platforms such as Meteosat, and coarse resolution sensors on polar orbiting satellites such as SSM/I. Some available products that report rainfall in the Congo Basin are described below:

- The National Oceanic and Atmospheric Administration/Climate Prediction Centre (NOAA/CPC) Famine Early Warning System (FEWS) African Rainfall Estimation Algorithm Version 2 product (RFE 2.0). This is a 0.1° resolution (pixel spacing of roughly 11 km x 11 km at the equator) daily rainfall product based on four sources: ground based rain gauge measurements at between 500 and 1000 weather stations (depending on maintenance and availability of data); up to four times daily radar measurements from the Special Sensor Microwave/Imager (SSM/I); four times daily microwave measurements from the Advanced Microwave Sounding Unit – B (AMSU-B) sensor on NOAA meteorological satellites; half-hourly GOES Precipitation Index (GPI) rainfall amounts derived from Meteosat IR cloud top temperatures. The algorithm

merges data using a maximum likelihood estimation method, with a weighting coefficient based on the input data's random error (NOAA Climate Prediction Center, 2006).

- NASA's Tropical Rainfall Measuring Mission (TRMM), specifically the product 3B42 version 6, is a monthly product at 0.25° resolution covering the globe between latitudes of 50° N and 50° S that merges 3-hourly estimates from IR and PM sensors on its satellite platform (also referred to as TRMM), and calibrates them to a monthly product using Global Precipitation Climatology Centre (GPCC) data.
- The NOAA/CPC morphing technique (CMORPH) uses similar input data to TRMM 3B42 but does not merge PM and IR data, instead replacing any absent PM data with an alternative source of half-hourly geostationary IR data (Joyce et al., 2004). The product is effectively an interpolated combination of several PM and IR datasets to improve resolution (up to 8 km and half-hourly temporal resolution), but archive products (before the previous month) are only available at 0.25° resolution (the resolution of the individual inputs used) and in 3-hourly intervals (CPC, 2017).

A comparison of 10 different African rainfall products showed a poor performance for the RFE2.0 product in areas of extreme topography (Dinku et al., 2007). The comparison used aggregated data, both spatially and temporally to account for differences between datasets. While the RFE 2.0 product did not perform as well as the other datasets, for the purposes of this study, its availability, and spatial and temporal resolution outweigh the lower performance in areas of higher topographical complexity.

The effects of different land surfaces on the scattering mechanisms of SAR at different wavelengths and polarisations are observable, and especially so at coarser resolution when speckle noise is reduced by filtering the images. With knowledge of precipitation patterns, season of image acquisition, and having dual polarisation (both co- and cross-polarised images), differences in vegetation moisture may have an impact on an automated SAR land cover classification. This is seen in forest extent products produced from SAR imagery, and the effects are pronounced in areas that experience seasonal flooding, particularly in tropical areas and over wide regional maps, as wide area

mosaics are often compiled from images that are acquired at different seasons and in some cases in different years. Flooded forests reflect strongly in L-band SAR at HH polarisation, a characteristic that may be useful in either improving the classification of new forest extent maps, or in adding depth of information to forest extent products. In the absence of field soil moisture measurements and local rainfall surveys, precipitation values derived from spaceborne meteorological observations and used to interpolate between other ground based measurements at the time of the image acquisition are important in order to model likely moisture conditions for a study of their effects on L-band SAR backscatter (Lucas et al., 2010).

There are several documented scattering mechanisms from flooded vegetation and other wetlands that should be taken into account when studying a returned SAR signal. Standing water in the absence of vegetation can appear dark due to specular reflection, but if there are stems, a double bounce effect may reflect energy directly back to the sensor. Waves or shallow running water can act as a rough surface. Incidence angle affects the penetration of the signal through upper canopy, with lower incidence angles facilitating the detection of flooding beneath the canopy and volume scattering in the canopy can mask this effect at higher incidence angles (White et al., 2015).

2.3.8 SAR Spatial Resolution vs Speckle

As mentioned in section 2.3.1, there is a trade off in SAR processing between the level of speckle and spatial resolution. Depending on the process used to remove speckle, as well as the forest structure and distribution in the study area, this can have effects on both the accuracy and validity of forest maps produced, as well as the overall area of forest reported (Santoro et al., 2009). Studying the changes in forest reporting at different spatial scales allows both the selection of the optimal resolution at which to generate forest products, and provides information about the uncertainty of forest products generated from sub-optimal spatial resolution data. There are three commonly used methods for removing speckle in SAR images. These are multilooking, single image filtering, and multi-image (or multi-temporal) filtering. A measure of efficiency of a filter is the Equivalent Number of Looks (ENL) which is determined by

calculating the square of the mean divided by the standard deviation of the multilooked image (Im):

$$ENL = \left(\frac{Im_{mean}}{Im_{std}} \right)^2$$

Equation 3

ENL corresponds to the effective number of statistically independent looks (de Vries, 1998), and indicates the level of speckle, although it does not contain information about the spatial resolution of the resulting image.

The reduction of speckle by multilooking involves selecting a number of looks in range and azimuth over which the original detected data will be averaged. In azimuth, the number of looks effects an increase in the synthetic aperture length . Typically, the number of looks in both directions will be chosen to produce an approximately square pixel sample size in ground range. This is calculated using the range spacing, azimuth spacing and mid range incidence angle as below (usually the number of azimuth looks are selected, and the number of corresponding range looks is rounded to the nearest integer):

$$\frac{nAzimuthLooks}{nRangeLooks} = \frac{RangeSpacing}{AzimuthSpacing \cdot \sin(IncidenceAngle_{mid})}$$

Equation 4

Multilooking reduces variance while maintaining the same average values as the single look data, which has the effect of increasing radiometric resolution while reducing geometric resolution (de Vries, 1998).

SAR processing software packages offer several single image filtering algorithms of varying complexity, and with applicability depending on the distribution of the speckle noise. A filter type is chosen, along with the size of the moving kernel which the filter is applied over. A larger kernel increases the ‘smoothness’ of the filtered image, but results in a greater loss of information. The following is a list of commonly used filters:

- Mean filter – The most basic filter, simply averages speckle over a kernel.

- Median filter – Also simple, removes signal spikes using the median value over a kernel.
- Lee / Lee Sigma – These filters assume a Gaussian distribution of speckle noise; the Lee filter assumes the mean and variance of the pixel of interest is equal to the local mean and variance of the pixels in the associated kernel and calculates a local variance- and mean-weighted average over the kernel; the Lee Sigma filter calculates the mean of values in the kernel that fall within a two standard deviation range (Mansourpour et al., 2006). The aim of these filters is to increase edge detection
- Frost – This filter also produces a local variance weighted average, but the value is also weighted by distance from the central pixel in the kernel. It assumes multiplicative-convolved noise in the speckle, whereby the “*desired information is multiplied by a stationary random process which represents the effects of coherent fading*” (Frost et al., 1982).

Differences between these methods are visually subtle on smaller kernels (i.e. a 3x3 moving window), but become more apparent with larger sized kernels. Generally more complex algorithms aim to preserve and enhance edges, while reducing loss of radiometric and spatial resolution.

A multi temporal speckle filter is an application of one of the above filters through a stack of coregistered SAR images. The algorithm used in this study to generate a speckle filtered image J is shown below, for a sequence of N images, at position (x,y) in the k th image (i.e. $I_k(x,y)$):

$$J_k(x, y) = \frac{E[I_k]}{N} \sum_{i=1}^N \frac{I_i(x, y)}{E[I_i]}$$

Equation 5

for $k = 1, \dots, N$, where $E[I]$ is the local mean value of pixels in a window centred at (x,y) in image I . This effectively calculates a mean value weighted by the sum of the means through the image stack, and can reduce speckle without a reduction in spatial resolution.

2.4 Classification Methods

There are many different methods for generating meaningful products from remote sensing imagery. These methods can broadly be broken into two types, supervised and unsupervised classification. Supervised classification requires the user to provide a training sample of known features in the area covered by a remote sensing dataset, which are then used to derive statistics to subdivide the feature space (the data being evaluated in each band of an image) into the classes given by the training sample (Singh, 1989). Unsupervised classification uses cluster analysis of the feature space to delineate the remote sensing data into a number of separable classes, which are then linked to a target class by the user (Lu & Weng, 2007).

The generation and analysis of statistics from a training dataset in supervised classification of remote sensing imagery directly overlaps with the problem of data analysis in machine learning, which has applications from biomedical imaging to automatic facial recognition. Fernández-Delgado et al. (2014) compared 179 available classifiers over 121 varied and commonly tested (in the field of machine learning) datasets, including remote sensing imagery, and found implementations of Random Forest (RF) (Breiman, 2001) and Support Vector Machine (SVM), from Chang & Lin (2011) based on Cortes & Vapnik (1995), to perform most reliably across the numerous datasets. Lu and Weng (2007) generated a taxonomy of classification methods used in remote sensing studies, noting the differences between parametric and non-parametric approaches. Parametric approaches assume the data are Gaussian (normally distributed), and can result in noise in complex landscapes (Lu & Weng, 2007). Non-parametric approaches do not require this assumption, instead separating classes without specifying statistical parameters.

Three methods of supervised classification were used throughout this thesis: in Chapter 4 a rule based thresholding classifier was used; Chapter 5 used an SVM classifier; 6 employed both RF and SVM machine learning classifiers. A description of these classifiers and other studies that employed them follows.

2.4.1 Thresholding Classifier

The simplest of the three classification methods used, a thresholding classifier is a basic method whereby a series of binary decisions separate classes using single or multilayer input. It requires user defined thresholds from each input layer, usually described with simple arithmetic and Boolean operators. As such an accurate decision tree classification is reliant on classes that are clearly separable with the input data, as well as knowledge of the thresholds to be used, through a priori information, training data and image statistics (Singh, 1989). Object based image analysis classifiers can use thresholds from additional features alongside pixel values, generated from segmented imagery, relating to segment shape, size and statistics. From Table 2.1 and Table 2.2, a thresholding approach was used in the GLS 2000 product to classify based on NDVI (Mayaux et al., 2004 – 68.6% accuracy across multiple landcover classes). It was an integral part of the JAXA Kyoto & Carbon Initiative FNF processing chain, with threshold values calculated in different global regions to delineate forest and non-forest classes, as well as to separate the urban and water body classes (Shimada et al., 2014 – 84.86% - 91.25% reported overall accuracies), as well as the study by Qin et al. (2015) in monsoon-affected Asian countries (95.9% overall accuracy reported).

2.4.2 Support Vector Machine Classifier

SVM is a supervised non-parametric classification algorithm that takes training data, typically from two classes, and generates a multi-dimensional decision surface referred to as an *optimal hyperplane* that best divides the input values, where the separability between classes is greatest (a simple two-dimension example is shown in Figure 2.1). A non-parametric approach like this is useful in remote sensing, where data tends not to be normally distributed (Mountrakis et al., 2011).

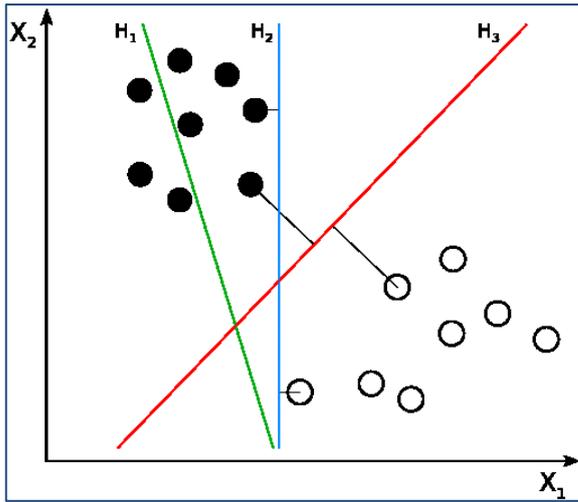


Figure 2.1: Graphic showing how a support vector machine would choose a separating hyperplane for two classes of points in 2D. H_1 does not separate the classes. H_2 does, but only with a small margin. H_3 separates them with the maximum margin. Image and caption by [user:ZackWeinburg](#) licensed under Creative Commons [CC-BY-3.0](#)

A more detailed description of SVM, as well as a description of its implementation in code as used by this study (LIBSVM) is published in Chang & Lin (2011). The LIBSVM classifier used offers the choice of four basic kernel functions, which in the case of machine learning relate to a function that maps training data to a higher dimensional space, allowing the SVM to find a linear separating hyperplane (Hsu et al., 2010). These four kernels are linear, Gaussian Radial Basis Function (RBF), polynomial and sigmoid, described in more detail by Hsu et al. (2010), and there are recommended situations where each are appropriate. For a study with a high number of features, a linear kernel is advised, and RBF is advised where there are fewer features. RBF is typically used as the default kernel. Each kernel type has tunable parameters that can be optimised by cross-validation using the testing data (*ibid*).

2.4.3 Random Forests Classifier

Random forests is a non-parametric tree-based model developed by Breiman (2001) that can be used for regression (in this case meaning classifying to a value on a continuous scale, such as percent tree cover or AGB) or discrete (classifying data into one of a fixed number of classes) classification, and is frequently used in remote sensing studies of biophysical parameters. It is considered a computationally fast classifier, especially for large noisy datasets and multiple classes with outliers. It is an evolution of

a method known as ‘bagging’ where bootstrapped samples (random samples, collected with replacement or potential reselection) of a training dataset are used to train a large number of classifiers (random trees), from which an aggregated classifier is chosen (Gislason et al., 2006). The bagging process ensures that outliers in the training dataset do not have a large effect on class separation. Each random tree is a series of binary decision nodes (the number of node levels, or tree depth, is a parameter selected by the user) that best separates the samples into their respective classes based on a random selection of features. All of the bagged random trees are tested on the remaining ‘out-of-bag’ (OOB) training data that have been withheld from the bootstrapping process, and from the ensemble of the trees a majority vote selects the class (or an average in the case of a regression) (*ibid.*). Various parameters are selectable that affect the processing time, generalisation, and predictive accuracy of the classifier (Criminisi et al., 2011). Table 2.4 describes the main parameters and their significance to the classifier. Random Forests has performed well in numerous remote sensing land cover classification studies, particularly in terms of computational efficiency, accuracy and ease of use (Ghimire et al., 2012; Fernández-Delgado et al., 2014; Gislason et al., 2006). Bwangoy et al. (2010) used a technique similar to random forests, classification tree bagging, with positive results (81.11% reported overall accuracy for Wetland/Non-wetland classification). Two other studies in Table Table 2.2 demonstrate the popularity and effective use of random forests in SAR classification of forests and wetlands (Ferreira-Ferreira et al., 2015 – 83% accuracy across five classes; Walker et al., 2010 – 92.4% accuracy for a FNF product).

Table 2.4: Random Forests Parameters; descriptions and significance adapted from OpenCV, (2017) online documentation

Parameter	Description/Significance
Tree Depth	Number of node levels permissible; low values may result in underfit classifiers, and high values may result in an overfit
Min. number of samples per node	Recommended small percentage of total training data; determines whether a node will terminate as a 'leaf', or split
Number of random features selected at each node	Determines how many random features (or input layers) will be selected to split each node in a tree; defaults to the square root of the number of features to reduce tree complexity
Maximum number of trees in forest	Determines the number of random trees that will be calculated; affects the processing time (linearly)
Sufficient accuracy	Sets an accuracy threshold for testing on out of bag training data

It is challenging to compare the accuracies and efficiencies of the studies described in Tables 2.1 and 2.2 with reference to the classification methods used, as they are mostly performed with different datasets, timeframes and locations. The accuracies are also in several cases not directly comparable, with different landcover classes being assessed, and varying types of validation data (from widely sampled ground truth data to analysis of optical imagery). Comparative classification studies in the SAR domain such as that of Tewkesbury et al. (2015), which investigated different change detection approaches using optical sensors, are rare.

Chapter 3: Study Sites

This chapter is a description of the study sites analysed in Chapters 4, 5 and 6.

The methodology throughout this study involves the analysis of trained supervised classifications (described in detail in section 2.4) of spaceborne radar and optical imagery. Due to the remote locations of data, and the lack of widespread field data in the Congo basin, high and medium resolution optical imagery was used to generate training data for several classes. In some areas, field data were available, but for forest classification the number of samples is too low and spatially clustered to provide adequate coverage. For optical based training data, random point samples were generated within the study areas, and 100 metre square boundaries were produced around the points, allowing for a fixed one hectare (ha) sample with each plot. This is also consistent with the format of the limited ground data available. In addition, an expert led sampling strategy was performed on smaller scenes, whereby training points (at individual pixel level) for the major classes were selected subjectively, and a region growing algorithm was automatically applied to increase the size of the training area from the selected points.

The Congo rainforest often refers to all of the forested areas of Central Africa, rather than specifically those contained within the drainage area of the Congo river (Justice et al., 2001; Hansen et al., 2008) and it either wholly or partially covers the following six countries: Cameroon, the Central African Republic, Equatorial Guinea, Gabon, the Democratic Republic of the Congo (DRC) and the Republic of the Congo. Of these, the DRC and the Republic of Congo are countries with active UN-REDD national programmes, and Cameroon, the Central African Republic, and Gabon are REDD partner countries. The general climate of the Congo basin is warm and humid, with a low temperature range (2°C), an average temperature of 25°C and rainfall averaging 1800mm/year (Bwangoy et al., 2010).

3.1 Description of Study Sites

The JAXA ALOS PALSAR imagery obtained for the study described in Chapter 4 covers a rectangular area of Central Africa, bounded in the North-West by coordinates (Lat. 008°00'N, Long. 008°00'E) and in the South East by coordinates (Lat. 010°00'S, Long. 038°00'E), seen in tiles A13-A16 and A19-A22 in Figure 4.2. The tiling constitutes the original format of data delivery for the K&C Initiative in Africa, and the tiles for this study were selected to completely cover the Congo Basin. This covers an area of around 6.7 million km², although a large proportion of this consists of water bodies. Apparent in Figure 3.1 is the fact that this study area is divided roughly equally on either side of the equator, leading potentially to seasonal differences throughout the dataset even at concurrent data acquisition. There are two rainy and two dry seasons; above the equator there is a milder rainy season between September and November and a major one from March to May, and below the equator the minor rainy season from February to May and a major one from September to December (Samba et al., 2008). From SRTM topographic data, the elevation ranges from 0 at the coast to peaks of over 4000m in the east of the DRC (the Virunga mountain range) with a maximum of 5034m above sea level. With the exception of a few mountains such as Mt. Cameroon (4070m), most steep terrain in the region is in the east of the study area. Since this is a very large area, and is covered by a variety of land cover types, and an optical product would be combined with a SAR dataset, it is appropriate to define the forest using the upper threshold of tree cover (30%) as stated in the FAO FRA. A minimum mapping unit (MMU) of 1 ha is appropriate given the validation dataset used.

Analysis for chapter 5 was performed on a 5° by 5° square with an upper left coordinate of Lat. 005°00'N, Long. 015°00'E. The more recent release (January 2014) of K&C Initiative JAXA ALOS PALSAR data is organised into grids of 5° mosaicked squares subdivided into 25 1° by 1° tiles. The selected area covers part of four countries: Cameroon, the Central African Republic, the Republic of Congo, and the Democratic Republic of Congo, and contains a wide range of landcover types, including: rainfed croplands, mosaic cropland/vegetation, broadleaved deciduous forest, shrubland, herbaceous vegetation (savannah), permanently and regularly flooded forest, urban

areas and water bodies¹. Topography is for the most part flat, with a relatively small elevation range, for such a large area, from 279m to 870m above sea level. The forest definition used for chapter 5 is based on the training dataset, which used 200 x 200m points classified using high resolution optical data and had a requirement that at least 50% of the plots were covered by trees. Since this was an empirically trained classification, forest was defined as greater than 50% tree cover for this study. A MMU of 1ha was used to make best use of the available resolution (50m) while removing single isolated class pixels.

Chapter 6 is intended to answer research questions 2 and 3 outlined in section 1.4. This therefore requires a L1.1 SLC SAR product to allow processing and analysis at multiple scales. To capture seasonally inundated forest and compare with other forest types requires a long enough time series with imagery from different months of the year, and close proximity to a water body. This was achieved by selecting a nine-image time series of a single scene of L1.1 data (rather than a larger area with a shorter temporal coverage) located in a wetland forest and rural complex spanning the Congo River in the Mongala district of Équateur province in the DRC. The GlobCover 2009 forest classes in this area are 30, 40, 160 and 210 (mosaic vegetation/cropland; closed to open broadleaved evergreen and semi-deciduous forest, closed to open broadleaved forest regularly flooded, and water bodies). The smaller study area facilitated processing and removed potential issues with mosaicking data in adjacent scenes. The topography of this study area is relatively flat, and ranges from 320m to 480m above sea level. The forest definitions for chapter 6 were fixed to the training and validation datasets, which were year specific, generated at the coarser of four resolutions (60m) and tested using higher resolution (15m) data, allowing a 50% canopy cover definition to be used. As well as this definition, the flooded forest class was constrained by distance to water bodies (within 4km). MMUs were dependent on scale, although the removal of isolated clusters

¹ Landcover classes taken from ESA's GlobCover 2009 product, (Bontemps et al., 2010)

of pixels with a majority filter meant that MMU values of 0.55ha, 1.13ha, 2.53ha and 1.62ha were used respectively for the scales of 15, 30, 45 and 60m resolution.

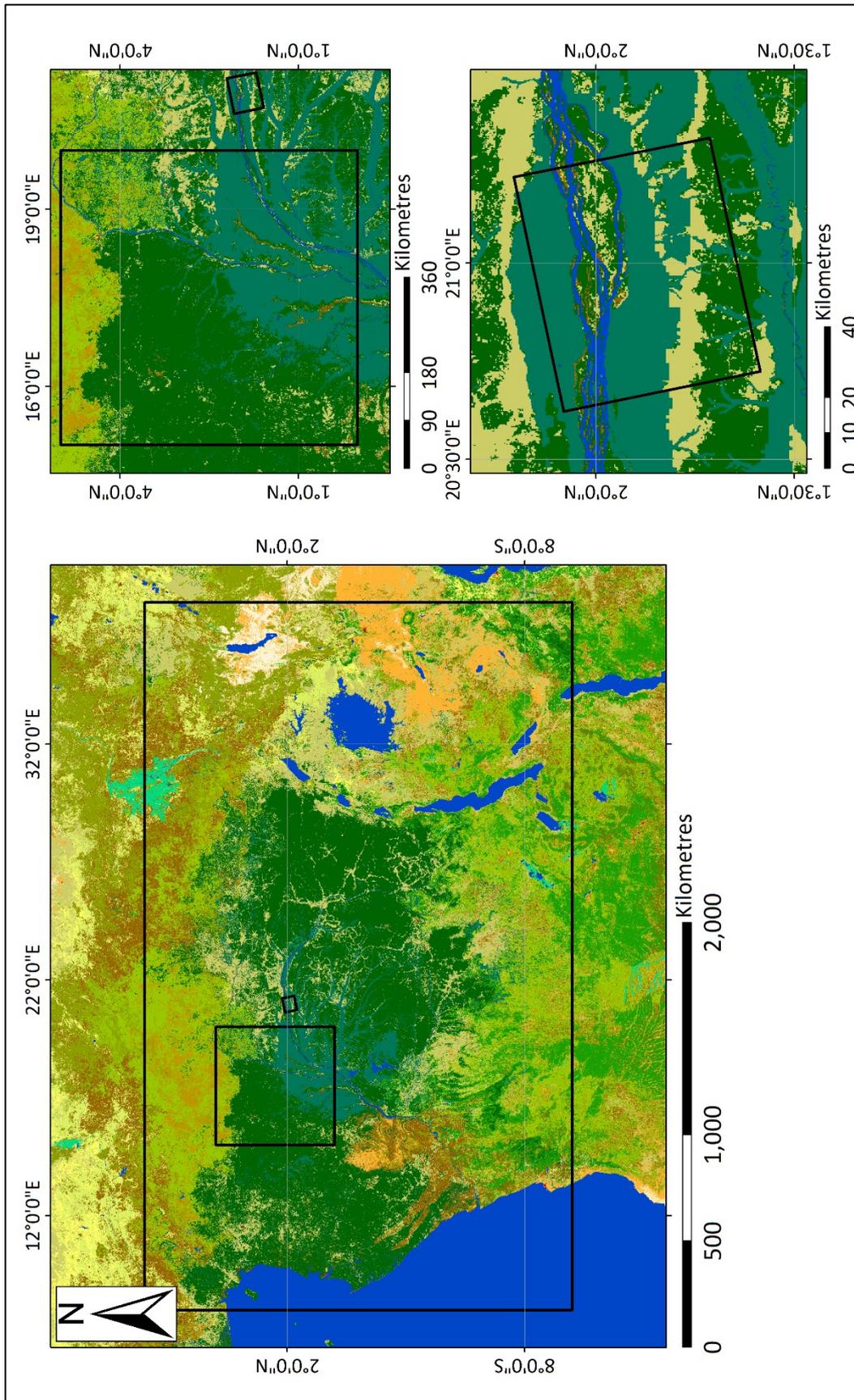


Figure 3.1: Study sites and relative location with corresponding GlobCover 2009 dataset (complete legend in Figure 4.5). Top left - Chapter 5 study site; Top right - Chapter 6 study site; Bottom - Chapter 4 study site.

Chapter 4: New Map of Forest Cover in the Congo Basin Derived from L-band SAR Mosaics and Existing Optical-based Forest Tree Cover Data

4.1 Introduction

This chapter is a description of a study of at-the-time available optical- (Landsat treecover continuous fields from Sexton et al., 2013) and SAR- (K&C Initiative from JAXA, 2010) based mosaic imagery. It is intended to inform the problems associated with both types of data in this region, as well as analyse the suitability of the K&C Initiative data products.

4.1.1 Aim and Objectives

The aim of this chapter is to partly answer the first research question from section 1.4:

“What are the ideal image acquisition parameters and classification techniques for an annual forest map in the Republic of Congo from spaceborne SAR data, given currently available data and regional seasonal effects on image quality?”

In particular, the aim is to explore available spaceborne L-band SAR data and through analysis suggest improvements in the acquisition and processing of the data. To do this the following three objectives are stated:

- To obtain and characterise available L-band SAR data in relation to other available map products in the Congo Basin
- To classify the data and compare with existing forest cover products
- To analyse the results and suggest improvements in data acquisition, release and processing

4.2 Datasets

4.2.1 ALOS PALSAR

Until January 2014, only version 1 of the K&C Initiative ALOS-PALSAR mosaic was available. This included two 50 m resolution products, one from 2008 and the other from 2009, both derived from the backscatter intensity of the fine beam dual polarisation (HH and HV) acquisition mode. The 2009 mosaic was used to fulfil the aims of this chapter. The imagery obtained covers a rectangular area of Central Africa, bounded in the North-West by coordinates (Lat 008°00'N, Lon. 008°00'E) and in the South East by coordinates (Lat. 010°00'S, Lon. 038°00'E), seen in tiles A13-A16 and A19-A22 in Figure 4.2. The tiling constitutes the original format of data delivery for the K&C Initiative, and the tiles for this study were selected to completely cover the Congo Basin. They were mostly acquired over a two-month period to reduce the effects of seasonal differences across the data strips, although in several cases data were used to gapfill from adjacent years. The data released by the K&C Initiative were pre-processed to backscatter intensity from raw data using SigmaSAR, a proprietary SAR processor.

The full processing steps taken are described in detail in De Grandi et al. (2011). The remainder of this paragraph outlines the steps relevant to this study. Long acquisition strips (70 km in range and up to 3000 km in azimuth) were multilooked to 16 looks in azimuth and 4 looks in range. Individual datastrips were radiometrically calibrated to account for linear signal drops from near to far range, and a separate, inconsistent complete signal loss in parts of some strips. For mosaicking, between-datastrip temporal differences in vegetation water content, soil moisture, and vegetation cover were corrected using an inter-strip balancing algorithm. This used the along-track profiles of the overlapping area of both adjacent strips and first corrected strips with anomalously high or low digital number (DN) values (decided using discrepancy analysis and an interactive procedure) using a gain function derived from the discrepancy in the left and right adjacent profiles. Profiles and discrepancies of corrected strips were recomputed and a linear gain function was applied across all strips for the mosaicking procedure. The mosaicked products were geocoded using the NASA SRTM 3 arcsecond global DEM product (described in section 6.2.4), but not radiometrically terrain corrected, to aid visual interpretation for thematic applications.

While the within- and inter-strip radiometric corrections resulted in a more homogenous signal across the mosaics, there are several marked backscatter differences between adjacent data strips throughout the data, as well as visible relief due to uncorrected topography (both demonstrated in Figure 4.1). Metadata from these earlier K&C PALSAR mosaics relating to the acquisition date and time of each individual data strip was not easily matched to available vector files showing the footprint of each acquisition. Without information relating to the footprint and incidence angle of each individual datastrip, terrain correction with this version of the K&C Initiative data was not possible, and it was not possible to match acquisition time with soil moisture measurements or other ancillary datasets that relied on a precise date.

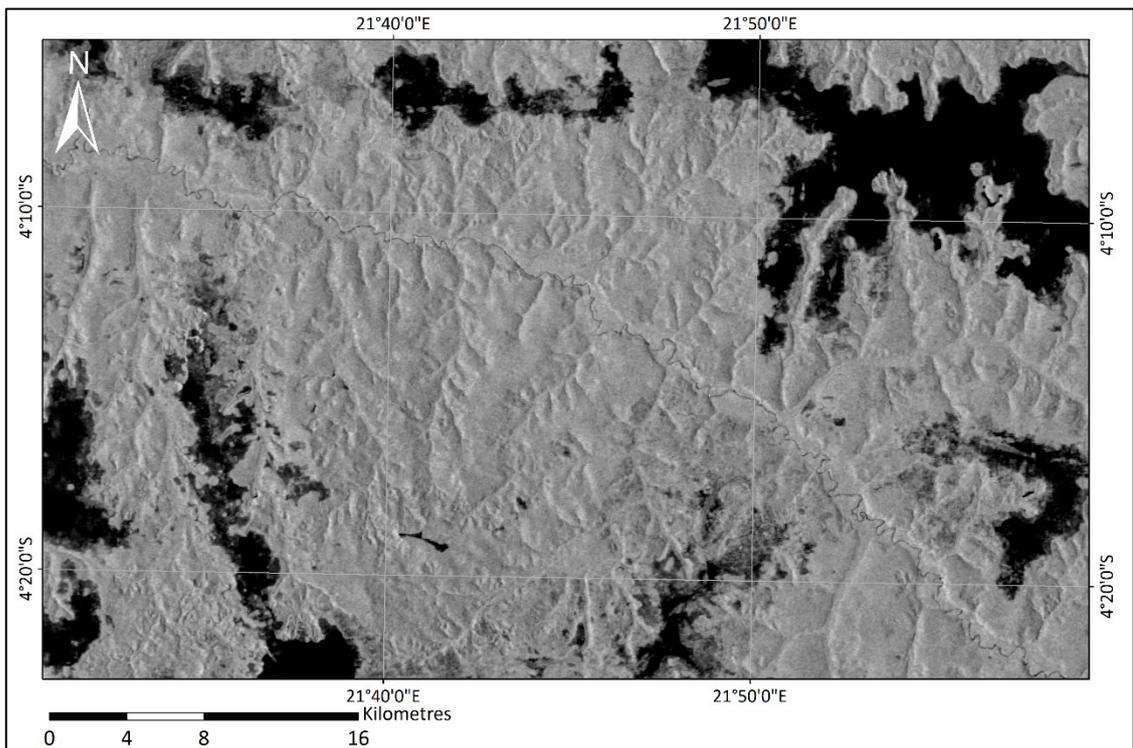
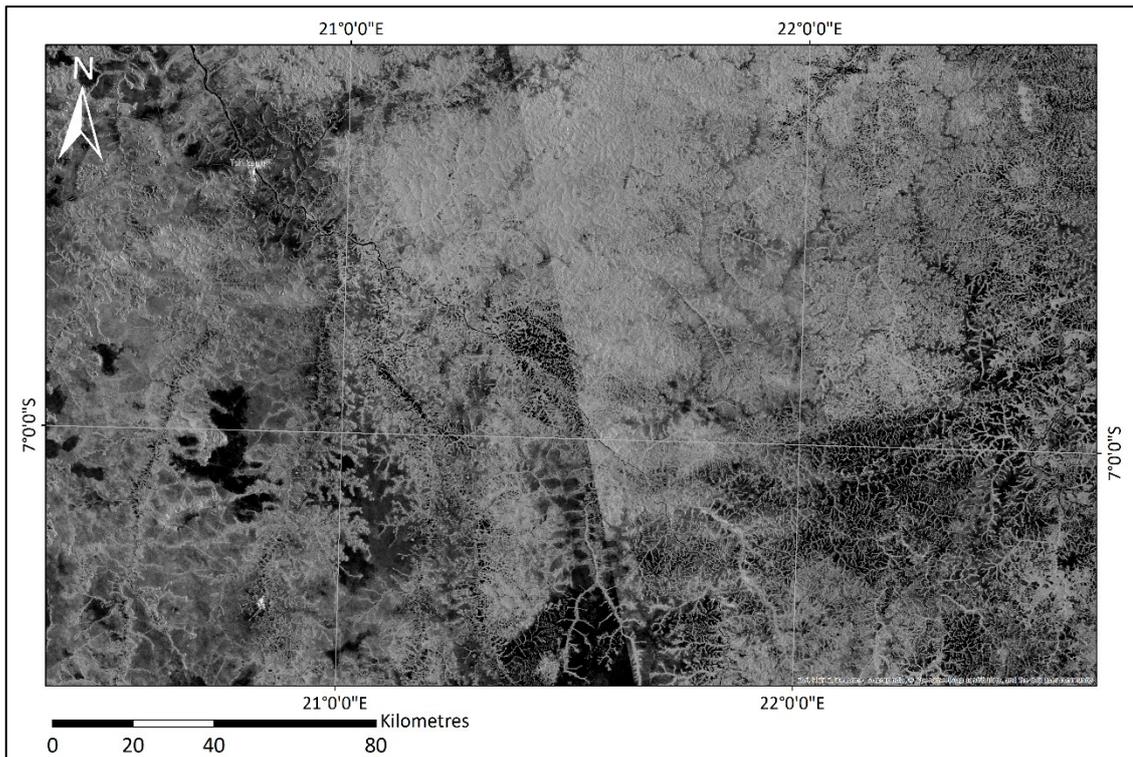


Figure 4.1: Samples (different scales) of the K&C Initiative ALOS-PALSAR 50 m 2009 HH mosaic; top showing a radiometric difference between data strips in Lulua Province, DRC; bottom showing uncorrected relief in an area of Kasai Province, DRC. The map projection is Universal Trans Mercator (UTM) 32.

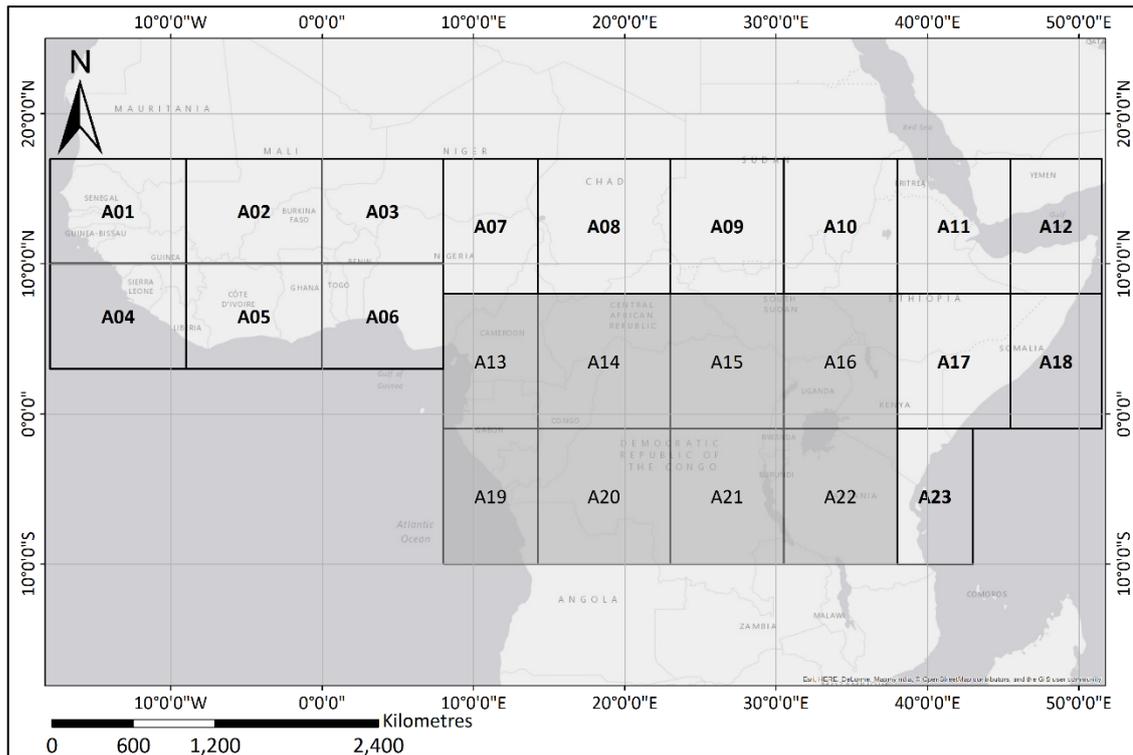


Figure 4.2: K&C initiative ALOS-PALSAR 50m mosaic Version 1 availability over Central and West Africa. Used tiles (A13-A16 and A19-A22) are shaded. Background maps from ESRI ArcGIS software used under academic licence courtesy of Esri, HERE, DeLorme, MapmyIndia, © OpenStreetMap contributors, and the GIS user community

4.2.2 Landsat Tree Cover Continuous Fields (LTC) and Moderate Resolution Imaging Spectroradiometer - Vegetation Continuous Fields (MODIS-VCF)

Freely available Landsat optical data at 30m spatial resolution is available from the United States Geological Survey (USGS) covering the Congo basin. A 30m spatial resolution Landsat and Moderate Resolution Imaging Spectroradiometer (MODIS) - derived dataset, Landsat Tree Cover Continuous Fields (LTC), was available with coverage for the year 2000 (Sexton et al., 2013). It was derived from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) surface reflectance data from the 2000 epoch Global Land Survey (GLS) dataset, which used input data from 1999 to 2002. A sample of Landsat data were rescaled to 250 m resolution, the same spatial resolution as the MODIS-VCF product (DiMiceli et al., 2011). This was used to generate a training dataset from a corresponding MODIS-VCF sample, and a regression tree model (a Cubist model

(Quinlan, 1993) as cited by Sexton et al., (2013)) was then fit and applied to full resolution Landsat data.

MODIS VCF (DiMiceli et al., 2011) is itself derived from 16-day surface reflectance composites at 250 m resolution. Training data from medium and high resolution sensors (Landsat 5, 7, Ikonos and Quickbird among others) are generated and then resampled to 250 m to create a 0-100 scale of vegetation cover. This is applied to 30 randomly selected samples of the MODIS data using an M5 linear model regression tree algorithm, and an average of the 30 model results is computed for each MODIS pixel (Townshend et al., 2011). MODIS VCF collection 5.1, version 1 was used for this study.

The LTC product was downloaded and mosaicked, and used MODIS VCF data from the year 2000 at a spatial resolution of ~250m to gap fill areas obscured by cloud cover (Figure 4.3). This data was at the time the highest resolution freely available, replicable optical mosaic of such a wide area, although it has since been superseded in coverage and scope by the Global Forest Change (GFC) product (Hansen et al., 2013), and more recent LTC products derived from the 2005, 2010 and 2015 GLS epochs have been produced.

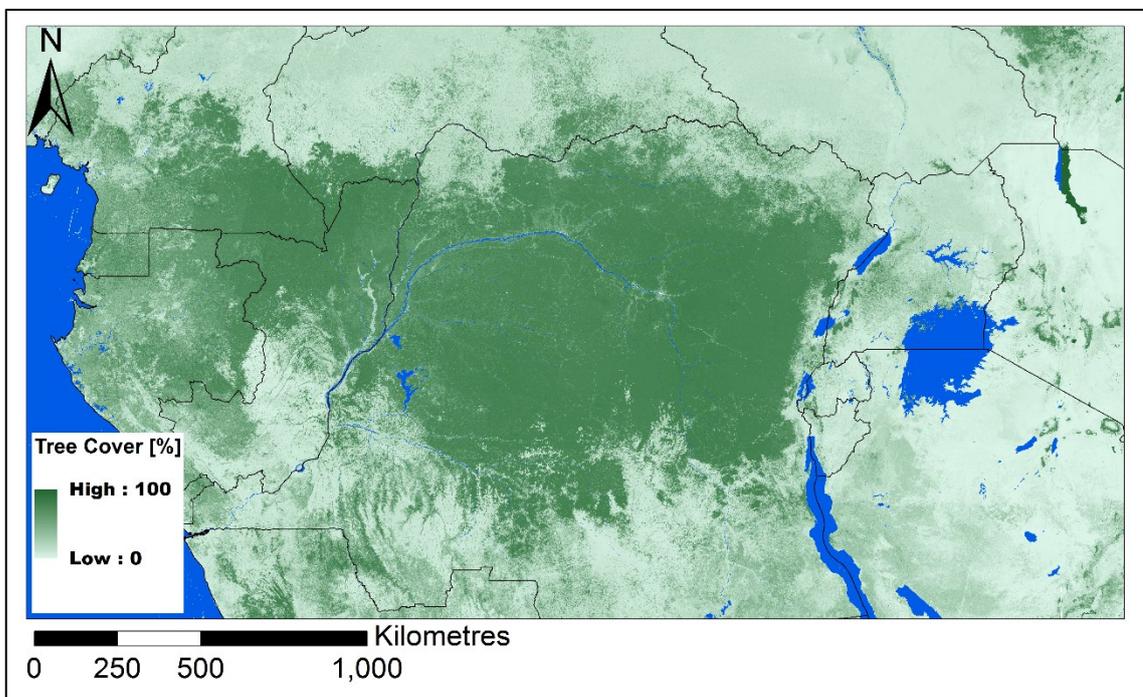


Figure 4.3: Mosaic of Landsat Tree Cover Continuous Fields with MODIS VCF gap fill, covering Central Africa.

4.2.3 JAXA K&C Forest / Non-Forest

A forest/non-forest map was produced by JAXA for the years 2007-2010 from ALOS-PALSAR data, also referred to in section 2.2.1. The technique used is described in detail in Shimada et al. (2014). The dataset was a second version of the K&C Initiative ALOS-PALSAR mosaics at 25 m spatial resolution, produced annually for the years 2007-2010, and processed in 500 km x 500 km tiles for the entire globe at HH and HV polarisations. The algorithm to radiometrically correct and mosaic the data was similar to that used in section 4.2.1, with 16 multilooks to reduce speckle, but with topographic correction applied using local incidence angle to produce backscatter expressed as gamma-naught (γ^0). Frequency distributions of HH and HV backscatter were generated from 5 x 5 median filtered and segmented mosaics, with reference to forest or non-forest objects discerned using Google Earth Imagery. Global and regional frequency distributions were calculated, with temporal differences due to seasonality and a very low power loss observed over the 4 year period. An object-based algorithm was applied to the segments, with the first node masking settlement areas using high HH and HV γ^0 values (> -1 dB and > -6.5 dB respectively). Forest/non-forest (FNF) regional thresholds were then applied to the HV polarisation, and a further water body node divided the Non-Forest (NF) class into Water body and NF using a low HH and HV threshold. For the African region, a segment mean γ^0 FNF threshold of -15.6 dB was selected, derived from a mean γ^0 of -12.76 dB and standard deviation of 2.01 dB in forest class regions of interest, and mean γ^0 of -19.38 dB and standard deviation of 2.70 dB in non-forest class regions of interest. The FNF map in Central Africa is shown below in Figure 4.4; the settlement class in the released product is combined with the non-forest class.

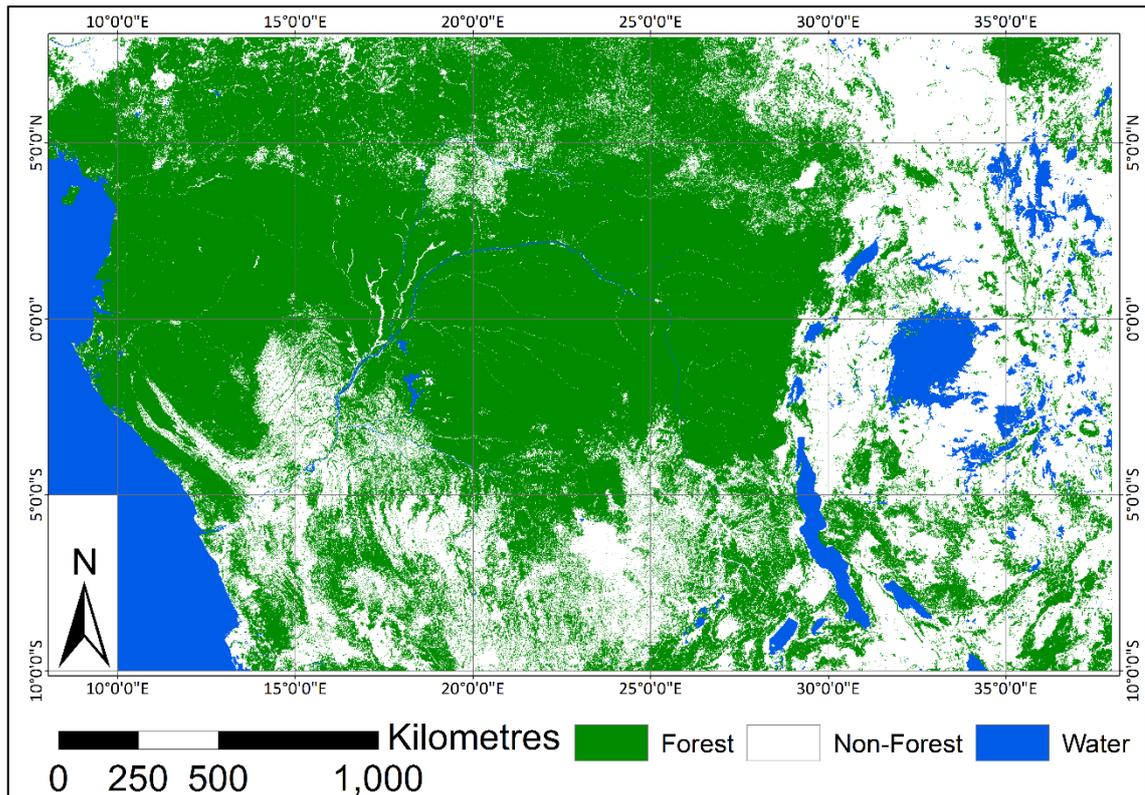


Figure 4.4: K&C Initiative FNF map 2009 for Central African region

4.2.4 ESA-GlobCover 2009

Derived from data from the Medium Resolution Imaging Spectrometer instrument (MERIS) on the ESA ENVISAT platform, GlobCover is a global thematic land cover map at 300 m spatial resolution described in detail in Bontemps et al. (2010). The MERIS data used for GlobCover 2009 were six bimonthly mosaics in four optical bands for the year 2009. 22 classes were defined using the UN Land Cover Classification System (LCCS). Data were geometrically corrected to within 150 m geo location accuracy using a DEM (Getasse 30 DEM, a coarse 30 arcsecond resolution product derived from interpolated altimeter points), and corrected for atmospheric interference and Bidirectional Reflectance Distribution Function (BRDF) effects (reflectance effects due to variations in the sun and view geometry for different acquisitions) to produce surface reflectance. In order to reduce surface reflectance variability and aid classification, the global dataset was split into 22 regions based on environmental and remote sensing discontinuities. The area covered in this study includes all or part of three of these regions. The classification algorithm began with a supervised per pixel classification for poorly

represented classes, followed by an unsupervised classification of remaining pixels to create spectrally similar clusters. These clusters were then analysed for phenological changes and then grouped based on spectral and temporal similarities. Finally clusters were assigned a LCCS class based on reference rules from international experts, seen in the legend in Figure 4.5.

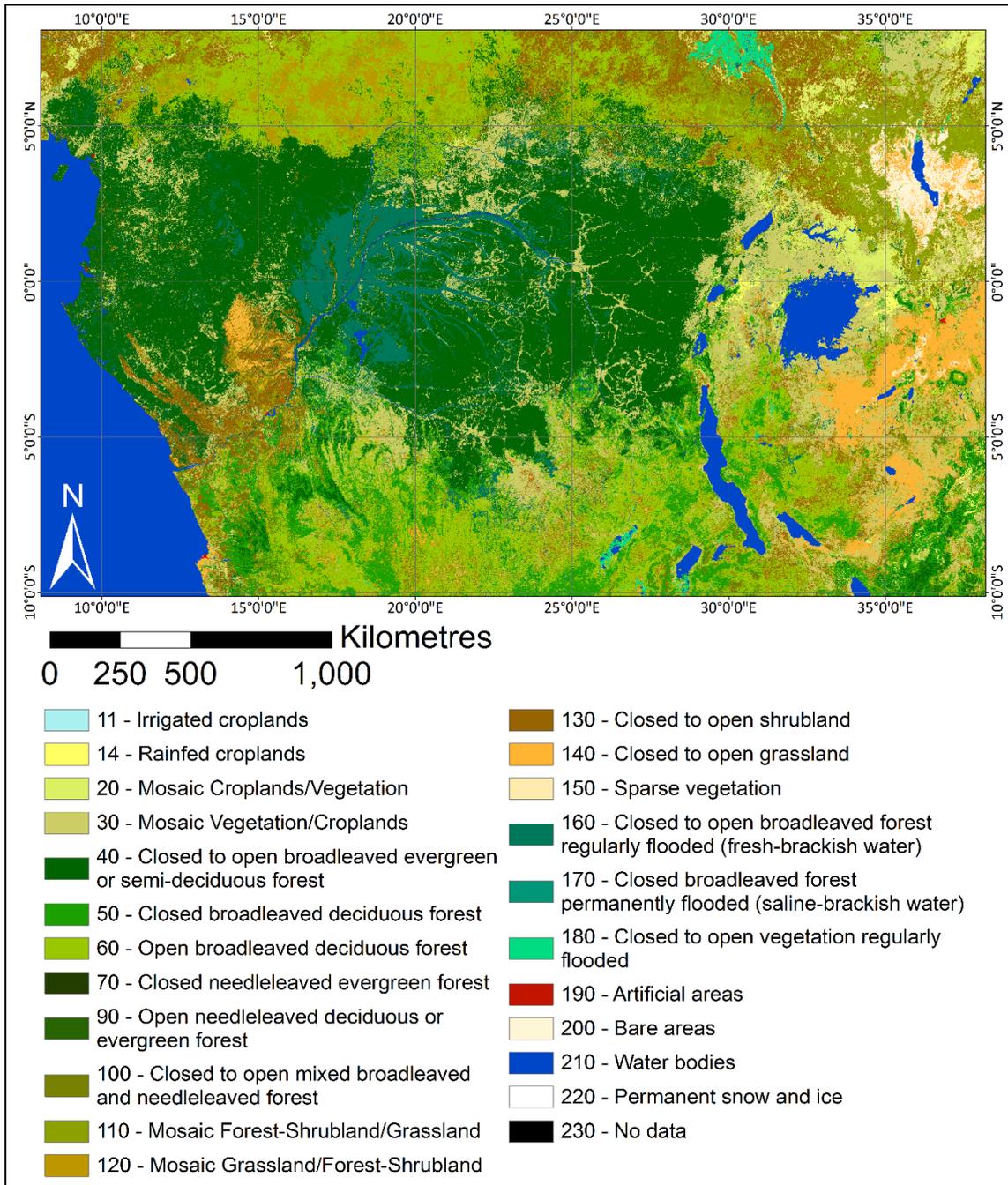


Figure 4.5: GlobCover 2009 with full GlobCover LCCS class legend

For the benefit of analysis with both this study and the JAXA K&C FNF product (both of which are effectively binary FNF classifications), this study aggregated the 22 GlobCover classes to just seven (Figure 4.6), in decreasing order of forest cover (class numbers from the legend in Figure 4.5):

- 1) *closed forest* (classes 50, 70, 170: >40% forest cover)
- 2) *closed to open forest* (classes 40, 100, 160, 180: 15-40% forest cover)
- 3) *open forest* (classes 60, 90: >15% forest cover)
- 4) *mosaic forest* (classes 30, 110: mosaic of forest, other vegetation, and croplands)
- 5) *other vegetated* (classes 11, 14, 20, 120, 130, 140, 150)
- 6) *other non-vegetated* (classes 190, 200, 220)
- 7) *water bodies* (class 210).

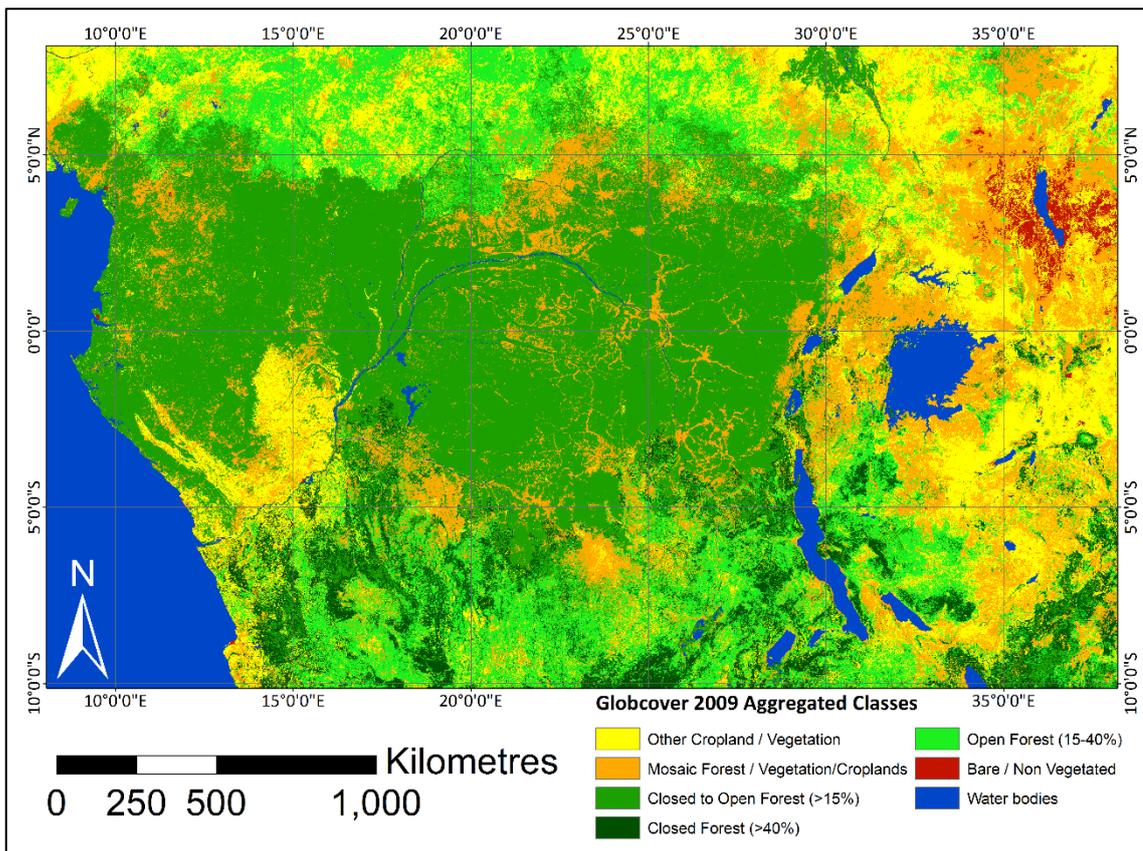


Figure 4.6: GlobCover 2009 Aggregated Classes

4.3 Methods

4.3.1 Data Pre-Processing

In order to coregister the PALSAR mosaics to, in particular, similar or higher resolution datasets such as Landsat, it needs to be precisely georeferenced, to at least one or two pixels. Without this precision, like is not being compared with like, especially in border areas of forest/non-forest. The 2008 and 2009 K&C Initiative PALSAR mosaics exhibited occasional projection problems that resulted in a varying offsets from tile to tile. The problem is somewhat apparent in the Google Earth-projected data provided on the K&C Initiative website (<http://www.eorc.jaxa.jp/ALOS/ge2/KC50/kc50top.kml>), where the data is sometimes shifted by four or five pixels from the Google Earth imagery in the base map. In order to correct for this error, the 2009 mosaic, in the HV polarisation, was subset and reprojected to its corresponding Universal Trans Mercator (UTM) projections (Figure 4.7). A manual correction to one subset's header file tie points (+3.5 pixels in x and +4.5 pixels in the y direction in the UTM zone 33 tile) was sufficient to prepare them for coregistration.

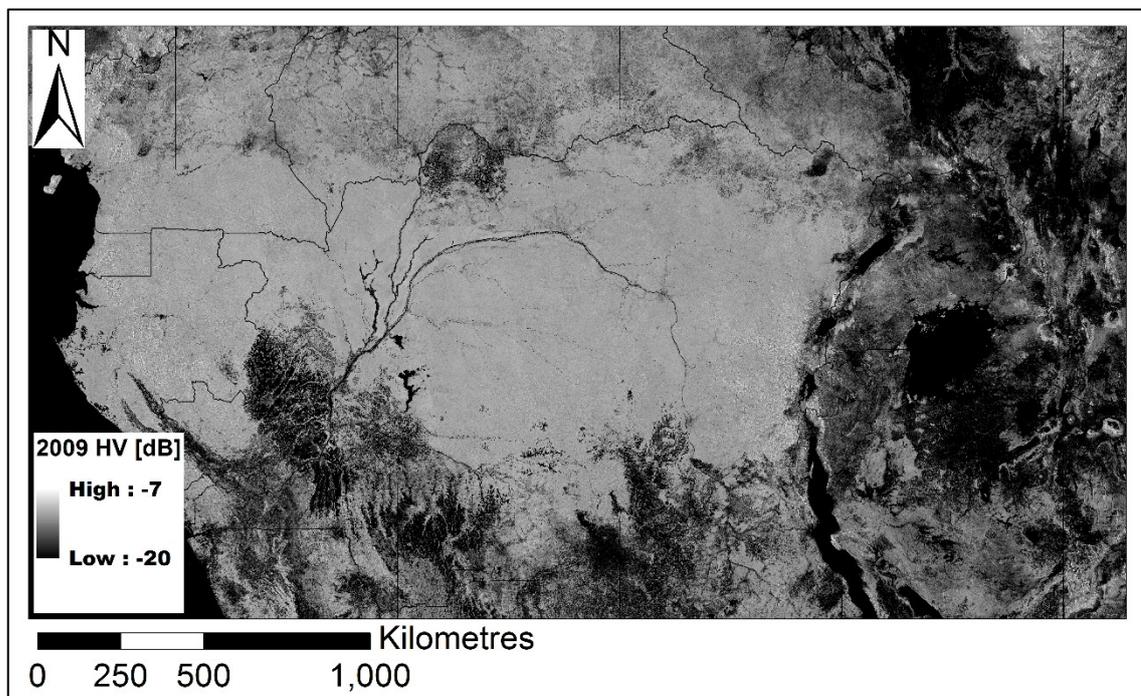


Figure 4.7: ALOS PALSAR HV mosaic (*sigma zero*) from 2009.

For a comparison between SAR-derived and optical-derived forest cover classification in the Congo basin, the 2009 HV polarised SAR data were selected. To begin with, each subset was converted to logarithmic sigma zero (σ^0) values from the backscatter intensity digital number (DN) values using a calibration factor (CF) of -83 as described in the accompanying K&C Initiative metadata file and using the conversion equation below:

$$\sigma^0 [dB] = 10 \log_{10}(\langle DN^2 \rangle) + CF$$

Equation 6

They were then resampled to the same resolution as the LTC data and automatically coregistered and combined in multiband layers according to UTM zone.

4.3.2 Classification

A decision tree classification was then run according to Figure 4.9, using thresholds of 30 % tree cover for the LTC data and σ^0 values of -13.5 dB in the L-band HV data. These threshold values were chosen through comparative analysis of binned pixel value areas in the case of the SAR data, described in full in the following paragraph, and from the upper limit of tree cover as described by the UN Framework Convention on Climate Change (Decision 16/CMP.1, UNFCCC 2006) for the LTC data.

A histogram of σ^0 values in 0.1 dB bins between -83 dB and 7.5 dB (the data minimum and maximum values) was generated for the entire study area. The cumulative area sum of these bins (calculated by multiplying the pixel count by the pixel area, 30 m x 30 m) from high dB values to low was calculated, to give the likely area of forest cover for a potential forest cover threshold. These values were compared to the aggregated GlobCover 2009 forest cover classes, also cumulatively summed from more forested (Class 1: Closed Forest) to less forested (Class 4: Mosaic Forest) to indicate the likely classes that may be captured by a particular σ^0 threshold. In the absence of reliable training data for this dataset, this method allows selection of a threshold that closely captures a similar area of forest, although the class distributions could of course differ. From the areas for the aggregated GlobCover classes 1 to 4, the σ^0 values with corresponding cumulative areas were retrieved as -10.8, -12.6, -14.5 and -17.9 dB respectively, illustrated in Figure 4.8. The Closed Forest class area is above the upper

limit of tree cover as described by the UN Framework Convention on Climate Change (Decision 16/CMP.1, UNFCCC 2006). The Mosaic Forest class is too broad a category and neither the Open Forest nor the Closed to Open Forest classes precisely describe this threshold, so the mean between classes 2 and 3 was selected, -13.5 dB.

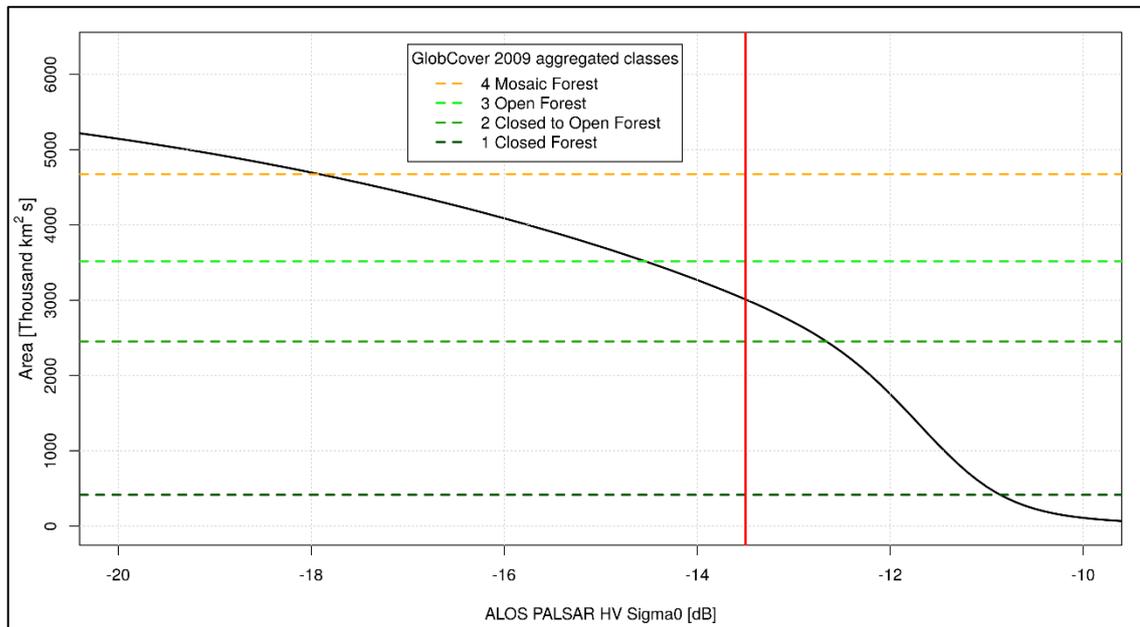


Figure 4.8: Comparison of cumulative area of HV σ^0 values with aggregated GlobCover forest cover classes. Red vertical line indicates FNF threshold value for HV σ^0 (-13.5 dB)

The results of the decision tree described in Figure 4.9 allow visual analysis of the potential of SAR with respect to an optical classification, by looking at disputed regions (coloured orange and light blue) and identifying which method has correctly classified the area as forest or non-forest.

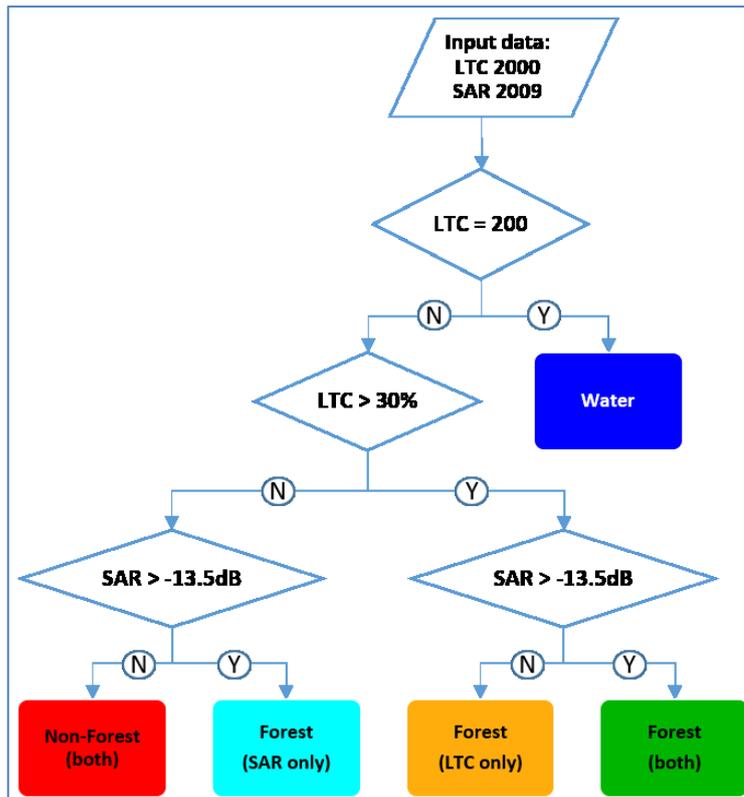


Figure 4.9: Decision tree classification schematic of SAR (2009 PALSAR HV sigma zero values) and optical (LTC) data; coloured boxes indicate classes and correspond to classes used in Figure 4.12 and Figure 4.13

4.3.3 Accuracy Assessment

For the purpose of accuracy assessment, a reference dataset was produced, manually, through visual analysis of high resolution Google Earth Imagery (GEI), as close as possible to the year of ALOS-PALSAR image acquisition (2009). GEI includes both high resolution optical imagery, as well as medium resolution annual true colour composites of Landsat data. It was also possible to check near infrared (NIR) bands using Google Earth Engine (GEE), an online resource that allows both visualisation and computation of large volume remote sensing datasets, including the Landsat archive. A stratified sample of 750 points was selected from a larger random sample of 5000 points from the study area in Figure 4.10. The stratification allowed selection of 150 points from each of the five classes from Figure 4.9, which resulted in 300 points each for Forest and Non-Forest from both LTC and SAR thresholds (150 shared points in each class). Bounding circles of radius 50 m centred on each point were generated and a four class reference dataset was built by exporting the dataset to Google Earth and manually assigned either *forest*, *non-forest*, *water* or *mixed* values (based on a visual assessment where no single

class occupied more than half of the bounding circle). Examples of bounding circles in GEI and their assigned classes are shown in Figure 4.11, with indication of which classifications correctly identify the class.

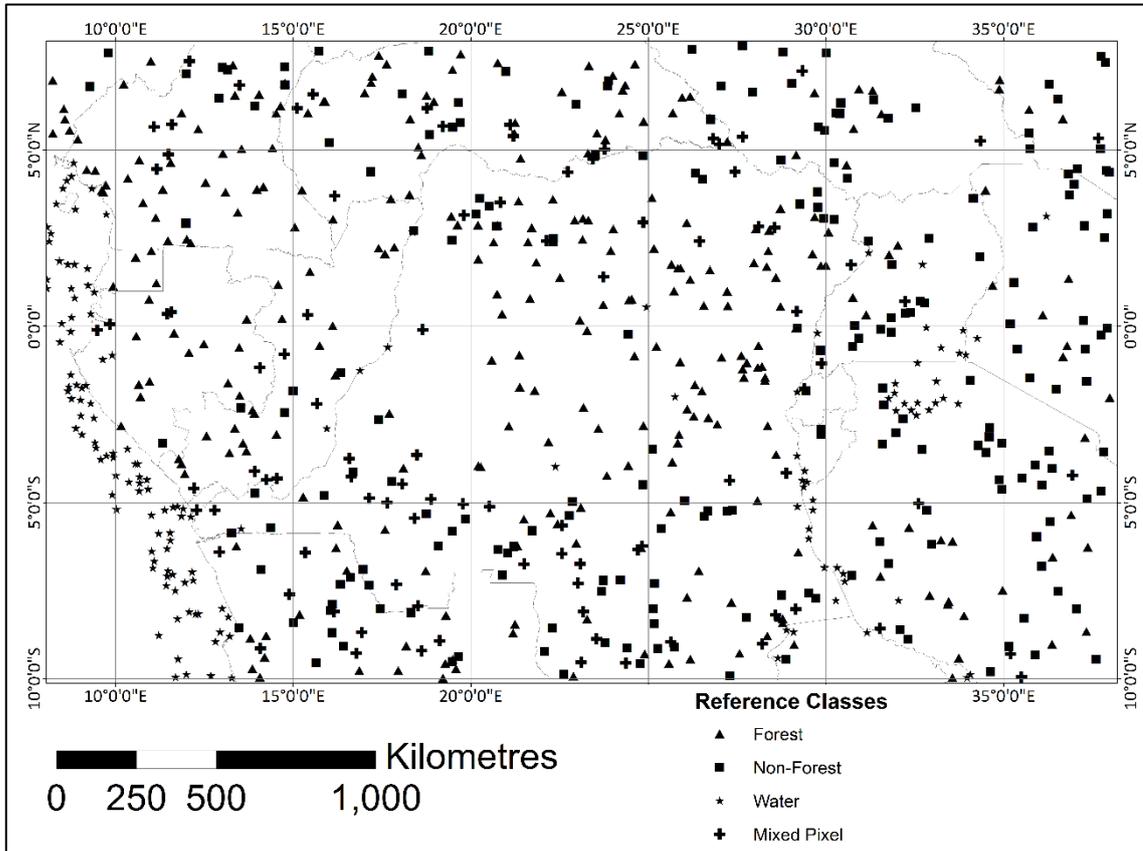


Figure 4.10: Distribution of validation reference points, with reference class legend. symbol sizes not to scale

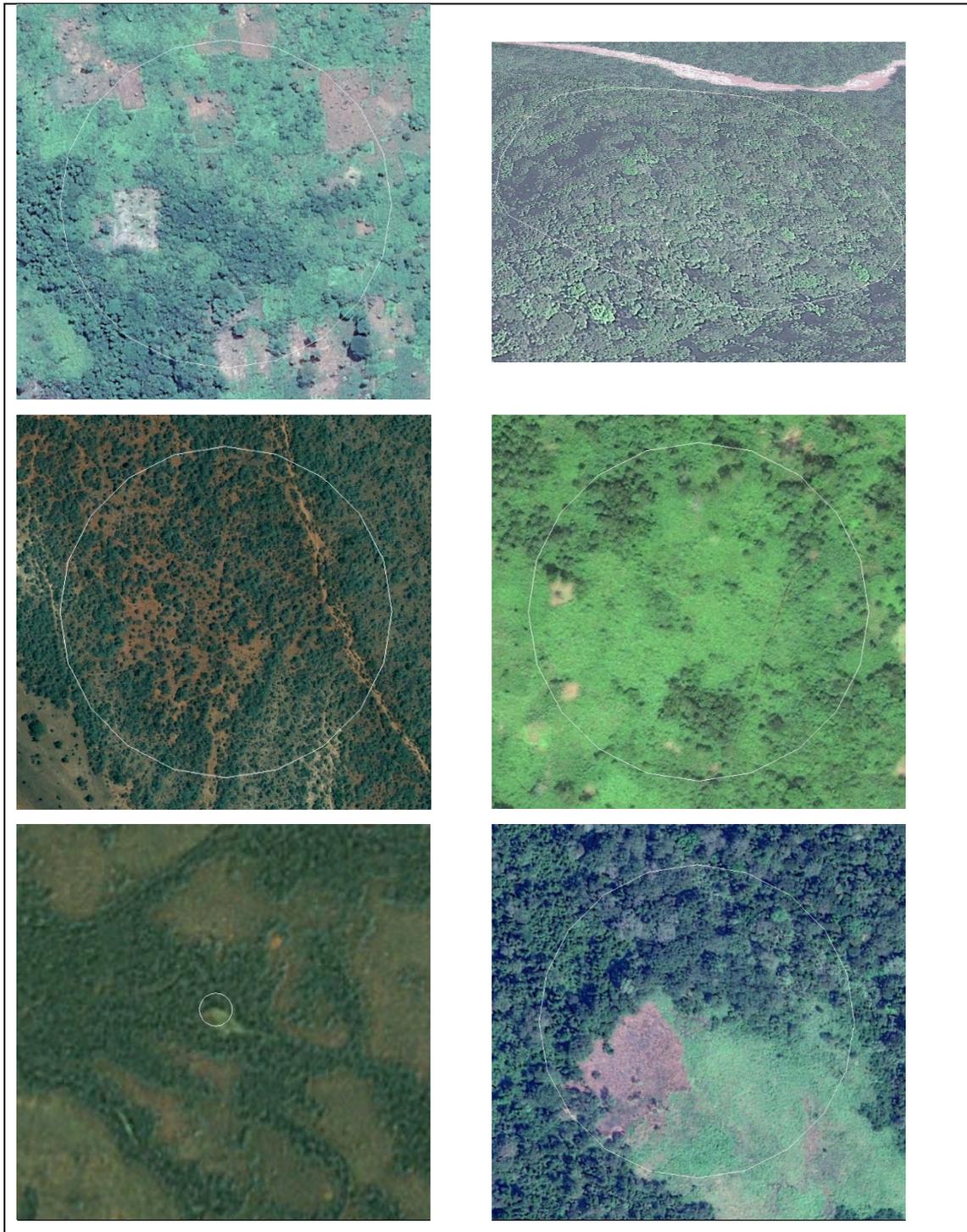


Figure 4.11: Screen grabs from Google Earth of reference dataset generation. **Top left:** mixed class, classified as forest in JAXA FNF, LTC threshold and SAR threshold; **top right:** oblique view of forest on steep topography, classified as forest in JAXA FNF and LTC threshold, but not by SAR threshold; **middle left:** non forest (shrub mosaic) correctly classified as non-forest by JAXA FNF, LTC threshold and SAR threshold; **middle right:** cropland mosaic, classified as forest in JAXA FNF and LTC threshold, but not by SAR threshold; **bottom left and right:** zoom out (Landsat resolution) and zoom in (high resolution GEI) of a mixed class, classified as forest by JAXA FNF, LTC threshold and SAR threshold

4.3.4 Software Used

Initial processing for used a combination of Matlab (version 2013a) and R (version 2.15) to process and recombine a considerable number of zipped SAR files contained on an ftp server. Scripts involved simple directory and file manipulation, and mosaicking using the R raster package (version 2.3-12) and as such did not require reproduction in this thesis. Decision tree thresholds were applied using EXELIS-ENVI (version 5.0). Reference data shapefiles were generated and manipulated within ESRI ArcGIS (version 10.2.2), which was also used to produce the final map figures.

4.4 Results

4.4.1 Visual Assessment

Figure 4.12 shows the resulting 2009 PALSAR HV mosaic-derived forest/non-forest classification comparison with the LTC 2000 epoch. The eight windows in Figure 4.13 illustrate some of the limitations and strengths of both datasets across the study area, with accompanying high resolution optical imagery. There are major discrepancies between the two thresholds used; in the north throughout most of the Central African Republic (CAR), in cropland areas, and in upland areas where there is a terrain effect on the SAR data. In the study area 837,196 km² is above the -13.5 dB HV SAR threshold but below the LTC forest threshold (i.e. < 30% tree cover), and 259,089 km² is classified as non-forest by the SAR threshold but above the LTC forest threshold. The total area of forest reported by JAXA's FNF product is 3,542,858 km²; the total area above the SAR FNF threshold is lower at 3,009,964 km² and the area above the LTC FNF threshold is much lower at 2,431,857 km². In the aggregated GlobCover 2009 product, the areas of each of the four forested classes (classes 1-4 in Figure 4.6), from most densely forested to less densely forested, is 416,846 km², 2,010,730 km², 1,114,535 km², and 1,132,889 km² respectively.

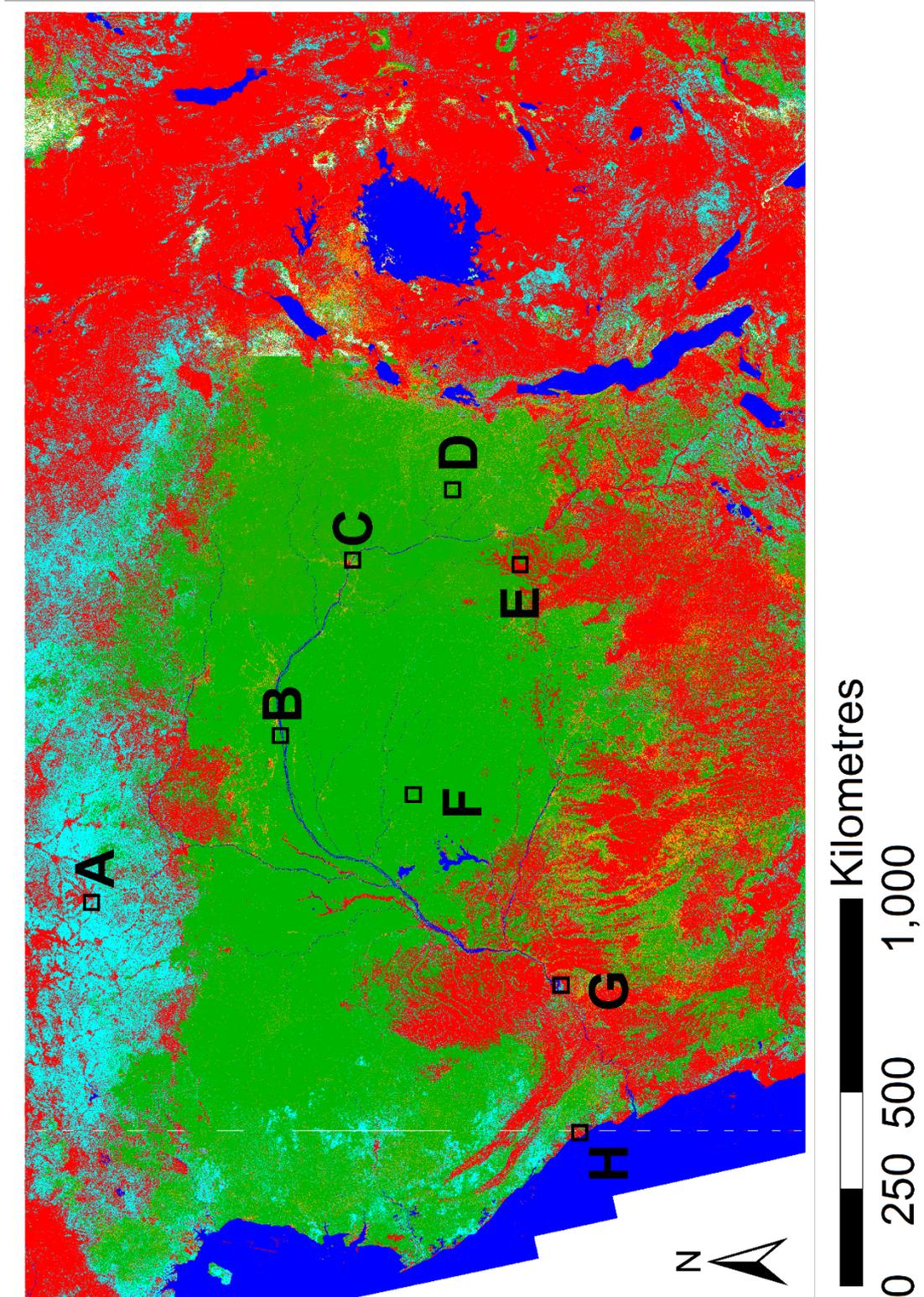


Figure 4.12: Full Forest/Non-Forest optical/SAR classification comparison with boxes showing location of inserts of Figure 4.13, which also contains legend

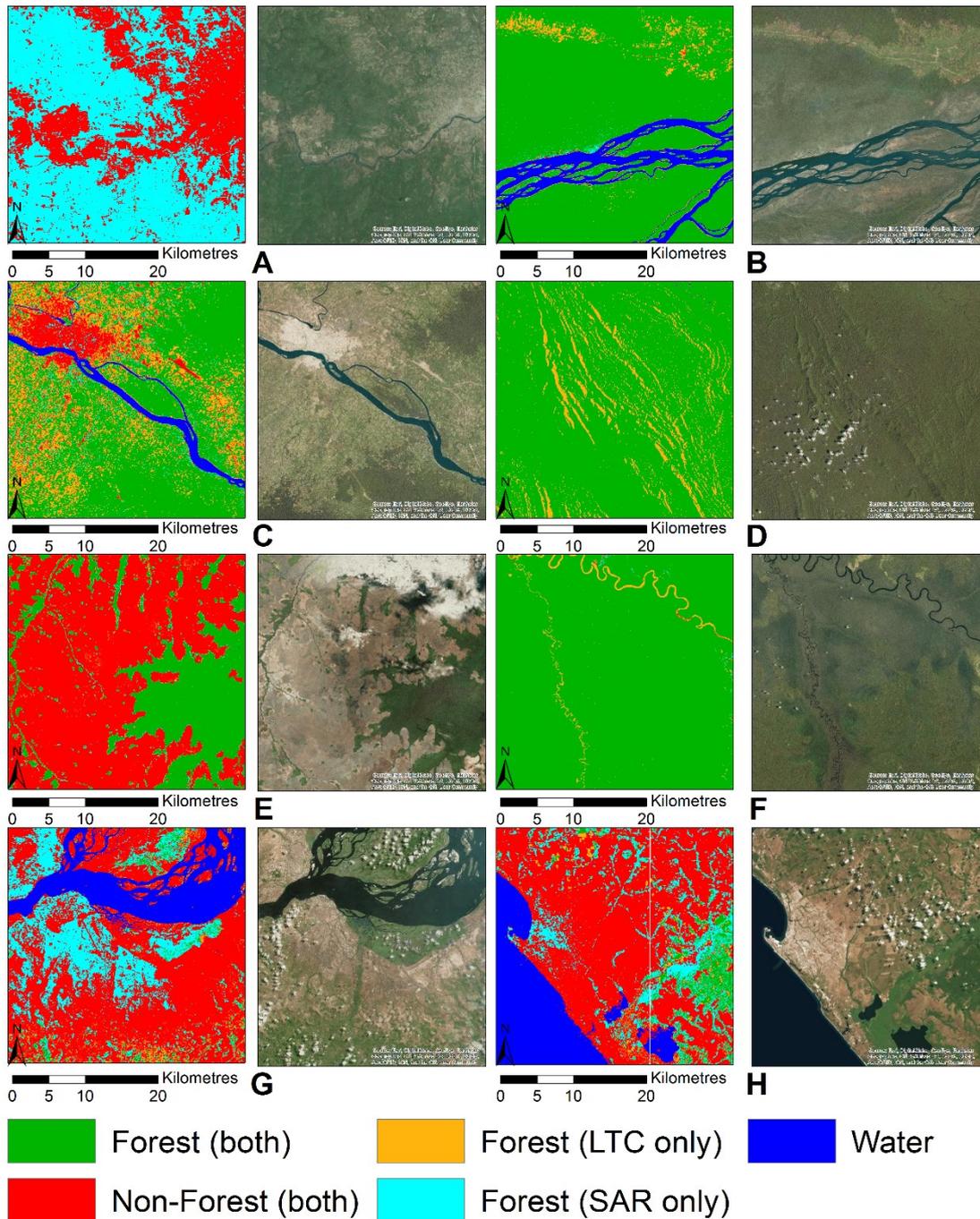


Figure 4.13: Forest/non-forest optical/SAR classification comparison, with eight selected areas of interest (A-G) and accompanying high resolution imagery detail for further comparison: **A)** forest under-represented by LTC, correctly classified by SAR; **B)** crops misclassified as forest by LTC; **C)** more accurate classification of forest, water and urban (non-forest) by both SAR and LTC; **D)** terrain effects in SAR data causing misclassification of slopes facing away from the sensor as non-forest; **E)** deforestation in the south eastern Congo Basin; **F)** misclassification of water by LTC; **G)** misclassification of urban (non-forest) area of Kinshasa as forest by SAR; **H)** missing line from coregistration process. High resolution imagery used under academic license from: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community

4.4.2 Error Matrices

Using the 750-point reference dataset, an error matrix was generated for the SAR FNF threshold, LTC FNF threshold, and JAXA FNF product (Table 4.1). Although the mixed pixel class from the reference dataset indicates classification difficulties and was a necessary reference class, it has no equivalent class in any of the FNF maps, so a second error matrix (Table 4.2) was produced with the mixed pixel reference class omitted (and with corresponding corrections to overall, producer's and user's accuracies). Each product differed in the distribution of classes assigned to the mixed pixel reference class, and it is not surprising to note that the combined LTC and SAR threshold classes for FNF (where both thresholds were met in each case) has both high accuracies and significantly fewer (4.7% vs 12.5% of the total number of points, respectively) *Mixed* reference class occurrences (Table 4.3). The JAXA FNF product has a higher overall accuracy (76.27%) with the SAR threshold slightly below (74.13%) and LTC threshold lowest (66.93%). The removal of the *Mixed* reference class from the error matrices increases the overall, producer's and user's accuracies. The *Water* class accuracies are extremely high, with the exception of the JAXA product, which buoys the overall accuracies. Removing the *Mixed* reference class and the *Water* classes results in overall accuracies of 80.59%, 69.90% and 85.54% for the SAR threshold, LTC threshold, and JAXA FNF products respectively. The errors of commission in the Forest class are higher in the JAXA FNF product than the SAR threshold method, and the error of omission in the Non-Forest class is higher in the JAXA FNF product than the SAR threshold method. This is an expected result as a lower threshold for tree cover (15%) is used by the JAXA FNF product's forest definition, meaning a greater forest area was reported.

Table 4.1: Error Matrices for SAR FNF, LTC FNF, and JAXA K&C FNF products, including Mixed reference class

SAR threshold	Reference Data					Overall Accuracy	Producer's Accuracy	User's Accuracy
	F	NF	Water	Mixed	Total			
Forest	228	34	0	38	300		0.78	0.76
Non-Forest	64	179	2	55	300		0.84	0.60
Water	0	0	149	1	150		0.99	0.99
Total	292	213	151	94	750	74.13		
LTC threshold								
Forest	191	51	1	57	300		0.65	0.64
Non-Forest	101	162	1	36	300		0.76	0.54
Water	0	0	149	1	150		0.99	0.99
Total	292	213	151	94	750	66.93		
JAXA FNF								
Forest	276	56	1	66	399		0.95	0.69
Non-Forest	16	150	4	28	198		0.70	0.76
Water	0	7	146	0	153		0.97	0.95
Total	292	213	151	94	750	76.27		

Table 4.2: Error Matrices for SAR FNF, LTC FNF, and JAXA K&C FNF products, excluding Mixed reference class

SAR threshold	Reference Data				Total	Overall Accuracy	Producer's Accuracy	User's Accuracy
	F	NF	Water					
Forest	228	34	0		262		0.78	0.87
Non-Forest	64	179	2		245		0.84	0.73
Water	0	0	149		149		0.99	1.00
Total	292	213	151		656	84.76		
LTC threshold								
Forest	191	51	1		243		0.65	0.79
Non-Forest	101	162	1		264		0.76	0.61
Water	0	0	149		149		0.99	1.00
Total	292	213	151		656	76.52		
JAXA FNF								
Forest	276	56	1		333		0.95	0.83
Non-Forest	16	150	4		170		0.70	0.88
Water	0	7	146		153		0.97	0.95
Total	292	213	151		656	87.20		

Table 4.3: Error Matrix of FNF from combined LTC and SAR thresholds (where both thresholds were met in each class)

Combined	Reference Data					Overall	Producer's	User's
	F	NF	Water	Mixed	Total	Accuracy	Accuracy	Accuracy
Forest	136	3	0	11	150		0.94	0.91
Non-Forest	9	131	1	9	150		0.98	0.87
Water	0	0	149	1	150		0.99	0.99
Total	145	134	150	21	450	92.44		

4.4.3 GlobCover class comparisons

To understand the difference in classification between the three methods beyond broader accuracy statistics, the distributions of aggregated GlobCover classes (Figure 4.14) across the study area in each method's Forest and Non-Forest classes were calculated. Total area distributions (Figure 4.15) and the areas expressed as a percentage of the total class areas (Figure 4.16) were calculated. A similar area distribution for all GlovCover 2009 classes is in Appendix A. As the classes are arranged in order of decreasing forest cover, it would be expected to see the Forest class have a higher distribution in classes 1-3 and lower in classes 4-7, and the opposite for the Non-Forest class distribution. This is generally the case for the JAXA FNF product, but not for the SAR and LTC thresholds in the Closed Forest class, and not for the LTC threshold in the Open Forest class.

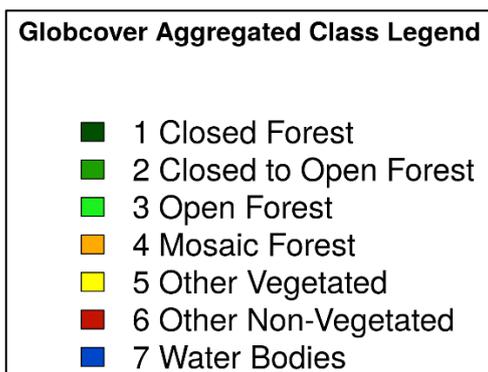


Figure 4.14: Globcover Aggregated Class Legend

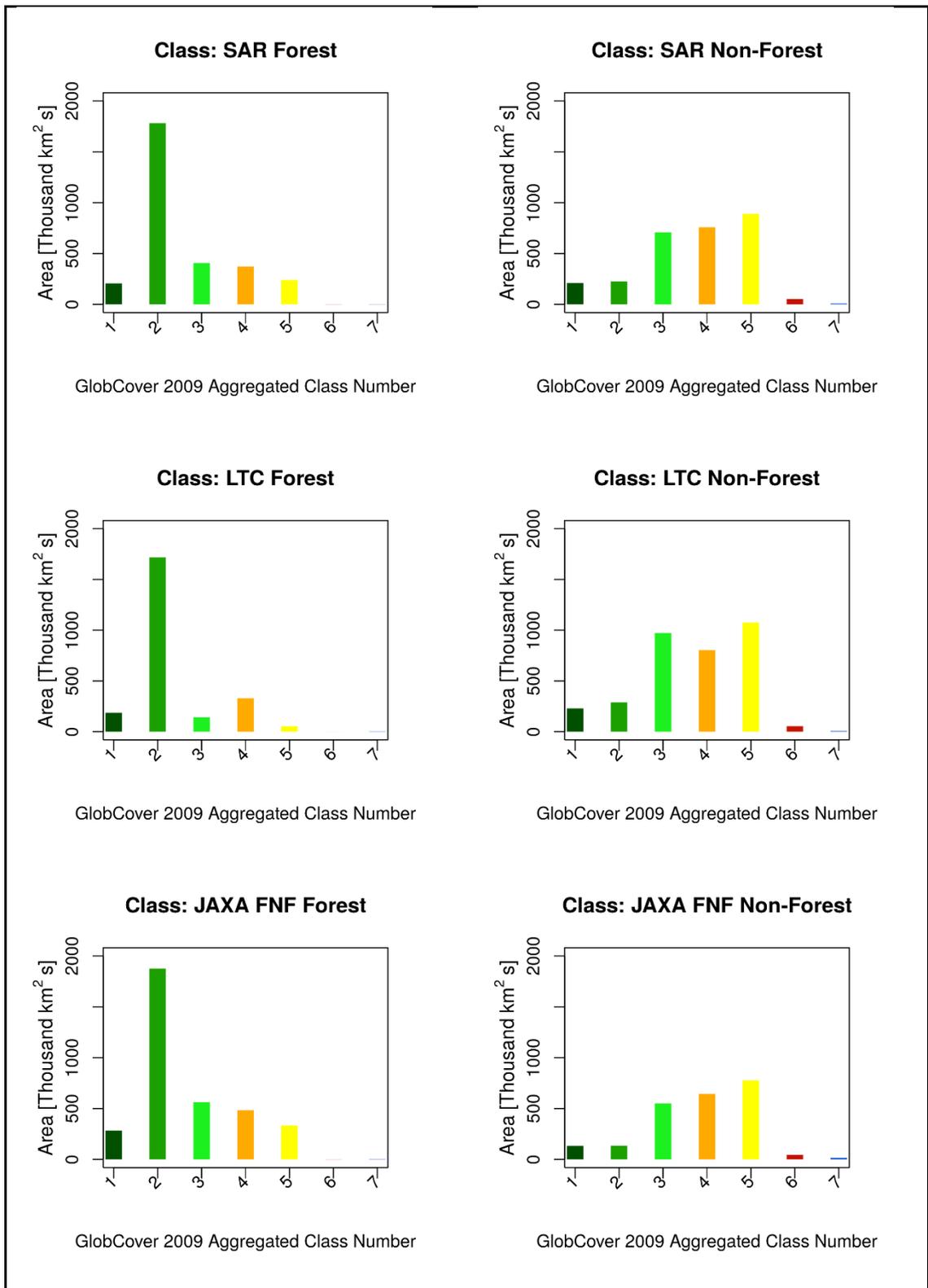


Figure 4.15: Comparison of FNF classes in terms of area distribution of aggregated GlobCover classes (class legend in Figure 4.14)

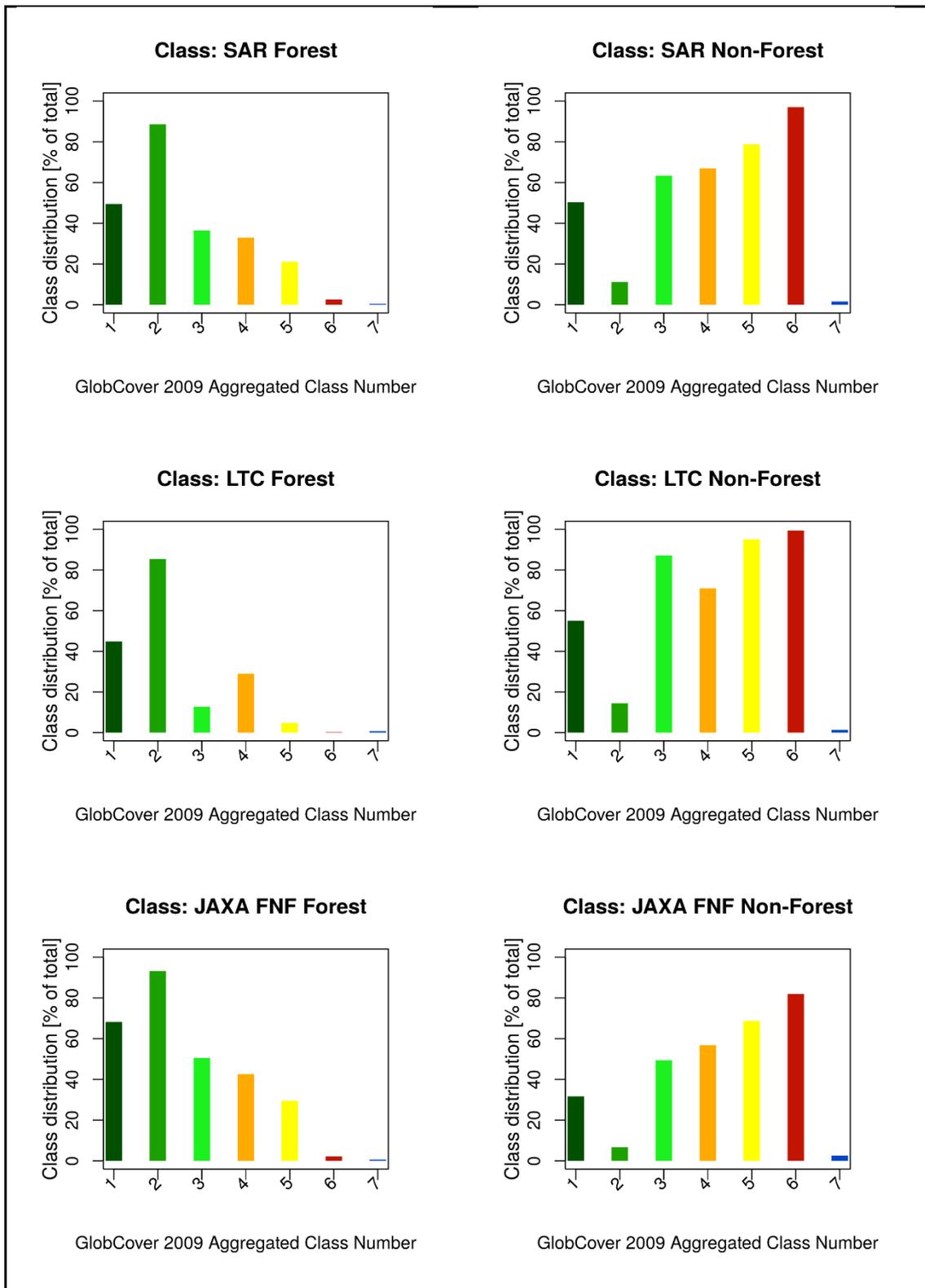


Figure 4.16: Comparison of FNF classes in terms of percent of the total area distribution of aggregated GlobCover classes (class legend the same as in Figure 4.14)

Combining the class distribution and the reference data for each class illustrates in which general areas each product has strengths and weaknesses. The graphs in Figure 4.17 show the results of this analysis in the areas where SAR and LTC thresholds result in opposing classes (Forest/Non-Forest in particular), as well as the sometimes low number of samples depending on the class combinations. The water class is omitted since there were too few cases of errors of commission or omission in this class, and also because both SAR and LTC threshold methods used the same water mask. The spatial distribution of these class combinations are seen in Figure 4.12. In these graphs, the total reference point count is given; a high total count indicates the level of class accuracy. It is also clear from these graphs that the GlobCover 2009 product does not always correspond precisely with the reference data. The GlobCover products are coarser resolution and for a comparable accuracy assessment would require a different validation method.

The top row of Figure 4.17 shows that the SAR threshold performs better than the LTC threshold in classifying Forest and Non-Forest, having higher total counts where the class agrees with the reference data. The distribution shows that LTC incorrectly classifies more of the *Open Forest* GlobCover classes as Non-Forest. The SAR threshold correctly classifies more Non-Forest in the *Closed Forest* GlobCover classes, and there were a greater number of *Mixed* Reference points in the SAR Non-Forest / LTC Forest class than in the SAR Forest / LTC Non-Forest class.

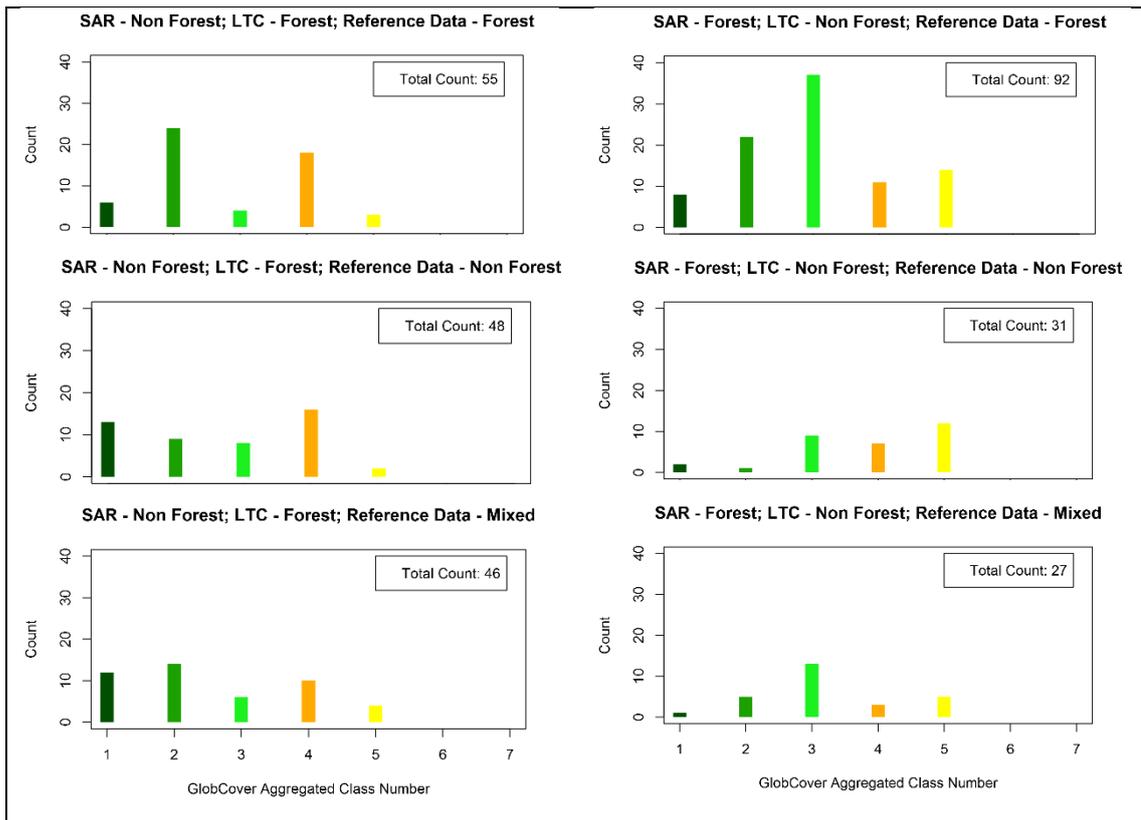


Figure 4.17: Reference data distribution of GlobCover 2009 aggregated classes for differing LTC and SAR threshold results; class legend in Figure 4.14

The graph in Figure 4.18 shows the cumulative area (from 100% to 0%) of the LTC product in the study area, giving an indication of the threshold changes required to report a similar area to the GlobCover 2009 classes, as well as the reported areas of the other two FNF products analysed in this chapter. This explains to some extent (as the distribution of the area after a threshold change would not exactly match the GlobCover 2009 class distributions) why the *Open Forest* aggregated class is not represented as forest as much as the SAR threshold by the LTC 30% threshold. It also shows clearly that the stated lower threshold of tree cover (15%) of the JAXA FNF forest definition matches both the LTC product's area at that threshold, as well as the GlobCover area of all forest cover aggregated classes above Open Forest. The higher threshold of this study's SAR product corresponds to the cumulative LTC area above 20% rather than the proposed 30% threshold aimed at with the forest definition in section 3.1.

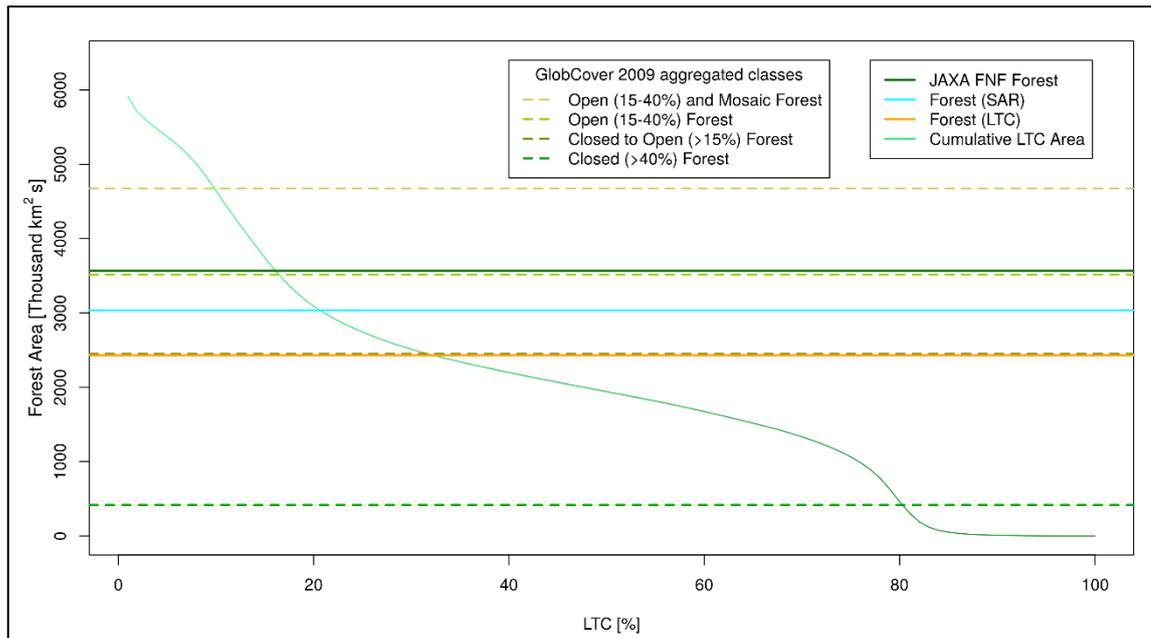


Figure 4.18: Cumulative LTC area comparison with areas of aggregated GlobCover 2009 classes, and the reported areas of the FNF products

The time difference between the SAR and optical datasets is a contributing factor towards class differences in certain areas, where deforestation or reforestation has occurred in the intervening period, and may explain the low overall accuracy from the validation procedure.

4.5 Discussion

The results from the accuracy assessments as well as the GlobCover aggregated class comparison appear favourable to the JAXA FNF product. However, the JAXA FNF product was generated from a superior terrain-corrected dataset released concurrently (De Grandi et al., 2011; Shimada et al., 2014). The relatively straightforward data analysis and classification of this study comes close to matching its overall accuracy, and indeed exposes an overfit in the JAXA FNF Forest class, evidenced by the lower user's accuracy and the higher percentage of mosaic and 'other' vegetation (aggregated Globcover 2009 classes) than this study's SAR classifier. Although it performs relatively well, it is important to users that a global forest/non-forest product be able to deliver accurate and consistent results in the second largest tropical rainforest regardless of difficulties in obtaining reliable ground truth data for calibration and validation. This study shows

that improvements can be made at low computational cost, and provides a reliable and unbiased validation dataset that may be used to supplement future forest mapping assessments in this region.

The validation dataset produced for this chapter adds to existing literature (e.g. Carreiras et al., 2014; Olofsson et al., 2013) demonstrating that with careful use of freely available high resolution data, and with reference to ancillary datasets providing information about elevation, and cues from now widely available medium resolution multispectral, well calibrated datasets, alternatives exist to costly, dangerous, and potentially inefficient ground truth gathering expeditions. There have been significant recent discoveries in the field in the Congo Basin, aided by and aiding remote sensing observations, in particular uncovering a potentially vast below ground biomass carbon sink in the form of an extensive peat dome in the Republic of Congo by Dargie et al., (2017), but these are rare and require concerted and persistent team efforts.

The problems faced by the optical datasets encountered at this stage illuminate the necessity for timely acquisition of medium to high resolution remotely sensed imagery. The LTC 2000 product was generated from 4 years of acquisitions, and was the only freely available 30 m resolution global tree cover product at the time of its release in 2012. The multiple years and mixed seasons of acquisition of the final product (which still required gap filling by coarser resolution MODIS VCF data), and the time lag between it and the SAR products it was compared with in this study, gave it an unfair disadvantage. Globcover 2009, which coincided with the SAR datasets of this study, is based on high temporal resolution but low spatial resolution acquisitions, and as a result it was again unsuitable for direct comparison.

The continuation of annual L-band datasets being released by the K&C Initiative, as well as the launch of the Sentinel 1 constellation, now provides a wealth of freely available SAR data, mostly unhampered by weather conditions, enabling a reduction in uncertainty in landcover classification due to the effects of seasonal variability. Before the launch and easy availability of the ESA Copernicus Sentinel 1 C-band SAR and ALOS2-PALSAR2 L-band sensors, the release of the JAXA K&C Initiative SAR mosaic in its second version gave an opportunity to focus on addressing the concerns encountered in this

chapter, namely the issues surrounding terrain correction, misclassifications in wetland areas, and also the chance to explore a longer time series of SAR data with improved metadata.

In a similar study, also using a decision tree thresholding approach on a similar but updated K&C Initiative mosaic dataset, Qin et al. (2016) achieved a high reported overall accuracy (95.9%).

In optical remote sensing, new opportunities were also arising. The Landsat Data Continuity Mission became Landsat 8 OLI (Operational Land Imager) on its launch in February 2013 (Rocchio & Ochs, 2007; NASA, 2017) ,and the ESA Copernicus Sentinel programme also produced the multispectral Sentinel-2A optical sensor in June 2015 with a 10-day repeat visit interval reducing to 5 days with the Sentinel-2B launch in March 2017 (ESA, 2017). However, the results of this chapter showed that the at-the-time the repeat visit interval of available medium and high resolution optical data was still too infrequent to overcome the challenges faced in the Congo Basin.

4.6 Conclusions

In the context of the aims and objectives outlined in Section 4.1, there are a number of conclusions that can be gained from this chapter. The overall aim was relatively modest, *'to explore available spaceborne L-band SAR data and through analysis suggest improvements in the acquisition and processing of the data'*, and was met in terms of the available SAR data that the aims were framed under. As is understood in academia, available data effectively means financially available, and the K&C Initiative mosaic FBD (Fine Beam Dual Polarisation) products are just that, and remain the most up-to-date freely available spaceborne L-band SAR products.

The three objectives were met: specifically by comparison of a SAR derived forest/non forest product with an existing land cover product (Globcover 2009); classification, validation and comparison with an existing forest cover product (JAXA FNF product); through analysis of the results and the proposal of ways to improve subsequent releases

of K&C Initiative Mosaic data. These are provision of ancillary datasets, in particular including Local Incidence Angle to allow processing and testing of terrain effects, and date of image acquisition – both of these metadata were unavailable, yet are crucial for an operational mosaicked SAR product; pre-processing of mosaics to account for terrain effects, for the same reason as releasing the metadata mentioned above – having a γ^0 product removes the potential for unwanted terrain effects on land cover classification, which is the principal requirement for K&C Initiative releases.

A novel result from this chapter came from the in-depth analysis of the HV σ^0 normalised radar cross section and how it related to the aggregated Globcover 2009 classes, such that the range of thresholds in that polarisation and their relationship to both forest area and landcover were presented (Figure 4.8 and Figure 4.16). This could assist future efforts to map forest cover and other land cover types in the Congo basin, particularly if more recent K&C Initiative products are used, with the caveat that the two relationships (HV σ^0 to forest cover and HV σ^0 to Globcover aggregated classes) are both taken into consideration, along with the issues associated with this dataset. There have been previous data fusions and comparisons between K&C Initiative ALOS PALSAR mosaics and Globcover 2009, such as Dong et al., (2014), but to the author's knowledge there are none which undertake such a comparison in Central Africa with the scope of this study.

The strengths and in some cases limitations of SAR with respect to optical data, as well as their areas of agreement, were illustrated by the eight areas of interest in Figure 4.13. These were calculated using relative differences in forest/non-forest classification from both datasets, and gave an indication of how a synergy of optical and SAR imagery might best be implemented. Principally, SAR and optical methods may complement each other in terms of correcting misclassification of urban and agricultural cropland areas as forest, respectively. Alternatively, a SAR-only classification could benefit from either a more thorough analysis of urban areas, or more simply with the use of an urban mask from ancillary datasets or the HH polarisation channel. The SAR terrain-related errors in classification were a cause for concern given the lack of sufficient metadata to resolve and correct for these issues, but the inclusion of local incidence angle information, as

well as processed terrain correction, in the January 2014 release of the K&C Initiative ALOS-PALSAR data resolved this issue.

At this point the limitations of working with such a large study area were noted. Due to the short time series in this study, processing such a large area was possible but time-consuming, but with the implementation of a supervised machine learning classification approach and an increase in the number of input data (both in terms of SAR polarisations and available time-series), for practicality a smaller area would need to be selected.

Chapter 5: Creating Annual Forest/Non-Forest Maps in a region of Central Africa from 2007-10 using SAR Mosaic data

Part of the work presented in this chapter has been published in a modified form in Wheeler et al. (2017).

5.1 Introduction

This chapter describes a study to generate annual forest cover and forest change (loss and gain) maps for a large section of the Congo Basin for each year from 2007 to 2010 from a supervised classification of the K&C Initiative L-band SAR HH and HV mosaics, and compare the results with the JAXA K&C Initiative's own forest/non-forest maps produced from the same dataset. The accuracy assessment of the classification using version 1 of the K&C Initiative mosaics from the work in Chapter 4 demonstrated that the data and classification method could be improved, and for practical reasons in terms of data volume and processing time, the study area needed to be reduced. This chapter is a progression, using an updated version of the K&C Initiative mosaics for 2008 and 2009, with two additional annual mosaics for 2007 and 2010. Optical remote sensing products were not used in this study, for the reasons outlined in Section 4.5 (principally the trade-off between obtaining frequent cloud free imagery and spatial resolution at the time of commencing the study), allowing the focus to rest on L-band SAR products.

5.1.1 Aims and Objectives

The aim of this chapter is to continue to answer the first research question from section 1.4: "What are the ideal image acquisition parameters and classification techniques for an annual forest map in the Congo Basin from spaceborne SAR data, given currently available data and regional seasonal effects on image quality?" With this in mind, the following objectives are outlined:

- To obtain and characterise available L-band SAR data in relation to other available map products in the Congo Basin

- To process and classify the data and compare with an existing forest cover product
- To analyse the results and suggest improvements in data acquisition, release and processing

5.2 Datasets

5.2.1 ALOS PALSAR

Dual polarised (HH and HV) K&C Initiative ALOS-PALSAR data were used, at a spatial resolution of 50 m. The data were pre-processed and released in a mosaic format, and geocoded consistently and accurately, allowing for easier comparison with other forest cover maps. The HH and HV channels were provided as normalised radar cross-section, gamma naught (γ^0), meaning the backscatter intensity has been adjusted for SAR geometric distortions (orthorectified) and topography (slope corrected) using the sigma naught (σ^0) value (generated from backscatter intensity using Equation 1 and Equation 6) and cosine of the local incidence angle (θ) as described in (Shimada et al., 2014) and in section 4.2.3.

$$\gamma^0 = \frac{\sigma^0}{\cos \theta}$$

Equation 7

The data were then converted back to DN for data delivery in the unsigned 16-bit data format (JAXA-EORC, 2016).

Four mosaics from 2007-2010 were acquired. Metadata at a pixel level is provided in the form of local incidence angle (the angle from nadir at which the sensor images a pixel, including the contribution of topography) and number of days after launch of image acquisition, allowing calculation of acquisition date (Figure 5.1). It is possible to see an indication of the scene relief in the local incidence angle (top right box in Figure 5.1), and how this is not visible in the HH and HV backscatter intensity images, unlike in

version 1 of the K&C Initiative mosaic releases analysed in Chapter 4 (Figure 4.1). A global forest/non-forest product based on the ALOS-PALSAR data is provided, at the same scale.

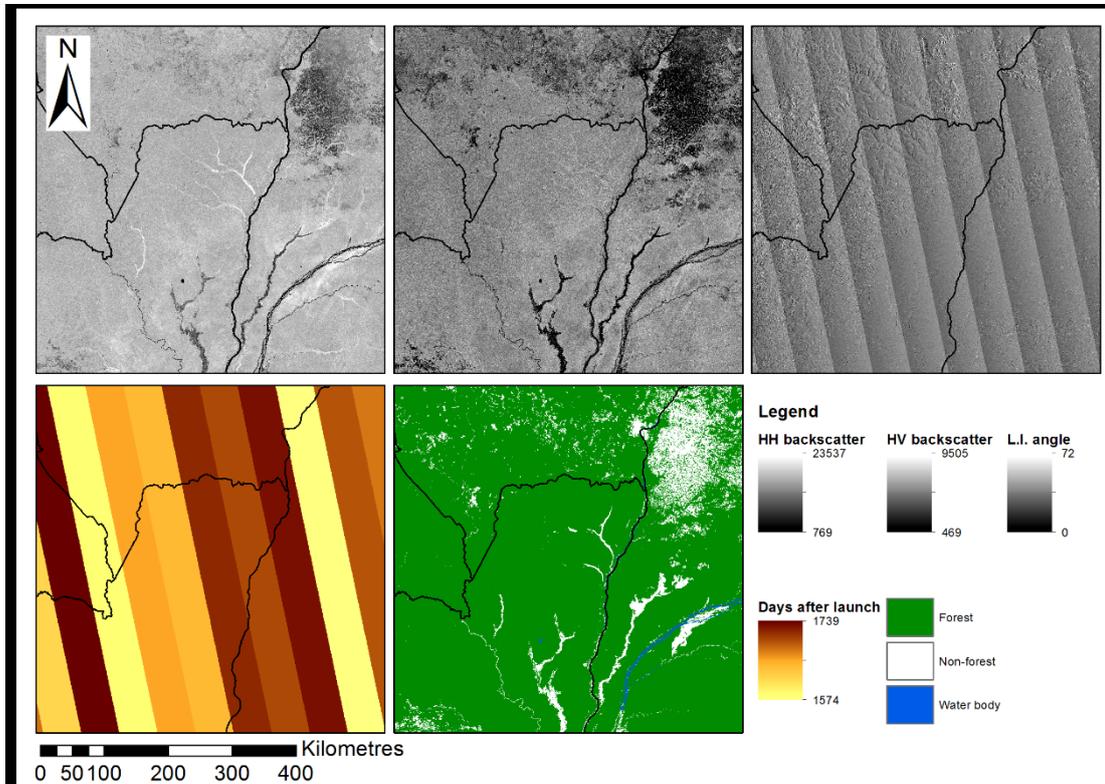


Figure 5.1: 2007 K&C Initiative products in the case study area. Clockwise from top left: HH gamma nought (γ^0) backscatter; HV gamma nought backscatter; local incidence angle; forest/non-forest map; date of data strip acquisition (days after launch)

5.2.2 JAXA K&C Forest / Non-Forest

The JAXA K&C FNF map is the same as described in section 4.2.3, but with four annual products analysed from the years 2007-2010 rather than just one from 2009, with the coverage seen in Figure 5.1 above.

5.3 Methods

The workflow in Figure 5.2 shows the processing steps taken, and a more detailed description follows.

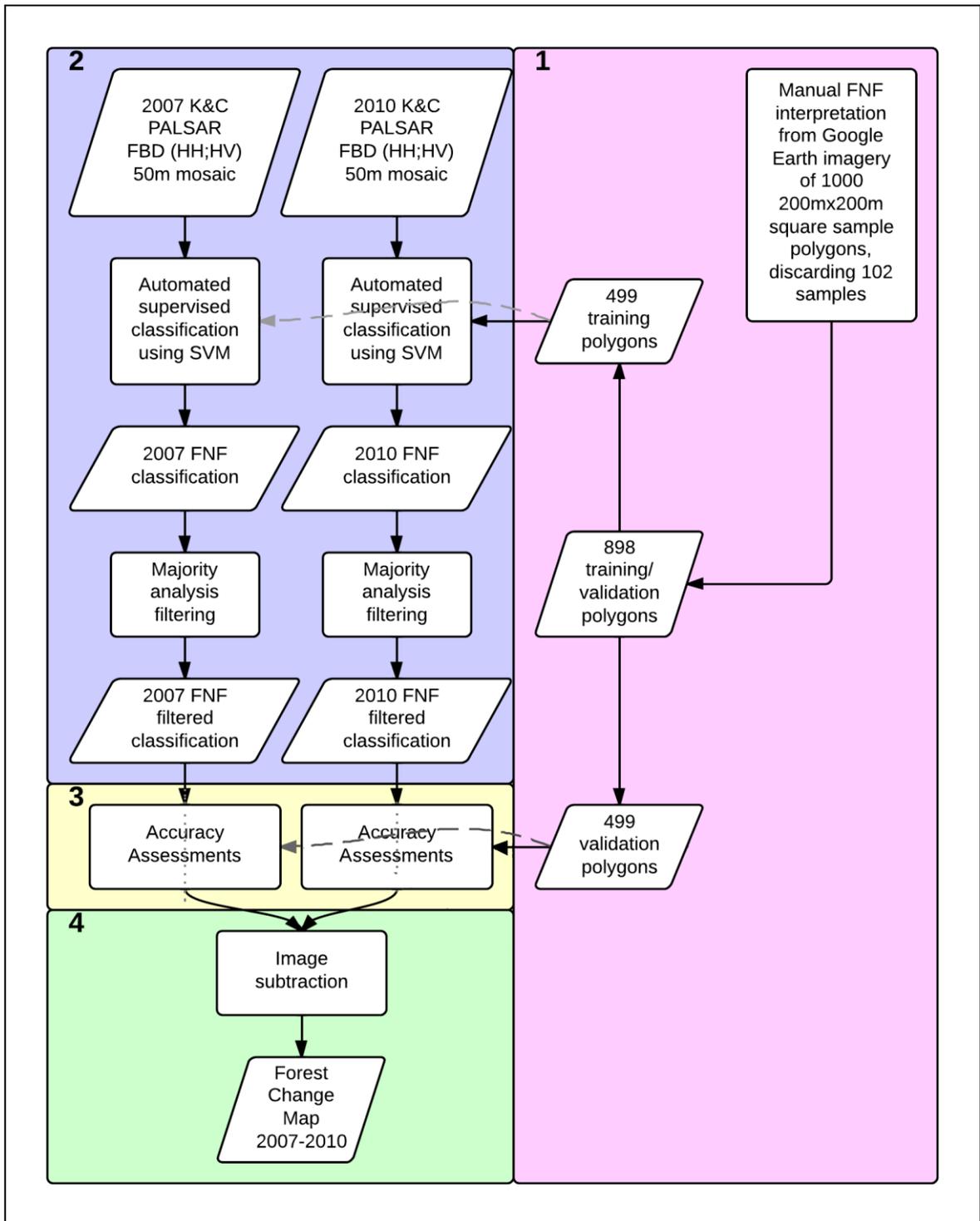


Figure 5.2: Processing steps for 2007-10 forest change map based on support vector machine classification. Numbered coloured boxes show 1: selection of training/validation samples; 2: SVM classification steps; 3: accuracy assessment; 4: calculation of forest change.

5.3.1 Data and Pre-processing

Analysis was performed on a 5° by 5° square with an upper left coordinate of Lat. 5°N, Long. 15°E (Figure 5.3). K&C Initiative JAXA ALOS PALSAR data is organised into grids of 5° lat/long mosaicked squares. The selected area covers part of four countries: Cameroon, the Central African Republic, the Republic of Congo, and the Democratic Republic of Congo, and contains a wide range of land cover types, including: rain fed croplands, mosaic cropland/vegetation, broadleaved deciduous forest, shrub land, herbaceous vegetation (savannah), permanently and regularly flooded forest, urban areas and water bodies².

² Landcover classes taken from ESA's GlobCover 2009 product, (Bontemps et al., 2010)

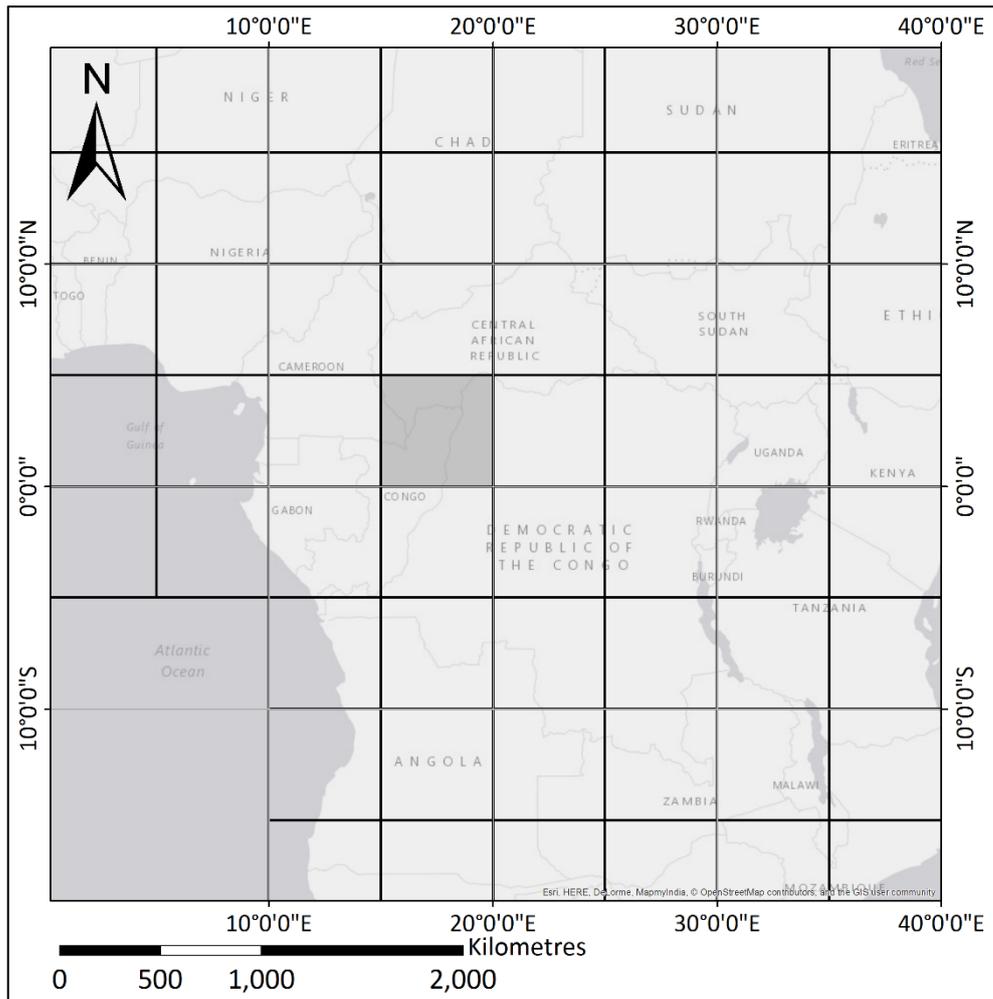


Figure 5.3: K&C Initiative ALOS-PALSAR mosaic (Version 2) tiles over Central and West Africa, with the selected tile for objective two shaded. Background maps from ESRI ArcGIS software used under academic licence courtesy of Esri, HERE, DeLorme, MapmyIndia, © OpenStreetMap contributors, and the GIS user community

K&C Initiative mosaic products are pre-processed as described in section 5.2.1. After download of 25 1° x 1° tiles were combined into the 5° x 5° tile seen in Figure 5.3. Logarithmic [dB] γ^0 values were converted from provided DN values using the same formula and calibration factor as in Equation 6. While the mosaic data has been multi-looked (16 looks; 2 in range and 8 in azimuth) and resampled to 50 m resolution, speckle filters were applied and tested to allow comparison with the JAXA K&C FNF product, which uses a median speckle filter in pre-processing, along with segmentation (Shimada et al., 2014). Three speckle filters were tested at this stage, a single image median speckle filter applied over a 3 x 3 pixel kernel, a boxcar (mean) multi-temporal speckle filter applied on the full 4-image timeseries on a 3 x 3 pixel kernel, and the same multi-temporal speckle filter on a 5 x 5 pixel kernel. Section 2.3.8 has more detailed

descriptions of the filters and the formulae for the multi-temporal filter. The smaller (3 x 3 is the minimum size) kernels were chosen as the noise of the mosaicked imagery is already reduced by the pre-processing steps mentioned, and it was intended to reduce the potential loss of spatial resolution. These three speckle filter methods were compared with an unfiltered version of the mosaics.

5.3.2 Generating Training and Validation Data

A training and validation dataset for the study was created from a random sample of 1,000 square (200m x 200m) plots covering the area of interest. Medium and high resolution Google Earth imagery (GEI) were used to determine the class in each sample square, and there was a loss of 102 samples due to mixing of classes, cloud cover, some poorly georeferenced high resolution imagery, and lack of clarity in the available training data. For the manual classification of sample plots, GEI was analysed throughout the period of data acquisition, where possible, and classes were only assigned where no change had occurred; this allowed the generation of one training/reference dataset over the 4 years of available data. After the reference plots were manually classified, it was found that the data were skewed: out of 898 plots, 778 were in the forest class. To allow an unbiased classification and accuracy assessment, a random sample of 120 forest plots were selected from the 778, thus creating a stratified sample with both classes equally represented. The resulting two sample sets (forest and non-forest) were each randomly divided 50:50 into training and reference vector data. The distribution of reference and training data in both classes is shown in Figure 5.4. While the forest plots are evenly distributed around the study site, non-forest plots are mostly clustered in the upper left quarter of the region, in CAR and DRC, although there is some representation in the Republic of Congo. This is due to the sampling strategy, as well as the predominance of forest classes throughout the study site. For the training and reference datasets, the Spatial Autocorrelation (Moran's-I) tool in ArcGIS returned a Moran's I value of 0.595. Moran's I is a spatial autocorrelation coefficient, where a value of -1 is given to evenly dispersed data, and +1 is given to completely clustered data (Legendre, 1993). This value points to some clustering of the sampled plots.

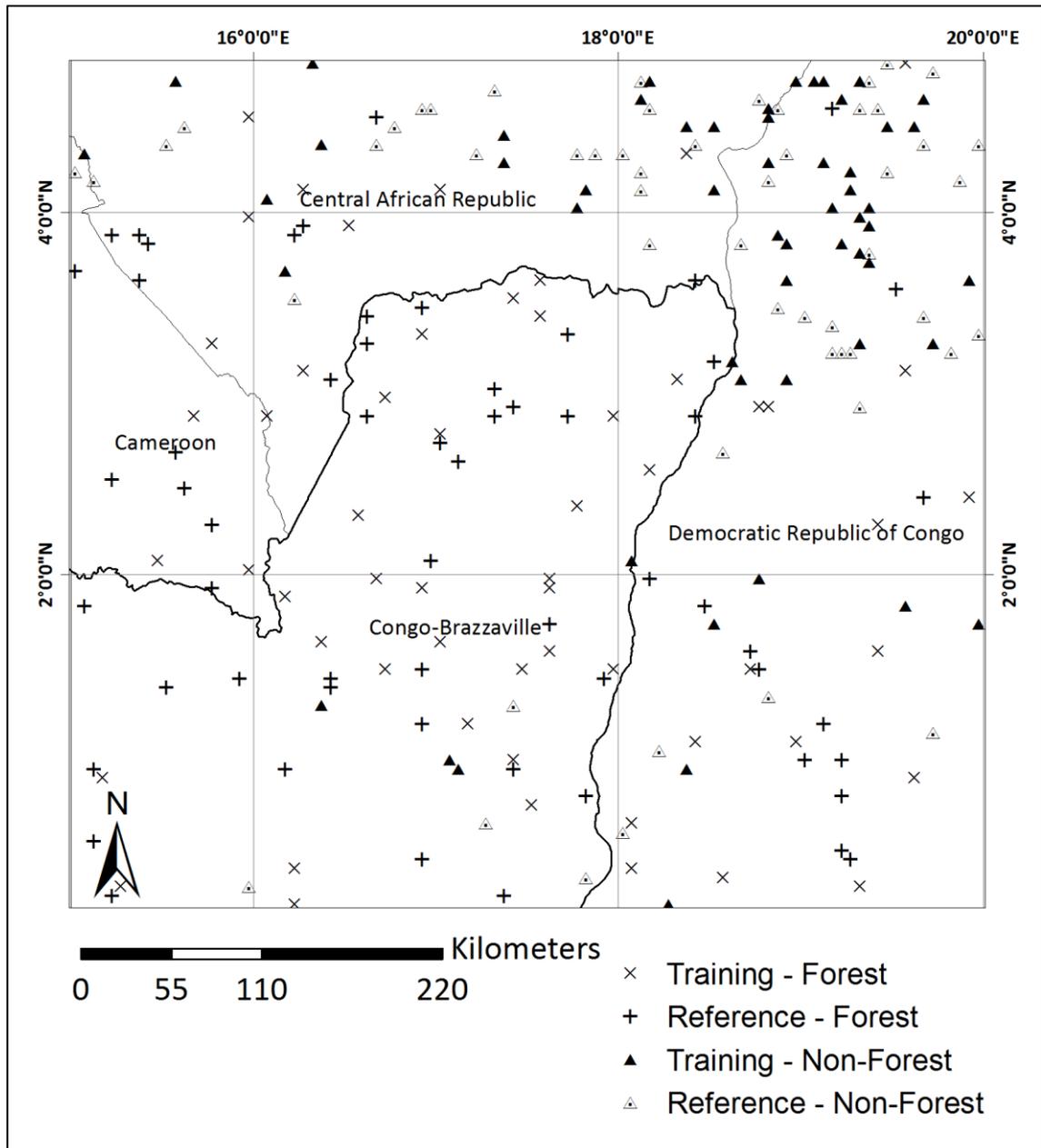


Figure 5.4: Distribution of training and reference/validation plots in the study site

5.3.3 Classification

As multiple classifications would be performed, on speckle filtered and unfiltered data over four annual products, totalling 16 classifications, a fast and efficient classifier was necessary. The training data described the classes, so a supervised classification method was sensible. Using the training data from both classes of sample plots, a support vector machine (SVM) classification, based on LIBSVM (Chang & Lin, 2011) and implemented through the Orfeo toolbox machine learning library, was applied to the HH and HV data

for each individual data year. SVM was chosen as it is an easily implemented (available in several mainstream remote sensing software packages, and with open source packages in multiple programming languages) classifier that has been shown to perform well and efficiently at separating a small number of classes from a small feature space, which describes the task required of this study. As part of the SVM implementation in Orfeo toolbox, data are automatically normalised (the algorithm requires global image statistics for mean and standard deviation), a prerequisite for SVM classification. For the unfiltered data, a filtering algorithm using a majority analysis kernel was applied to the resulting class image to remove and reclassify lone pixels to the dominant surrounding class, effectively creating a minimum mapping unit of 4 pixels (~1ha). Parameter optimisation is performed automatically in the Orfeo toolbox LIBSVM Image Classifier to refine the input parameters, in particular the γ parameter in the linear and RBF kernel functions, using a k-fold cross validation on subsamples of the reference dataset. The RBF kernel function for the SVM classifier (see section 2.4.2) was selected after testing all four available kernel options (linear, radial basis function, polynomial and sigmoid) on a sample area. Figure 5.5 shows the results of this test in a sample area (analysis was performed over the entire study site for the 2010 K&C Initiative mosaic dataset in HH and HV channels, using the multi-temporal speckle filtered product). There is little to distinguish RBF, linear, and polynomial kernel functions, while the sigmoid function performs relatively poorly. Overall classification accuracies using the full reference plots dataset were 91.42%, 91.69%, 91.96% and 83.79% for RBF, linear, polynomial and sigmoid kernel functions respectively. The parameter optimisation resulted in a γ value of 12.1257 for the RBF classifier in the sample dataset.

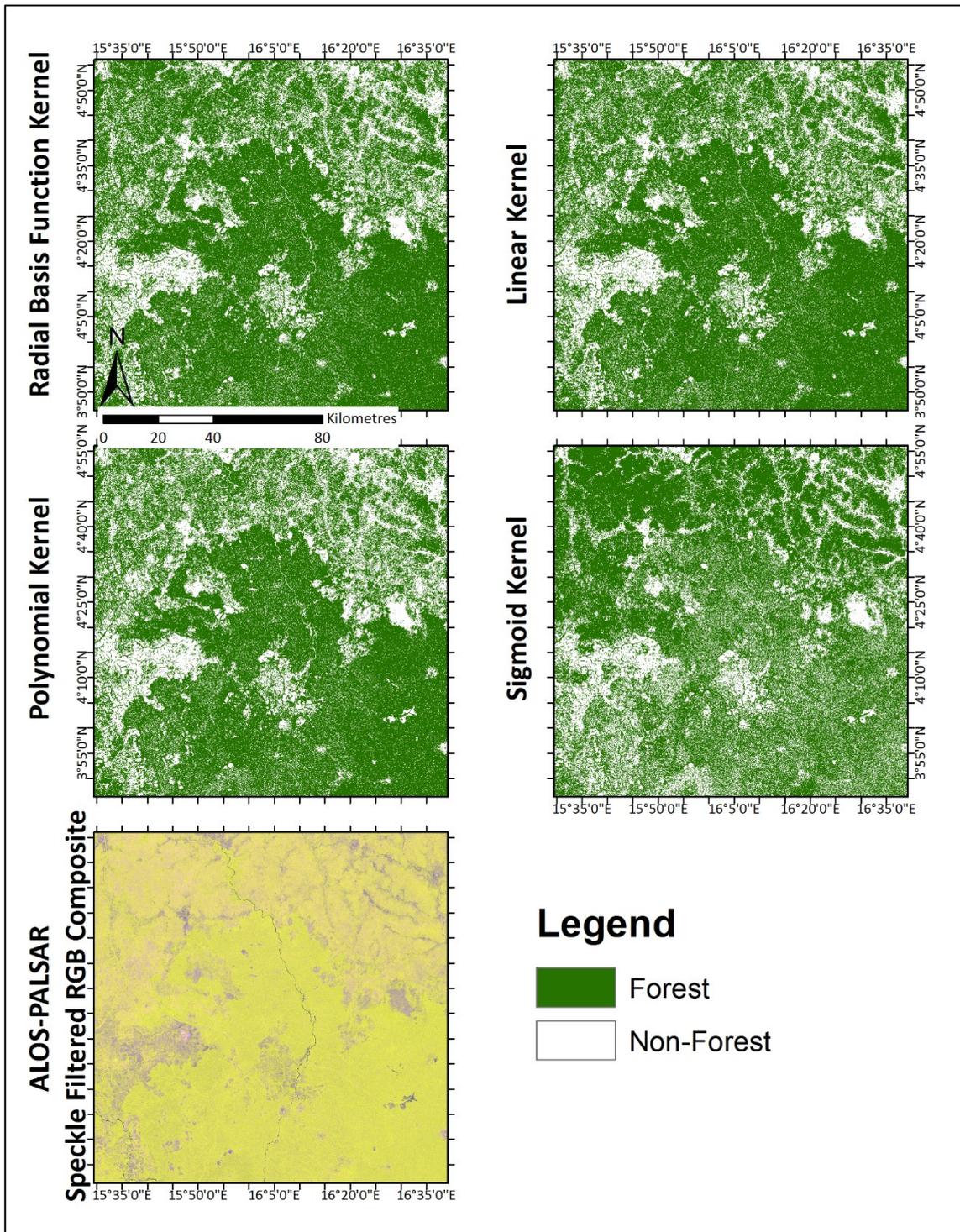


Figure 5.5: SVM Kernel function classification comparison on sample area in the upper left part of the study site: (top four images) with 2010 mosaic of ALOS-PALSAR γ^0 (RGB composite of HH, HV and HH-HV) for comparison

5.3.4 Change Analysis

Change analysis from the four years available was performed post-classification, using class differences from year to year. As part of the change analysis, areas that changed class (from forest to non-forest or vice versa) more than once were noted and masked from the final change product, similar to the technique of Carreiras et al. (2014), which used a system of disallowed transitions between cover types. This is based on the assumption that regrowth of forest following a deforestation event is unlikely to take place in two years, and as a result any pixels where multiple changes occurred were likely to contain classification errors. Areas of class change (forest change) was computed between consecutive years, as well as from 2007-2010 with the removal of multiple change pixels.

5.3.5 Accuracy Assessment

Following classification, an accuracy assessment was carried out using the reference samples from both classes (60 200 m x 200 m plots in each of the two classes, forest and non-forest). The distribution of the reference plots is seen in Figure 5.4, and like the training data, there is some spatial bias in non-forest reference plots due to the extensive forest coverage throughout the areas of the Republic of Congo and Cameroon contained by the study site. While a greater number of samples is advised in the literature – 75-100 per class for accuracy assessment in areas greater than 1,000,000 acres (Congalton & Green, 2009 p75) – the 60 sample plots per class cover at least four pixels each, increasing the number of samples used. In addition, the potential errors at the edges of the square plots used, in terms of inconsistent overlap with the 50 m x 50 m ALOS-PALSAR pixels, was accounted for in the reference data manual classification. This accuracy assessment was also applied to the JAXA K&C Initiative FNF products. A Kappa analysis was performed on each error matrix; the resulting Kappa coefficient estimate (κ) is a multivariate measure of agreement between the correctly classified samples (diagonals in an error matrix) and the row and column totals, giving a statistical measure of whether the classification performs better than chance allocation of classes (Congalton & Green, 2009 p105). The formula for estimating κ is given below:

$$\kappa = \frac{\rho_o - \rho_e}{1 - \rho_e}$$

Equation 8

where ρ_o is the actual agreement in classes (the accuracy) and ρ_e is the chance agreement. The κ coefficient allows more meaningful comparison between error matrices rather than just comparing overall accuracies.

In addition, the error matrices were used to derive confidence statistics from Olofsson et al. (2013); map category (i) and reference category (j) data were used to derive error adjusted areas for forest and non-forest (i.e. \hat{A}_j) for each classification

$$\hat{A}_j = A_{total} \sum_i W_i \frac{n_{ij}}{n_i}$$

Equation 9

where A_{total} is the total study site area, W_i is the proportion of A_{total} reported as map class i , n_{ij} is the cell position in the error matrix for map class i and reference class j , and n_i is the total of the cells in the map row for class i . The stratified error adjusted area includes errors of omission rather than commission. The importance is thus to understand the statistical effect of classification errors on the reported area of individual classes in a final classification, which may otherwise be obscured by the sum of omission and commission errors. An estimate of standard error of the area proportion ($S(P_{.j})$) was computed to give upper and lower 95% confidence intervals (Olofsson et al., 2013):

$$S(\hat{P}_{.j}) = \sqrt{\sum_{i=1}^q W_i^2 \frac{\frac{n_{ij}}{n_i} \left(1 - \frac{n_{ij}}{n_i}\right)}{n_{i\cdot} - 1}}$$

Equation 10

The standard error is converted to error of area proportion by multiplying the value of $S(P_{.j})$ for each class by the total area, and upper and lower 95% confidence intervals are double that value.

5.3.6 Software Used

Initial processing used a combination of Matlab (version 2013a) and R (version 2.15) to download and process zipped SAR files contained on the JAXA K&C Initiative FTP server. Scripts involved simple directory and file manipulation, and mosaicking using the R raster package (version 2.3-12) and as such did not require reproduction in this thesis. The open source Orfeo Toolbox machine learning software was used to apply the SVM classifications and generate error matrices. Training and reference data shapefiles were generated and manipulated within ESRI ArcGIS (version 10.2.2), which was also used to produce the final map figures. Error matrices and error adjusted values were computed in Microsoft Excel.

5.4 Results

5.4.1 Error Matrices

Error matrices were produced first to compare the filtered and unfiltered classifications in a single year, 2010, in order to select the most appropriate filter to use. Table 5.1 shows the results of this analysis. There is a marginal increase in accuracy between the post-classification majority filtered product and the unfiltered products, but neither post-classification filtered products compare favourably with the maps that were speckle-filtered in pre-processing. Between the speckle-filtered products, the multi-temporal speckle filtered products exhibit lower overall accuracies and kappa coefficients than the median (single image) speckle filtered product. For further analysis with the JAXA FNF product, the median speckle filtered product was selected.

Table 5.1: Error matrices for different SAR filtering options from 2010 data. Units for F (forest), NF (non-forest) and Total columns are number of sample points

SAR Filter Used		Reference Data	Reference Data		Total	Overall Accuracy	Producer's Accuracy	User's Accuracy
			F	NF		[%] Kappa coefficient		
Majority Filtered post-classification	F	960	0	960	82.7083	0.743	1	
	NF	332	628	960	0.6542	1	0.6542	
Unfiltered	F	959	1	960	82.6042	0.7423	0.999	
	NF	333	627	960	0.6521	0.9984	0.6531	
MultiTemporal	F	939	21	960	89.8438	0.8437	0.9781	
Spk Filtered 3 x 3	NF	174	786	960	0.7969	0.974	0.8188	
MultiTemporal	F	937	23	960	90.3125	0.8518	0.976	
Spk Filtered 5 x 5	NF	163	797	960	0.8062	0.972	0.8302	
Median	Spk Filtered 3 x 3	F	952	8	960	94.7396	0.911	0.9917
		NF	93	867	960	0.8948	0.9909	0.9031

Error matrices were computed for all 4 years from the median speckle filtered and JAXA FNF products and are reproduced in Table 5.2. The median speckle filter applied individually to images outperformed the JAXA FNF product in overall accuracy and kappa coefficient. There were very few instances where the forest class was misclassified as non-forest in the JAXA FNF product, and far more non-forest classed as forest, suggesting the model is overfit for the forest class. The same is true of this study's median speckle filtered SVM approach, but to a lesser extent.

Table 5.2: Confusion Matrices in sample numbers for each class – Forest (F) and Non-Forest (NF) – from this study’s median speckle filtered SVM classifier and JAXA’s FNF data against the validation reference dataset. Percentage overall accuracy, producer’s and user’s accuracies (for both classes) and the estimated Kappa coefficient are also presented

		2007		2008		2009		2010	
		Reference Data		Reference Data		Reference Data		Reference Data	
		F	NF	F	NF	F	NF	F	NF
Median Spk	F	950	10	956	4	933	27	952	8
Filtered 3 x 3	NF	59	901	73	887	58	902	93	867
Producer's Acc		0.9415	0.989	0.9291	0.9955	0.9415	0.9709	0.911	0.9909
User's Acc		0.9896	0.9385	0.9958	0.924	0.9719	0.9396	0.9917	0.9031
Overall Acc		96.406		95.99		95.573		94.74	
Kappa		0.9281		0.9198		0.9115		0.8948	
JAXA FNF	F	958	2	960	0	956	4	944	16
	NF	290	654	284	660	417	527	422	522
Producer's Acc		0.7676	0.997	0.7717	1	0.6963	0.9925	0.6911	0.9703
User's Acc		0.9979	0.6928	1	0.6992	0.9958	0.5583	0.9833	0.553
Overall Acc		84.664		85.084		77.889		76.996	
Kappa		0.6925		0.7009		0.5561		0.5382	

5.4.2 Forest Area Change

Forest areas, in km² as well as percentage of total study site area, were calculated for this study’s supervised classification and the JAXA FNF product, for each year. These, along with total forest gain and loss (from 2007 to 2010), and the area of pixels with multiple class changes (which were masked from the total forest gain and loss figures) are reported in Table 5.3 below.

Table 5.3: Forest area in km² and as percentage of total study site area, with class change areas for this study’s supervised SVM classification and the JAXA FNF product, and the area of multiple class changes (>1 change over the 4 year period). Loss and Gain figures exclude pixels with multiple class changes

		Year				2007-2010 Change		
		2007	2008	2009	2010	Loss	Gain	Mult. Change
Median	Area [km2]	249291	254398	247412	252936	6542	8512	15240
Spk Filter	%Total Area	81.09	82.75	80.48	82.28	2.13	2.77	4.96
JAXA FNF	Area [km2]	274476	273252	279207	278438	4355	8377	11710
	%Total Area	89.3	88.9	90.84	90.59	1.42	2.73	3.81

The difference between areas reported each year for both classifications is shown in Figure 5.7. There is no agreement between both products in area reported, with the exception of the value for forest gain between 2007 and 2010. The values for loss and gain are relatively high, and could be a function of changing image statistics between acquisitions; it is unlikely to have such a high value for forest gain over as short a time span as 4 years.

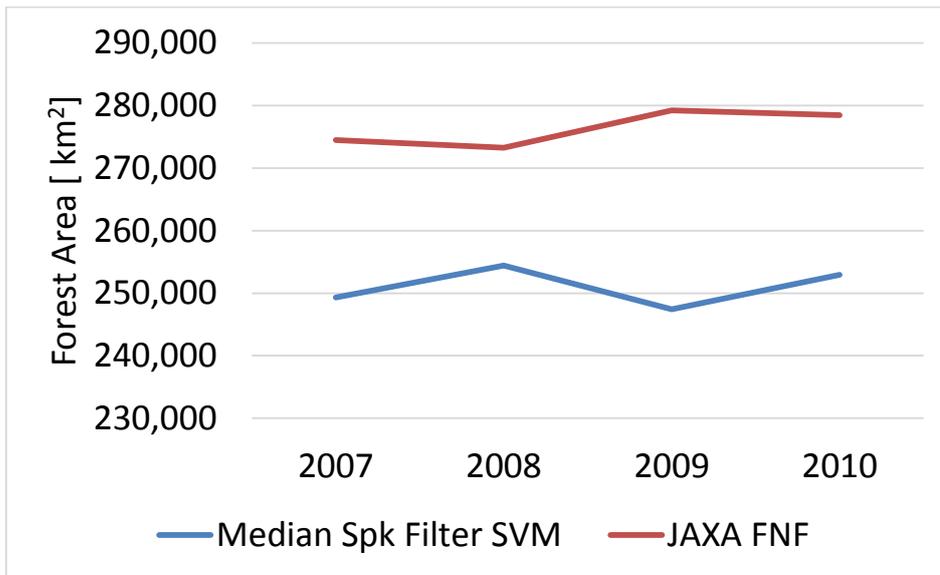


Figure 5.6: Reported forest area in km² by both classifications over the timeseries (2007-2010)

The error adjusted area values for forest and non-forest classes in Figure 5.7 and Figure 5.8 were calculated using the formula described in section 5.3.5, and the figures also show error bars (standard error). The values represent the area for each class, adjusted for errors of omission, so are completely different from the values in Figure 5.6, in that values are considerably lower, showing only the areas that can be reliably reported for each class, with the error bars showing upper and lower confidence intervals. The missing areas from the error adjusted values represent the area of errors of commission in the map.

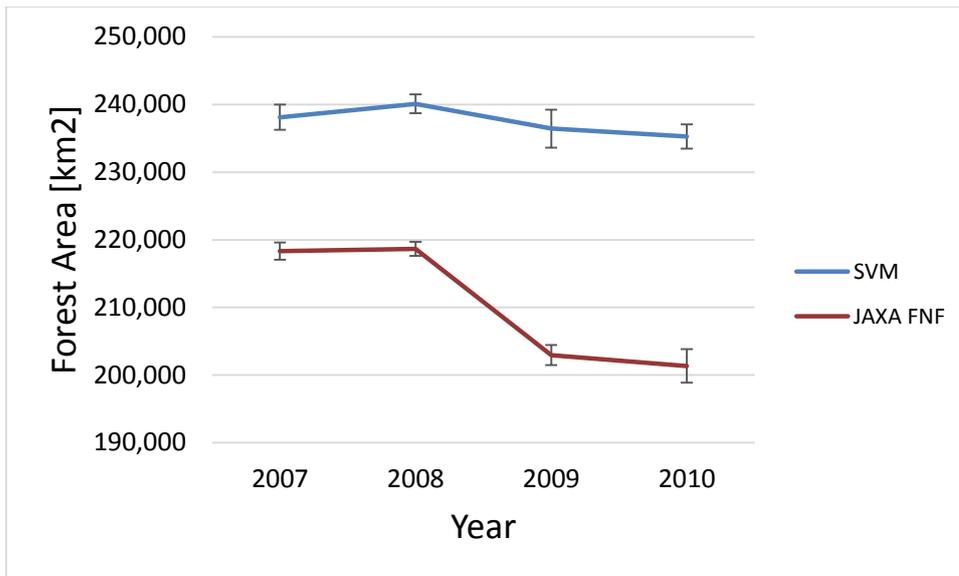


Figure 5.7: Graph of error adjusted forest area by year, showing differences between this study's supervised classification (Median speckle-filtered SVM) and the approach used for JAXA's FNF map; the figure shows the area that is classified as forest with 95% confidence error bars, computed using the formulae from equations 9 and 10

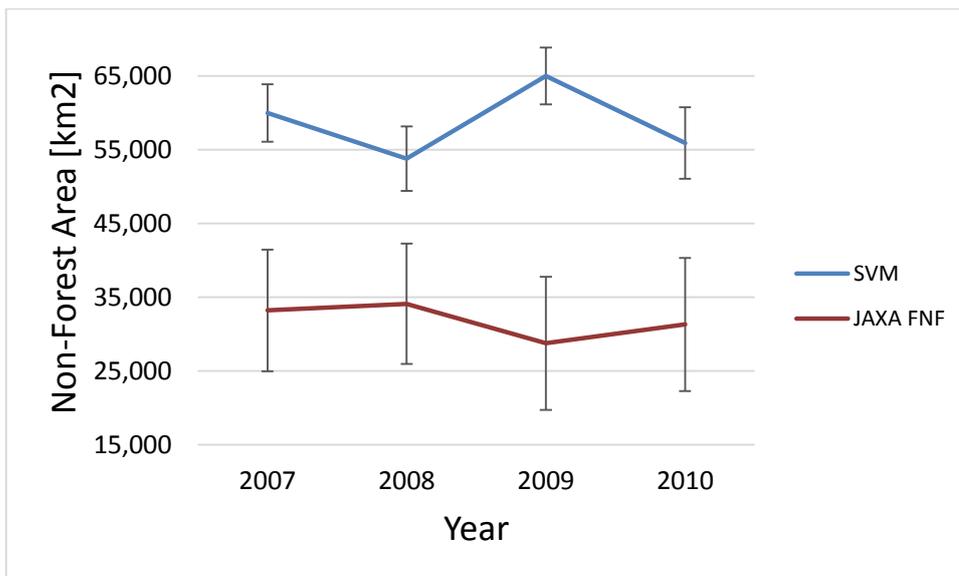


Figure 5.8 Graph of error adjusted non-forest area by year, showing differences between this study's supervised classification (Median speckle-filtered SVM) and the approach used for JAXA's FNF map; the figure shows the area that is classified as non-forest with 95% confidence error bars [km²], computed using the formula from equations 9 and 10

To give a better indication of the proportion of the study area reported, the adjusted areas for both classes are shown in Table 5.4 as the proportion of the total study site area. From this table it is clear that the forest changes from year to year in Figure 5.6 cannot be described as significant. In fact, the area of forest that is accurately mapped

by the Median Speckle Filter year to year remains very stable, while the adjusted area for the JAXA FNF product reduces between 2008 and 2009.

Table 5.4: Error adjusted areas and 95% confidence margins of error, expressed as percent of total study site area

		2007	2008	2009	2010
Median Spk Filter	F	0.77 ± 0.006	0.78 ± 0.005	0.77 ± 0.009	0.77 ± 0.006
	NF	0.2 ± 0.013	0.17 ± 0.014	0.21 ± 0.013	0.18 ± 0.016
SVM	F	0.71 ± 0.004	0.71 ± 0.003	0.66 ± 0.005	0.65 ± 0.008
	NF	0.11 ± 0.027	0.11 ± 0.027	0.09 ± 0.029	0.1 ± 0.029
JAXA FNF	F	0.77 ± 0.006	0.78 ± 0.005	0.77 ± 0.009	0.77 ± 0.006
	NF	0.2 ± 0.013	0.17 ± 0.014	0.21 ± 0.013	0.18 ± 0.016

5.4.3 Class Stability of Pixels

The class stability of pixels was used as a measure to remove probable classification errors from the reported forest change. From Table 5.3 it is clear that a high percentage of pixels exhibit multiple (2-3) changes between 2007 and 2010 in both products, almost equalling the totals for one change (either forest gain or forest loss) in the same period. The spatial extent of class stability is seen in Figures 5.11-5.13. Pixels with multiple class changes are more prevalent at class boundaries, particularly in the JAXA FNF product, and are a feature of noise that was not dealt with sufficiently by the speckle filter in this study.

5.4.4 Classification Differences

Figure 5.7 shows the forest area by year calculated from the two classification methods examined. A rise in forest area is observed using both methods, though a smoother change is seen using the SVM classifier (Figure 5.7). Annual fluctuations described by the K&C Initiative FNF method appear to be more indicative of seasonal changes in moisture or flood conditions from different acquisition dates from year to year. This is illustrated by the comparison between the two approaches in Figure 5.9. Analysis of the 2007–10 forest loss/gain product generated (Figure 5.10) gave a net increase in forest area of approximately 1,970 km², the equivalent of 0.64% of the study area. While these figures are difficult to verify without ground data, and may therefore not be a reliable indicator of actual forest area change, the class changes (both gain and loss) from year

to year seen in Figure 5.9, and in closer detail for the entire period in Figure 5.10 - Figure 5.12, often occur in a pattern consistent with clearing along logging and other access trails, and also reflect the fragmented nature of the landscape, particularly in the CAR and DRC areas of the study site. The scale of disturbance seen in this data does not appear in coarser resolution datasets, and is largely absent from JAXA's K&C Initiative FNF maps, due to the segmentation approach used. In addition there are several large features of class change (such as in Figure 5.12) that are due to the division of the dataset into 500 km x 500 km processing tiles during classification (Shimada et al., 2014), with separate image statistics driving the classification threshold, but in particular segmentation boundaries intersecting features at the edge of a processing tile.

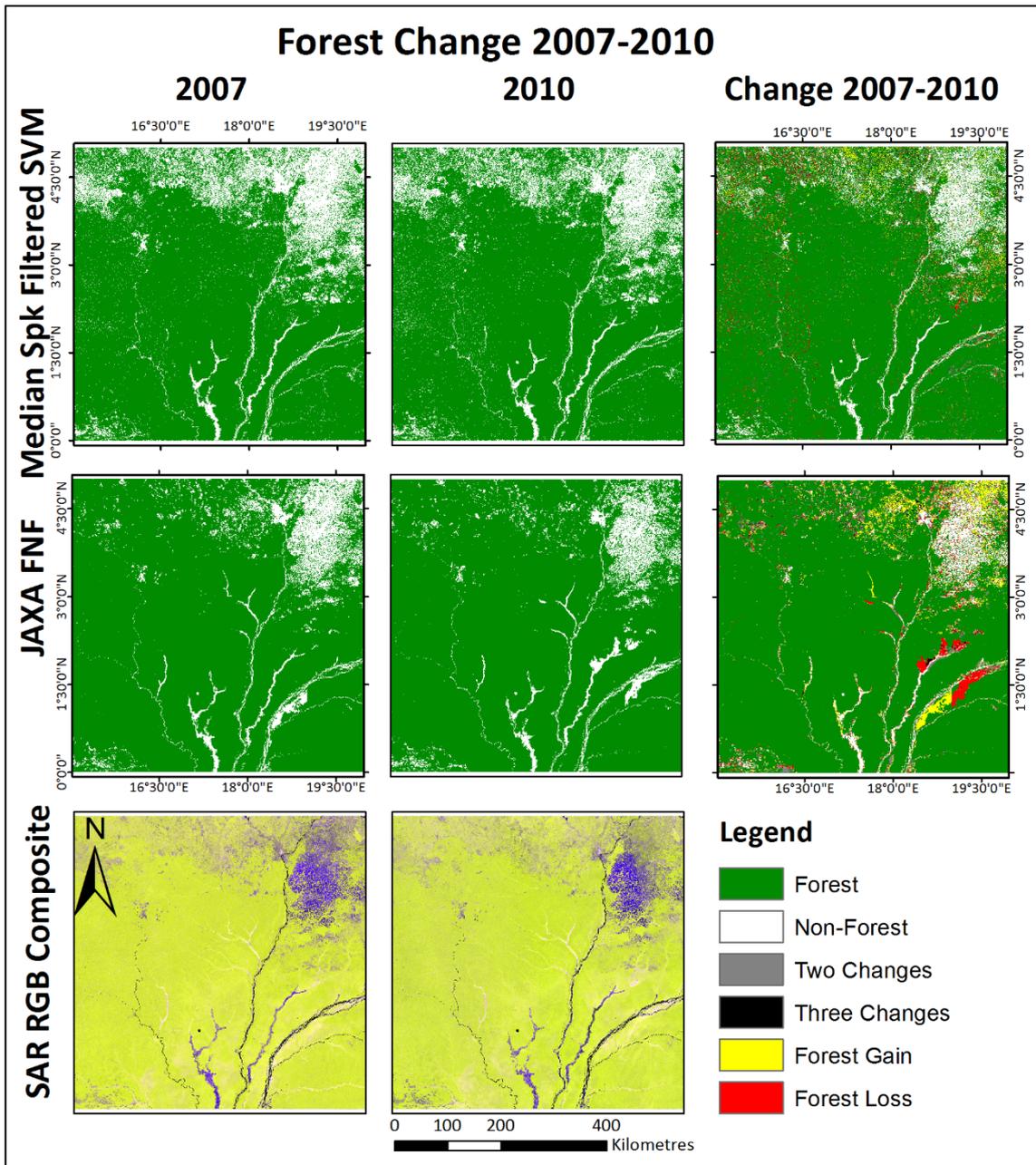


Figure 5.9: Comparison of forest gain/loss derived from SVM classification and JAXA FNF on the study site, with 2007 and 2010 ALOS-PALSAR ψ^0 RGB composites (HH, HV, HH-HV); relatively small scale changes detected in SVM classification, while large clusters of change dominate the JAXA FNF object based classification

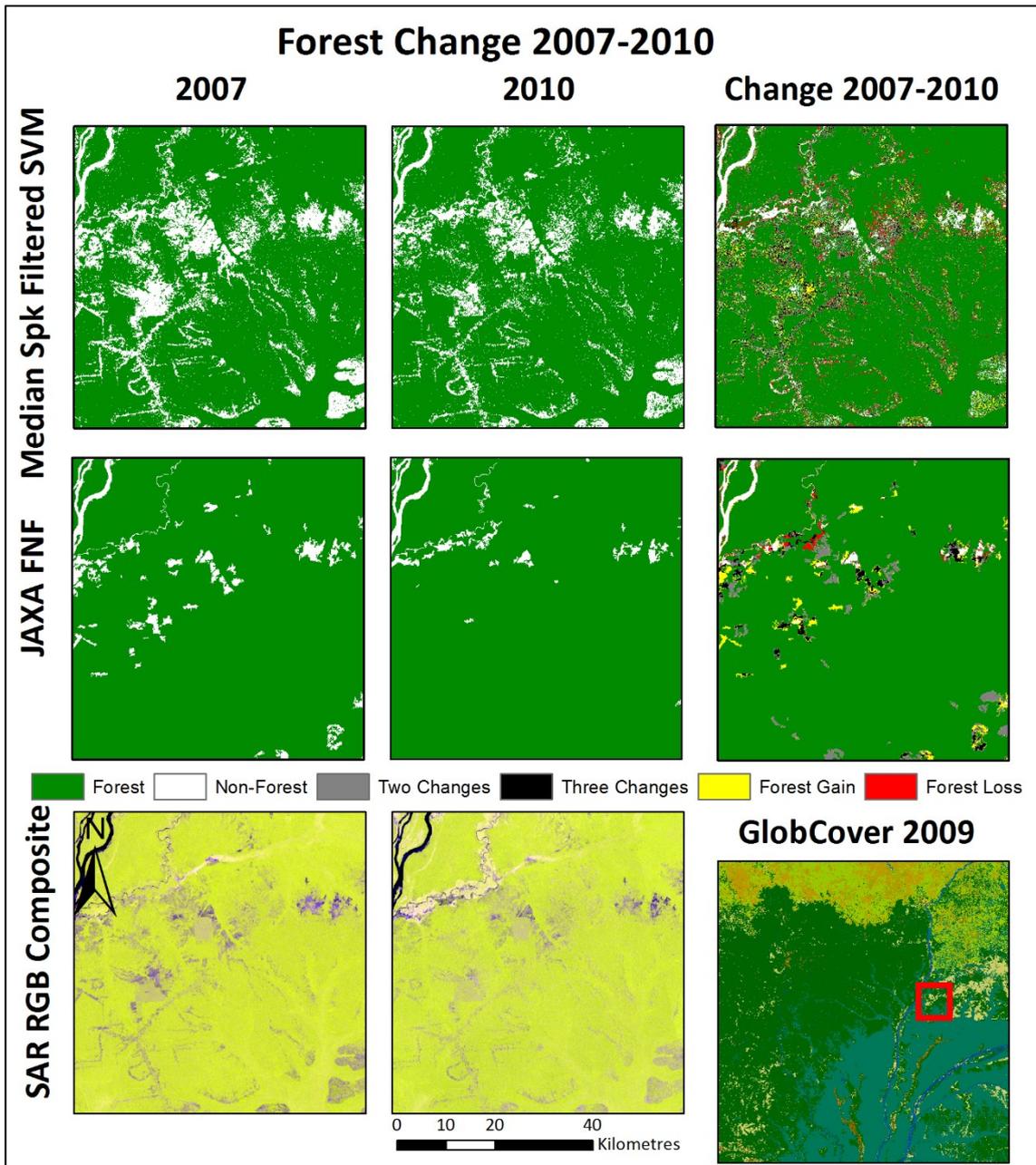


Figure 5.10: Comparison of forest gain/loss derived from SVM classification and JAXA FNF on a subset of the study site, with 2007 and 2010 ALOS-PALSAR γ^0 RGB composites (HH, HV, HH-HV) and extent indicated on a GlobCover 2009 thematic map of the study site (legend in Figure 4.5); logging and agriculture detected in this study's SVM classification, but missed by JAXA FNF segmentation approach; 2010 image brightness due to changing wetness conditions along river in top left of subset does not affect SVM classification but is listed as deforestation by JAXA FNF product

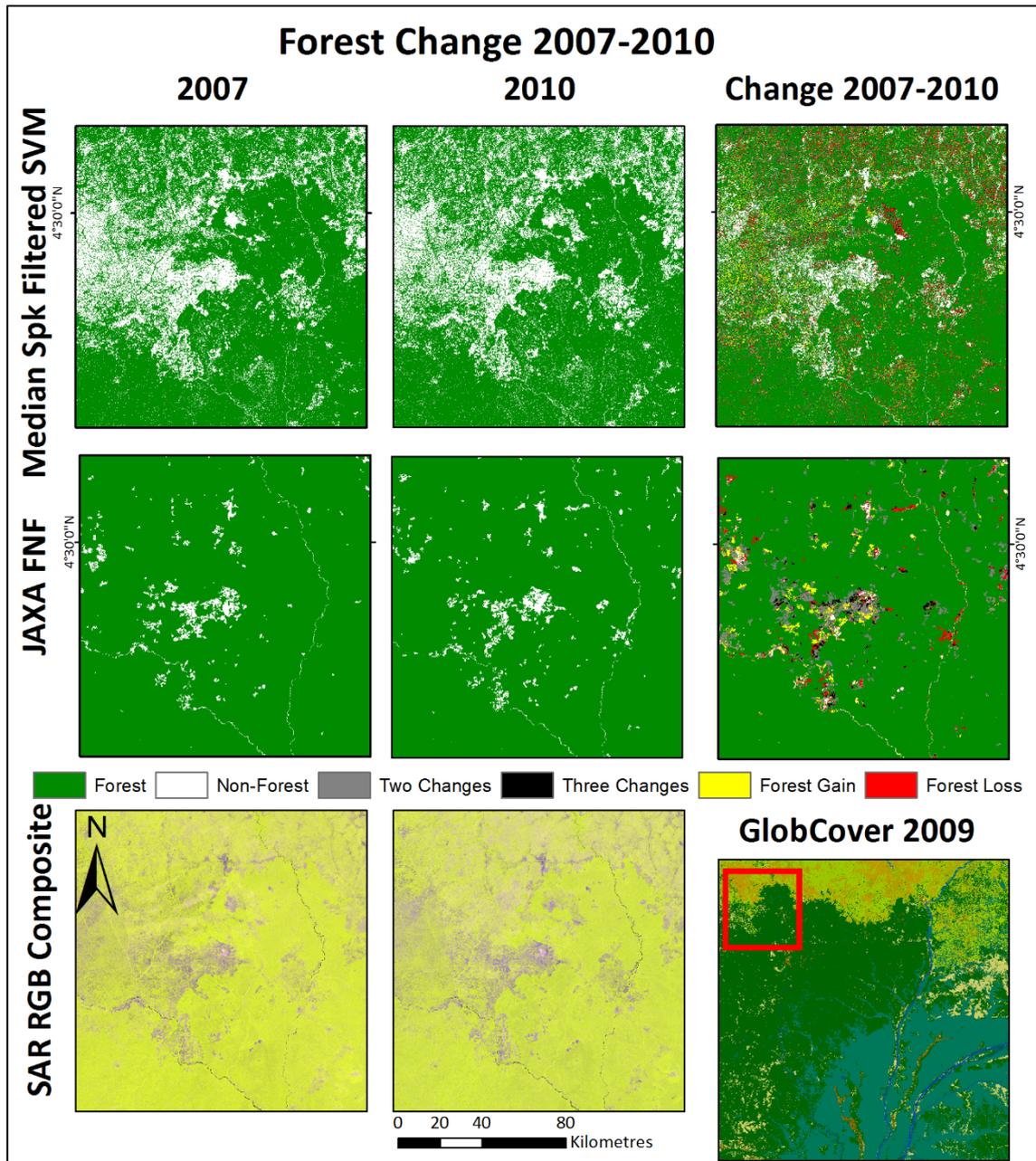


Figure 5.11: Comparison of forest gain/loss derived from SVM classification and JAXA FNF on another subset of the study site, with 2007 and 2010 ALOS-PALSAR γ^0 RGB composites (HH, HV, HH-HV) and extent indicated on a GlobCover 2009 thematic map of the study site (legend in Figure 4.5); fragmented landscape is better represented and an area of deforestation is detected by this study's SVM classification but not in the JAXA FNF product, although both classifications appear noisy

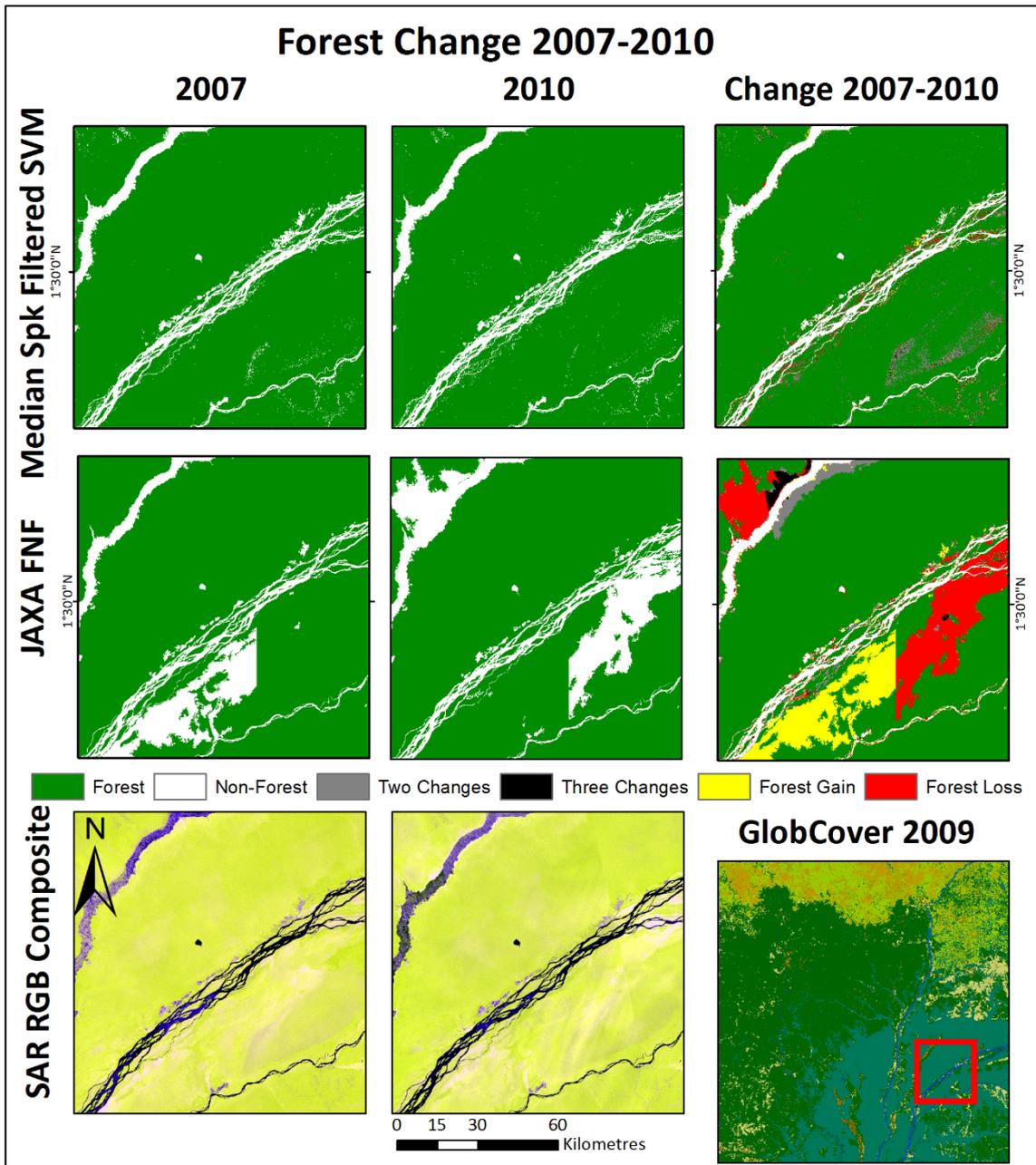


Figure 5.12: Comparison of forest gain/loss derived from SVM classification and JAXA FNF on a further subset of the study site, with 2007 and 2010 ALOS-PALSAR γ^0 RGB composites (HH, HV, HH-HV) and extent indicated on a GlobCover 2009 thematic map of the study site (legend in Figure 4.5); area shows marked linear features due to a classification error in the JAXA FNF product

5.5 Discussion

There is an improvement in accuracy of the classification method used in this study over JAXA's K&C Initiative FNF product in the same area. This may be attributed to their application of a global algorithm, compared with the supervised regional approach described in this study, particularly in forested areas that are prone to flooding. While their forest class uses a regionally changing threshold from the HV channel, their decision tree methodology uses a global HH backscatter threshold for a non-forest 'settlement' class that is within the seasonal range of backscatter from Congolese swamp forest (from preliminary results in Einzmann et al., (2012)). As a result, large areas of forest are classified as non-forest in wet conditions and forest in dry conditions (see Figure 5.9). In addition, the use of segmentation in their approach limits the detection of finer scale forest change that the 50m pixel resolution could otherwise be sensitive to. While the exact figures generated by the study for forest area, and by extension annual forest change, are unreliable without verification from ground data, the method described could be a useful indicator of finer scale change. The results from this study provide justification for applying this study's methodology across the whole Congo Basin, with refinement of the supervised classification parameters, and incorporating available field data from forestplots.net (Lopez-Gonzalez et al., 2011) in training and validation where possible. This will allow a broader analysis of forest change across the entire region, and it will provide further insight into the results of this study, depending on the change seen across the entire Congo Basin.

For a global forest/non-forest classification the JAXA FNF product performs well globally, and it is meeting spatial and temporal resolution requirements for user communities (Bontemps et al., 2012; Ban et al., 2015).

Comparing results to other SAR studies in Table 2.2, the MMU of this study (50m x 50m) is able to detect finer scale changes than that of studies using the CAMP dataset (De Grandi et al., 2000a) at 200m, partly because of the interim advances in processing power and data availability, which allows a higher resolution dataset to be easily distributed to researchers, and allows faster processing of those datasets. The GRFM dataset is also limited by using a single year of data, although the low and high water acquisitions offer an advantage due to the ability to map based on changing seasonal

conditions north and south of the equator. Accuracies from studies using GRFM (Sgrenzaroli et al., 2004; Bwangoy et al., 2010) are not directly comparable to this study due to these differences; there are improvements between this study and those in terms of overall accuracy (95.68% vs 87% and 81.11% respectively), but in each case one or more of the scope, location and/or range of the studies are different.

Similarly, for the other listed studies which used ALOS-PALSAR (Betbeder et al., 2014; Ferreira-Ferreira et al., 2015; Thomas et al., 2015; Qin et al., 2016; Walker et al., 2010), there are some issues which prevent direct comparison. In the case of Betbeder et al. (2014), the methodology was completely different, focusing more on MODIS EVI to identify forest areas, and used the ScanSAR acquisition mode of ALOS-PALSAR, which is a coarser resolution than the strip map mode that was used to create the K&C Initiative mosaics. Only one polarisation was used (HH), and it was specifically used to map below canopy flood extent. In the same way, Thomas et al. (2015) mapped a very selective region of Central Africa, within a fixed distance from the sea (in order to quantify mangrove extent), with no overlap with the study area for this chapter. Both of these studies could help to explain the misclassifications in the JAXA FNF product in areas of changing moisture and wetness conditions, as described on the previous page, by their acknowledgement and use of ALOS-PALSAR to map areas of flooded forest. Ferreira-Ferreira et al. (2015) and Walker et al. (2010) both produced maps closer to the one in this study, in terms of using a similar dataset and techniques (machine learning classifier, ALOS-PALSAR FBD mode, backscatter), although they were both studies in the Amazon basin. Walker et al.'s reported overall accuracy is in line with that reported in this study (92.4% compared with this study's average of 95.68%), and Ferreira-Ferreira et al.'s reported accuracy is lower (83%), but was across 5 classes. Qin et al. (2016) achieved a similarly high reported overall accuracy (95.9%), at a similar spatial resolution and much larger extent; their study extended from tropical humid forests in South East Asia to boreal deciduous forests in North East China.

5.6 Conclusions

In the context of the aims and objectives of this chapter, this study fulfilled the three objectives. The first two objectives, of obtaining and characterising available L-band SAR data, classifying and comparing to other available map products in the Congo Basin, was achieved through the selection of the study site and ensuing speckle filter analysis, SVM classification, and comparison with the JAXA K&C Initiative FNF product. In terms of the third objective, of analysing results and suggesting improvements to data acquisition, release and processing, as well as answering the first research question, a number of points may be highlighted.

First of all, when compared visually with the JAXA FNF product, a smaller MMU for a forest extent map is desirable. The 50m x 50m MMU in this study picked up finer scale changes in forest extent in this region, which were not seen by the JAXA FNF maps. An obvious question leading from this result is whether a finer resolution provided by later releases of the K&C Initiative mosaic would offer further improvements in forest extent and forest change.

Identifying the likely effects of seasonally flooded forest on the JAXA FNF product urban HH channel mask was a positive outcome of this research, and will allow improvements in global FNF products using these or similar datasets.

Following on from the conclusions of Chapter 4, this study illustrated the utility of appropriate metadata and high quality accessible L-band mosaics. Having information about exact date of acquisition and local incidence angle could also assist in identifying and classifying flooded forest areas and other wetlands, as well as allowing the use of ancillary climatic datasets such as rainfall, moisture conditions, and temperature.

The novelty of this chapter can be described by the identification of the faults of the wider area FNF product, and the achievement of a higher overall accuracy of the FNF classification. Although the methodologies themselves are not novel, it has been an achievement in this area of the Congo Basin to perform a rigorous testing of the available speckle filters, apply more robust area error and confidence statistics for forest and non-forest classes, and generate a training and validation dataset which may be made available for future comparative studies.

Chapter 6: Multi-scale Analysis of Forest and Inundated Forest from Small Area SAR Time Series Data

6.1 Introduction / Aim and Objectives

Following on from the work of Santoro et al. (2009) in Swedish boreal forests, this chapter aims to answer how different spatial and temporal resolution of ALOS-PALSAR data can affect the forest area reporting in a region of the DRC. The Kyoto and Carbon Initiative mosaics have been released at differing resolutions (initially 50m for 2008-2009 and an improved 50m version from 2007-10 released in January 2014; a 25m version from 2007-10 in September 2014) and these releases are themselves resampled from a 10m resolution product. The potential differences in forest area reporting using this dataset, as well as in the Forest/Non-Forest product that accompanies the mosaics, originally prompted this research to establish whether this may be significant. It also aims to quantify potential improvements in classification accuracy at different resolutions, and thus inform on error reporting in ALOS-PALSAR based forest map products.

The accompanying Forest/Non-Forest product from the Kyoto and Carbon Initiative mosaic showed several unlikely annual change events in multiple wetland areas in the Congo basin (see Figure 5.9). The decision tree classification algorithm used by JAXA contains a node which used the HH channel to separate urban areas of non-forest (Shimada et al., 2014), without considering the sensitivity of the HH channel to changes in moisture below canopy. Using proximity to water bodies, daily rainfall estimates, regional water basin products from NASA, and HH channel changes between SAR acquisitions, a secondary aim of this chapter was to establish whether flooding was indeed causing these supposed change events and investigate the addition of a class for 'flooded forest'.

The study area for this chapter is described in section 3.1, with a rationale for its selection. To summarise briefly, it was selected for the mixture of regularly flooding forest and mosaic vegetation and croplands, allowing a chance to investigate the flooding effects seen in Chapters 4 and 5, as well as having a reasonable opportunity for investigating forest cover change at different scales through the 4 years that data were acquired. As it is a smaller study site than the previous chapters, there is also greater scope to investigate optimal classification techniques, without complications arising from mosaicking data, such as strip balancing.

6.2 Datasets

6.2.1 ALOS PALSAR L1.1

The ALOS-PALSAR scenes used for this chapter are listed in Table 6.1. A L1.1 (single look complex) time series of ALOS-PALSAR data was used, which was obtained through a research data application by a supervising author (Prof. H. Balzter)³. ALOS-PALSAR data is now available free of charge for research purposes through the Alaska satellite data facility portal (<https://www.asf.alaska.edu/sar-data/palsar/>). They cover an area of wetland forest and rural complex spanning the Congo River in the Mongala district of Équateur province in the DRC, west of the district capital Lisala. The area experiences regular flooding and has a climate characterised by a single short dry season (one month in January/February) and otherwise high rainfall (~150mm per month (Consortium Congo 2010, 2014)). The scene extent is presented in Figure 6.5.

³ ALOS-PALSAR data is now available free of charge for research purposes through the Alaska satellite facility data portal at <https://www.asf.alaska.edu/sar-data/palsar/>

Table 6.1 Description of ALOS-PALSAR scenes acquired; scenes corresponding to the K&C mosaics are in **bold** and one scene discarded after a processing error is in strikethrough and italics.

Date	Polarisations	Range Spacing	Scene ID	Orbit
2007-06-19	Dual (HH;HV)	9 m	ALPSRP074660020	Ascending
2007-09-19	Dual (HH;HV)	9 m	ALPSRP088080020	Ascending
2008-05-06	Dual (HH;HV)	9 m	ALPSRP121630020	Ascending
2008-06-21	Dual (HH;HV)	9 m	ALPSRP128340020	Ascending
2009-06-24	Dual (HH;HV)	9 m	ALPSRP182020020	Ascending
2009-09-24	Dual (HH;HV)	9 m	ALPSRP195440020	Ascending
<i>2010-05-12</i>	<i>Dual (HH;HV)</i>	<i>9 m</i>	<i>ALPSRP228990020</i>	<i>Ascending</i>
2010-06-27	Dual (HH;HV)	9 m	ALPSRP235700020	Ascending
2010-09-27	Dual (HH;HV)	9 m	ALPSRP249120020	Ascending

6.2.2 RFE 2.0

RFE 2.0 data is acquired at a coarse spatial resolution (~11km pixel spacing at the equator) for the whole of Africa and is derived from four sources, described in section 2.3.7. RFE 2.0 daily rainfall estimate data were downloaded for the entirety of the ALOS-PALSAR mission, from January 26th 2006 until June 6th 2011.

6.2.3 HYDRO-1K

HYDRO1K data is a 1km cell resolution global hydrology aide that contains raster and vector derivatives of the GTOPO30 digital elevation model (USGS, 2015b). For this study the basin vector data were extracted for the Congo Basin. The HYDRO1K dataset contains several layers, but the one which was of interest was basin catchment area. These are vector polygons showing basin catchment or sub-catchment areas, with an accompanying Pfafstetter code. The Pfafstetter code is a six level unique identifier, where each level has a corresponding number of digits which describes the flow from sub-catchment to catchment area (Verdin & Verdin, 1999). The dataset was produced using flow analysis of the GTOPO30 30 arcsecond (roughly 1km resolution) DEM, assigning flow direction from a cell using the steepest slope towards a neighbouring cell (USGS, 2015b).

6.2.4 AMSR-E

The daily soil moisture (SM) and vegetation water content (VW) products from the AMSR-E sensor on the joint JAXA/NASA Aqua platform were downloaded, like the RFE 2.0 dataset, for the entire ALOS-PALSAR mission duration. The soil moisture product, measured in g/cm^3 , is derived from the ratio of horizontal and vertical polarisations of X- and C-band passive microwave brightness data received by AMSR-E (Njoku et al., 2003). The VW product (measured in kg/m^3), is based on the polarisation ratio at C-band, but is not considered decoupled from the effects of surface roughness, so is more accurately described as an effective vegetation water content (Lucas et al., 2010).

6.2.5 SRTM

NASA's Shuttle Radar Topography Mission (SRTM) DEM was used for automatic orthorectification of geometric distortions, and for topographic slope correction. The SRTM dataset is a near global (56° S to 60° N) DEM generated from analysis of data gathered by dual wavelength (C- and X- band) SAR sensors onboard the Space Shuttle Endeavour over an 11-day mission in 2000, allowing repeat imaging of most landmasses (most of Africa was acquired twice, with a small area imaged just once) (Hennig et al., 2001). Single pass InSAR analysis (made possible by a second signal receiver on the end of a 60 m long boom) was performed on the data to produce a 3 arcsecond (~ 90 m) DEM (SRTM-3). The design requirements for the SRTM-3 DEM included a 16 m absolute vertical error limit, and it has generally exceeded that requirement (Rexer & Hirt, 2014).

6.3 Methods

The approach used in this chapter involves a detailed comparison of classifications of an eight-date time series of ALOS-PALSAR data, to determine how the use of different spatial resolutions affects the reporting of forest area. They were also compared with existing forest/non-forest classifications at different spatial resolutions from the JAXA K&C Initiative.

6.3.1 ALOS PALSAR Pre-processing

The nine available scenes represent all that were acquired by ALOS in the study area in this mode (fine beam dual polarisation, ascending orbit). Four scenes (one per year from 2007 to 2010) corresponded to the K&C Initiative mosaics, in that the mosaics were created from the same raw data over the study area.

The processing chain is presented in Figure 6.1, and a detailed description follows.

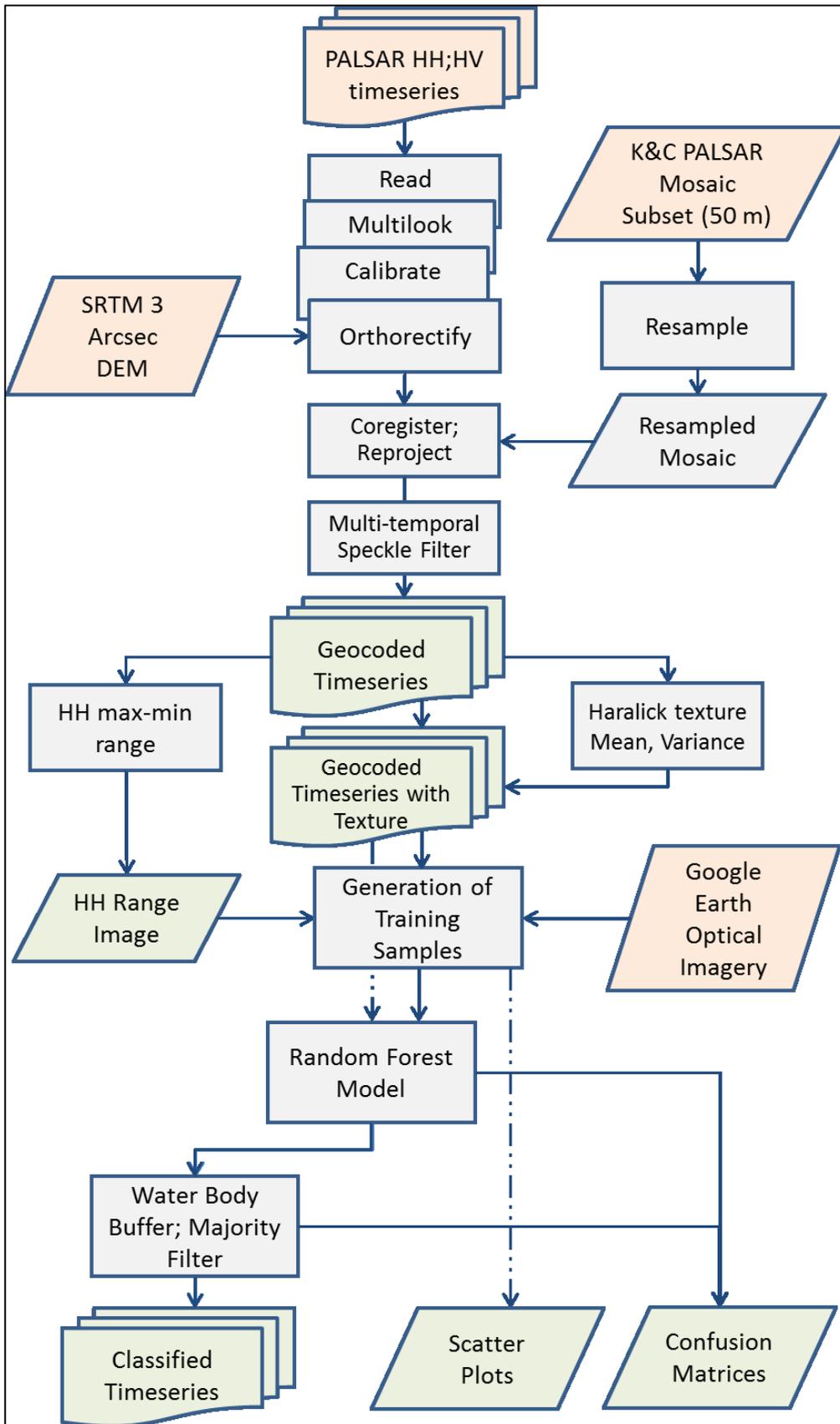


Figure 6.1: Flow chart of ALOS-PALSAR processing chain, including pre-processing of input data and training, classification and outputs. Several different multilook values were processed for all scenes.

To begin with, L1.1 ALOS-PALSAR fine beam dual polarisation (FBD) data were downloaded and processed according to standard protocol. Using the Sentinel 1 Toolbox (S1TBX) the L1.1, complex (containing phase and intensity information) data were read and then multilooked using four different multilook values resulting in pixel resolutions of approximately 15m, 30m, 45m and 60m, designated ML1, ML2, ML3 and ML4 respectively (indicating the number of looks in azimuth, not the equivalent number of looks). All images were calibrated using an absolute calibration constant correction from Lavalle & Wright (2009), which is specific to ALOS-PALSAR. They were then roughly orthorectified using the 3 arcsecond SRTM DEM and then finely geocoded by coregistering the data to an accurately geocoded Kyoto and Carbon Initiative mosaic subset of the study area (which had itself been resampled to the appropriate resolution depending on the multilook value used previously). During this stage, resampling was completed using bilinear interpolation.

One scene was discarded at this point (acquired on 2010-05-12) due to errors in the downloaded data, leaving eight images, two per year. There was one similar acquisition date each year, in June, and three years (2007, 2009 and 2010) had both images from the same months.

All remaining images were then processed to remove speckle (noise created from the random scattering within a resolution cell) using a multi-temporal speckle filter from the S1TBX, based on the algorithm of (Quegan et al., 2000), described in section 2.3.8.

All images were then reprojected into a Lambert Azimuthal Equal Area projection, using metres rather than degrees as units, and allowing for more consistent reporting of forest area. At this stage the images were cropped to their maximum intersection.

6.3.2 Generation and Selection of Derived Texture and Difference Layers

Eight Grey Level Co-occurrence Measures (GLCM) of texture from Haralick et al., (1973) were computed using the open source 'glcm' package in R, which allowed the mean of all texture directions to be used. These texture measures are commonly used to discern information from noisy data, and can improve classification accuracy (Hall-Beyer, 2017).

4-bit quantization levels were selected for the GLCM calculation (16 grey levels). Figure 6.2 shows a selection of class specific data value histograms produced from each of the 8 Haralick texture measures (histograms for all dates and polarisations are in Appendix B1). The mean and variance measures both exhibit different histograms for each class, and are sufficiently different to provide additional information. The homogeneity measure is very close to the variance measure, and the other 5 measures display too much similarity in 2 or more of the 4 classes and as such just mean and variance measures were used.

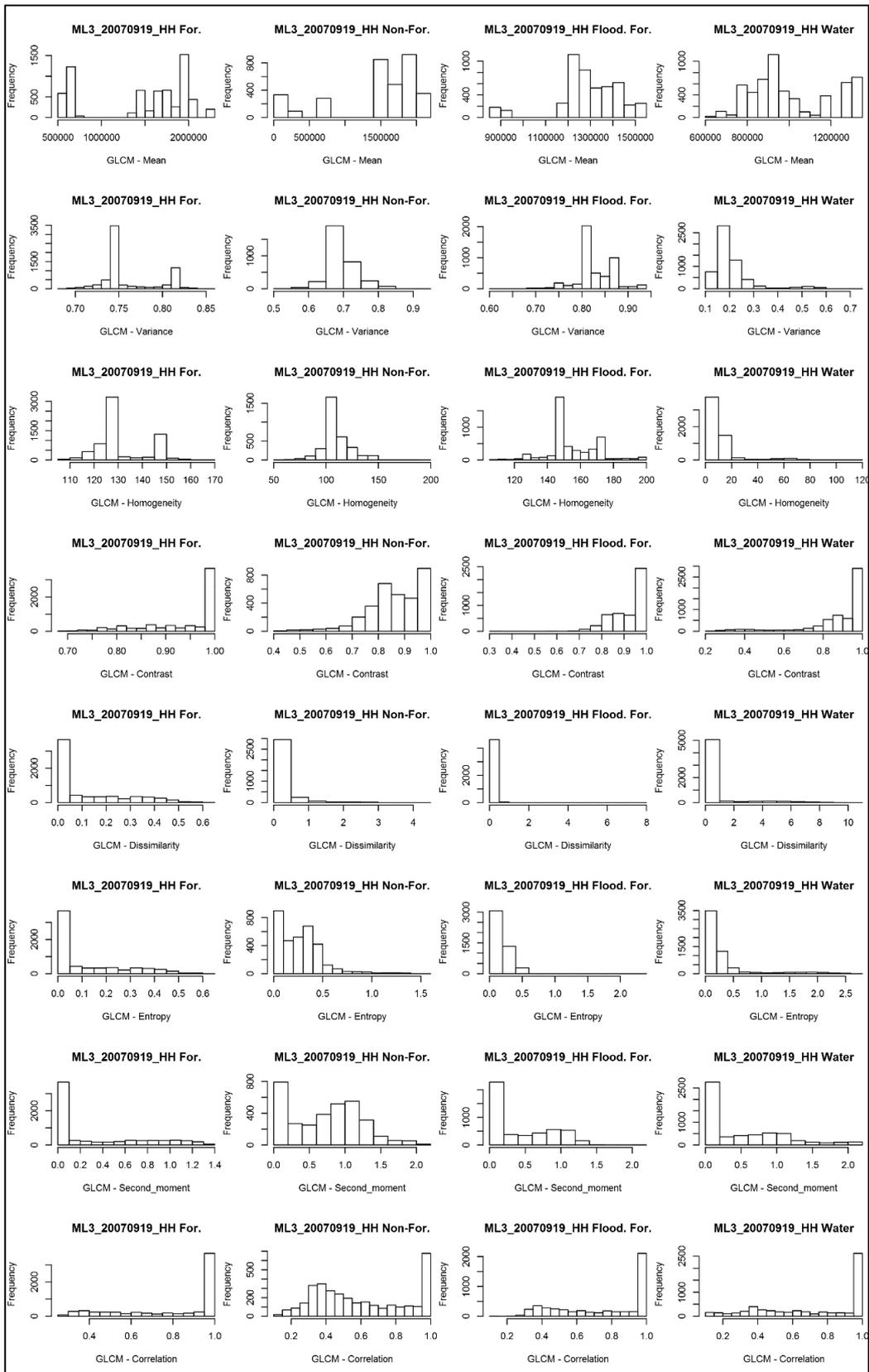


Figure 6.2: Histograms of the eight Haralick GLCM texture measures from Training and Validation sample areas; labels indicate the date, polarisation channel and class; frequency scales are different in each histogram to better illustrate range of values in more homogeneous texture measures. Texture generated from 45 m ALOS-PALSAR data; continued in Appendix B1

Both texture measures were computed for each of the two polarisations (HH and HV), and a difference layer (HH-HV) was also added.

For all eight images, the total range for HH backscatter was computed on a pixel level, to produce a single image of HH range from 2007-2010. The areas covered by water were masked out of this image, in order to only highlight areas of land with a high variability in HH backscatter.

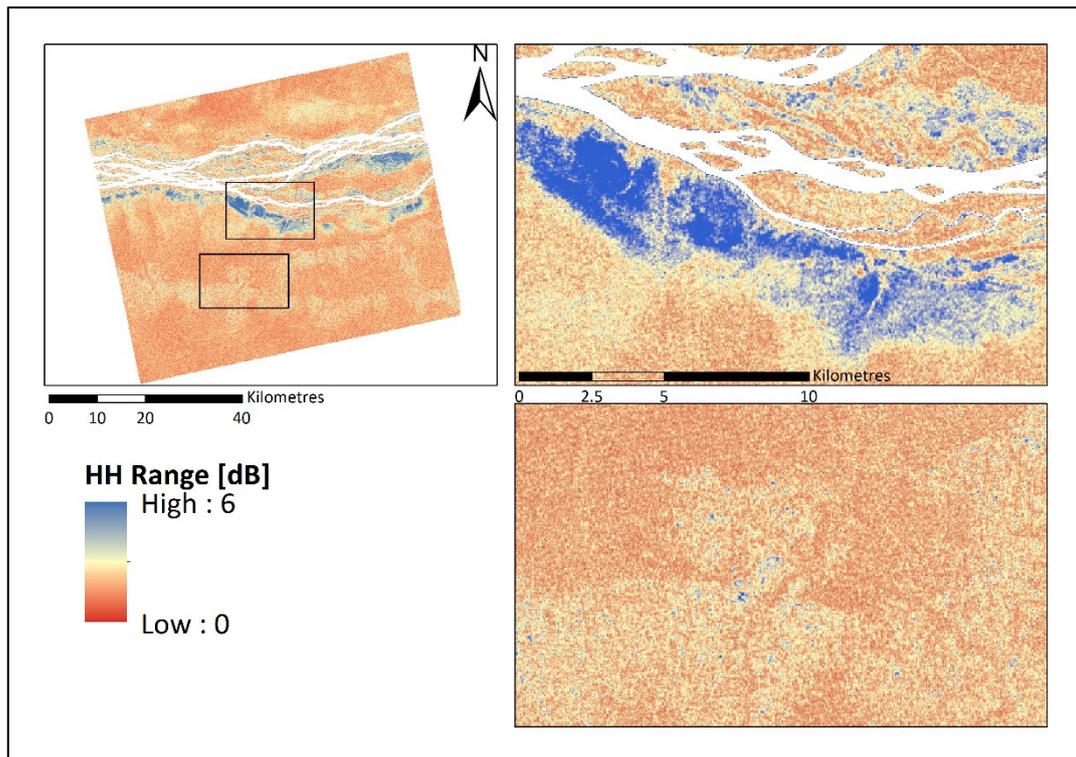


Figure 6.3: ALOS-PALSAR HH σ^0 eight-scene range image used to classify flooded forest areas; water bodies are masked and the data is stretched to illustrate the difference between flooded areas (top right) and non-forest areas (bottom right)

6.3.3 Generation of Reference Data

Training and validation samples were prepared ex situ for each year using a combination of high resolution optical imagery from google earth, ALOS-PALSAR RGB composites (HH;HV;HH-HV) and the HH-range image for all 8 images (Figure 6.3). An open source tool, from the “Semi-Automatic Classification Plugin” installed within Quantum GIS, provided a customisable region growing algorithm for generation of classified polygons. The region growing algorithm operated on individually selected pixels from the input

image, with several adjustable parameters including the maximum radius and tolerance thresholds (labelled as 'radius in the multispectral space'). After the selection of a clearly classifiable single pixel, a polygon forms around this 'seed' based on radiometric similarity. An area with a homogenous backscatter signal produced a larger polygon, up to the point that the maximum radius from the seed pixel was reached. More heterogeneous land cover types (as characterised by the radar backscatter) resulted in smaller, web-like polygons. Input layers to base the region growing algorithm on for each class were selected depending on the separability of the class by that layer. As such, the forest and non-forest classes used the HV channel and the flooded forest and water classes used the HH channel. The input parameters used were the following: maximum spectral distance from seed pixel to surrounding pixels – 0.2 dB; minimum area of region of interest (ROI) – 50 pixels; maximum width of ROI – 25 pixels. It should be noted that without the coordinates of the seed pixel, this sampling strategy, while a fast method of generating large reference datasets, is difficult to replicate.

In the absence of reliable ground truth data, the tool and method was chosen for its ability to combine the benefits of expert visual and contextual interpretation of multiple data sources with quick automatic generation of large numbers of neighbouring classified pixels. For each training polygon, the seed pixel was carefully chosen from a central area of the land cover type it represented, thus reducing the chance of a misclassification in edge areas. The resulting polygon was then visually assessed using high resolution imagery from google earth, the original speckle filtered ALOS-PALSAR imagery at its highest resolution (~15m), and in the case of the flooded-forest class, the HH range image. In the event that the polygon contained incorrect pixels, another seed pixel was selected. For the validation dataset, a similar procedure was followed. Figure 6.4 illustrates the spatial distribution of training and validation polygons generated, and the characteristic shapes of the polygons.

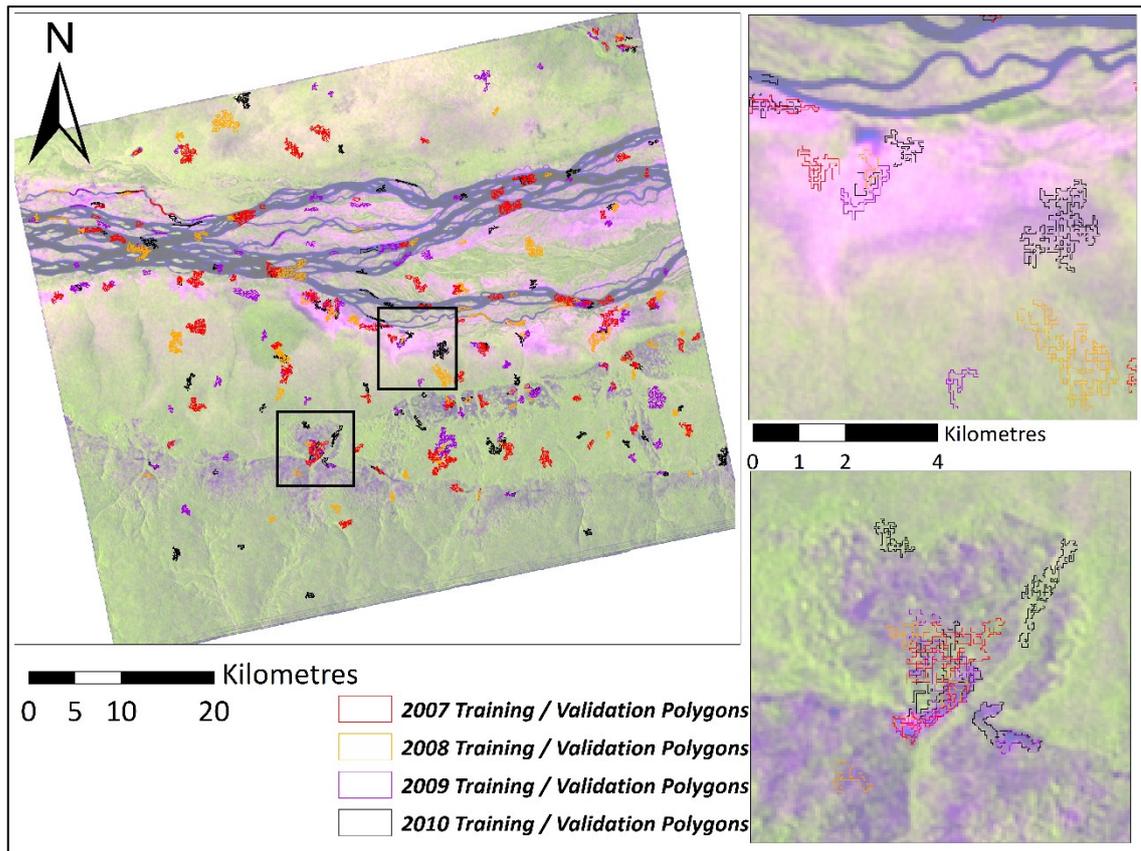


Figure 6.4: Spatial distribution of reference polygons (left), with zoom windows (right) indicated by black boxes; underlying image is an RGB composite of HH, HV and the polarisation difference from 2007-06-19

Training and validation polygons were grouped and assessed for spatial autocorrelation using the Moran's I tool (as in section 5.3.2) tool in ArcGIS. This returned Moran's I values of 0.5046, 0.3653, 0.6549 and 0.5308 for 2007, 2008, 2009 and 2010 datasets respectively. Z values were all above 2.58 (6.75, 6.54, 7.60, and 9.62 respectively), indicating that the likelihood of non-random clustering is statistically significant at the 99% confidence level. This may not reflect the actual variability in samples, but is less than ideal. During the model building step of the classifier, pixels in the input image stack which fell within any of the training/validation polygons for that year were split 50:50 into independent and separate training or validation samples.

6.3.4 Class Separability

Scatter plots were produced in HH and HV channels for each date, at each resolution, from the training data, in order to visually assess class separability. Matrices of transformed divergence in each of the 4 classes for HH and HV from the training dataset

were also produced to give an indication of class separability. Transformed divergence was calculated using the *separability* function in the *spatialEco* package in R according to the formula of Swain & King (1973). Values for transformed divergence between two classes range between 0 (perfectly inseparable) and 2 (perfectly separable), where values above 1.9 are considered to be well separated, and values below 1.7 are considered to be poorly separated (Regos et al., 2015; Jensen, 1996).

6.3.5 Classification Procedures

Two machine learning supervised classifications were compared, Random Forests (RF), and Support Vector Machine (SVM) (both described in section 2.4), to produce thematic maps with forest, non-forest, flooded forest, and water classes. To implement these algorithms, the open source Orfeo Toolbox (OTB) machine learning library was chosen and accessed through a batch script in the OSGeo4W (Open Source Geospatial [foundation] for Windows) environment. From section 2.4, RF was expected to perform better than SVM in terms of processing speed, as it is better suited to multiple classes and a large feature space. Following model training, the OTB image classifier routine was applied and an initial classification was produced for each scene, at each resolution and using both classification methods, to produce 64 classified images. As in Chapter 5, the SVM classifier was run using an RBF kernel, with automated optimisation for the γ parameter, using a k-fold cross validation on subsamples of the reference dataset, all after data were automatically normalised. The RF model parameters were tested on one date (2010-09-27) at all scales, to find the optimal classifier. Four of the five parameters from Table 2.4 were tested individually with other parameters at their defaults (tree depth – 5; minimum number of samples per node – 10; maximum number of trees in forest – 100; sufficient accuracy/OOB error – 0.01). Two values above and below the default thresholds were tested. The final RF classifier was tested on the dataset with just HH and HV polarisation (without mean and variance GLCM texture layers).

6.3.6 Post-Classification

A two-step post classification procedure was used to improve each classification, using contextual information (proximity to river features) and majority filtering to remove outliers, particularly at higher resolution. A water buffer was created to allow flooded forest pixels classified at more than 4km from the river to be reclassified as forest pixels. The second step was a majority filter (centre pixel takes the value of the mode in a moving window, remaining the same in the case of a tied value) using a moving window at a dimension appropriate to the resolution (windows of 7x7, 5x5, 5x5, and 3x3 for 15m, 30m, 45m, and 60m resolution respectively).

6.3.7 Scaling Thematic Classification

To check whether the effects of scale on the classification were different from a resampling of the classification from the highest resolution, a resampling of the classification at its highest resolution (15 m) to the coarser resolutions was undertaken. Resampling to a coarser resolution was undertaken using the gdal translate function, using the mode of the higher resolution pixels.

6.3.8 Validation

Confusion matrices were generated using the validation dataset for all classified images, both before and after the post-classification step, and from this an accuracy assessment was produced. To ensure fairness between scale levels, the number of pixels to validate at each class was restricted to 1000. However, since at the coarsest scale (60 m) there was a class (non-forest) with only 460 validation samples, the maximum number of validation samples for all classes in this scale was restricted to 460.

The area covered by each class was extracted and analysed for each scale, from the maximum intersecting area of the 8 scenes covered.

6.3.9 Basin Excess Rainfall Maps / Normalised Cumulative Rainfall

For the study area, the HYDRO1K subbasin catchment and daily RFE 2.0 data were used to generate local basin catchment measures of dekadal cumulative rainfall for each day during the ALOS-PALSAR lifecycle. While Basin Excess Rainfall Maps (BERM) for Africa

are available online (USGS, 2015a) through the Famine Early Warning Systems Network portal (FEWSNET), the data are difficult to analyse (they are published as scaled down .png map documents) and are based on dekadal cumulative rainfall sums that do not correspond to the ALOS-PALSAR acquisition dates. The FEWSNET BERMs are also produced using HYDRO1K sub-basin catchment and daily RFE 2.0 data, and have an additional river reach product. Their product scored sub-basin catchment rainfall based on long-term average conditions and also used upstream catchment rainfall in the calculation of the river reach product.

Given the scope of this project, and the intended use of the products, normalised cumulative rainfall (NCR) values were generated instead of BERMs, using a simpler method. Within each sub-basin catchment, the dekadal cumulative rainfall sum R was calculated for each day of the year, and the NCR values on each date d were generated as the dekadal sum feature (R_d) scaled to a range of 0-100 based on the minimum (R_{min}) and maximum rainfall (R_{max}) sums for the time period of the ALOS-PALSAR lifecycle (2006-02-02 to 2011-06-30) as shown in Equation 11 below.

$$NCR_d = 100 * \frac{R_d - R_{min}}{R_{max} - R_{min}}$$

Equation 11

The normalising step was undertaken since the sub-basin catchment areas are of varying shapes and sizes. Using values from the sub-basin catchments that contribute to the study area, the NCR values corresponding to the dates of ALOS-PALSAR scene acquisition for this study were then generated. The study location and an example of the NCR values from 2008-05-06 are presented in Figure 6.5.

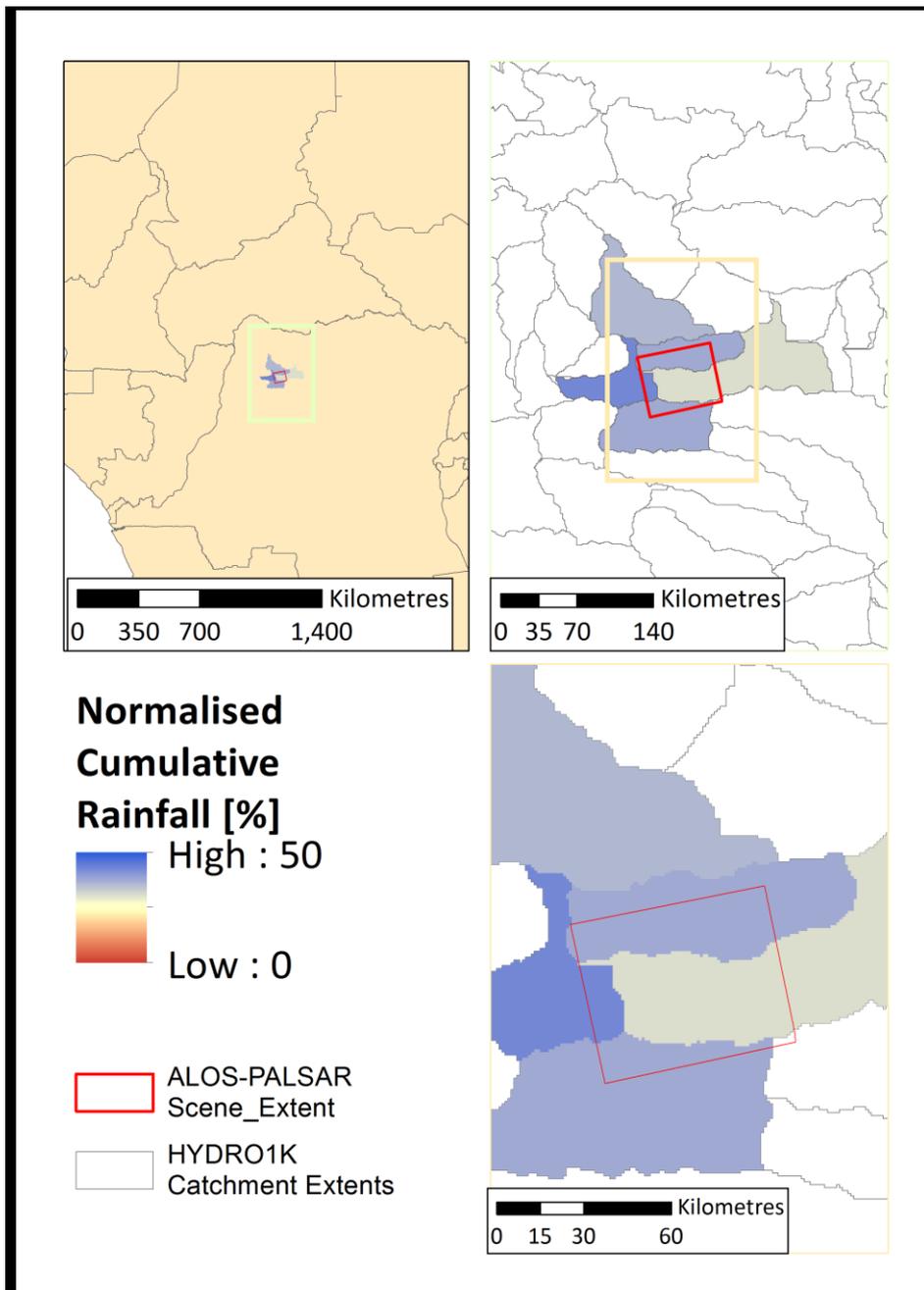


Figure 6.5: Study area scene extent, HYDRO1K basin catchment areas (right images), and example NCR values (from 2008-05-06)

Figure 6.6 shows how the mean (of the five contributing sub-basins) NCR values compare with those from the entire period against which they are normalised. Ideally, a greater range of NCR values would be desirable, but there was sufficient variability in the available eight points to compare relatively wetter and drier scenes.

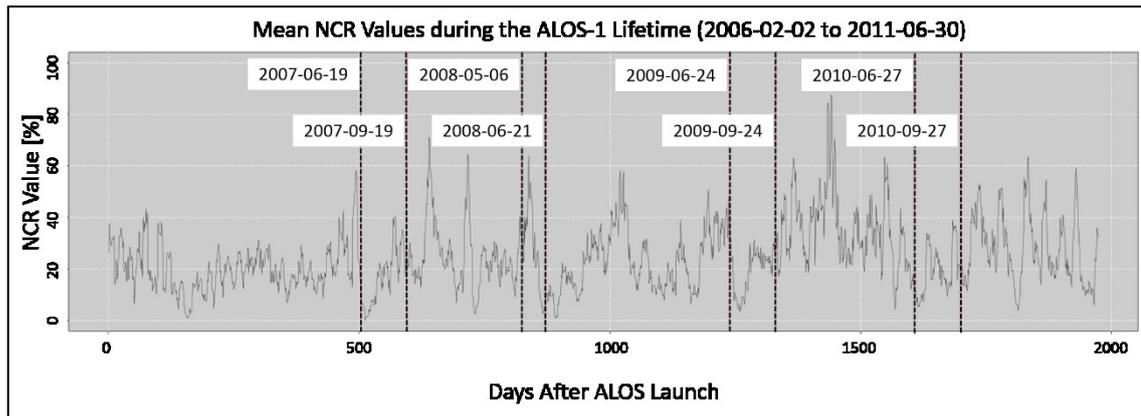


Figure 6.6: Graph of mean NCR values from the ALOS-1 lifetime (from launch to mission failure); vertical dotted lines correspond to image acquisition dates from this study (labelled)

6.3.10 Use of AMSR-E SM and VW

AMSR-E Soil Moisture (SM) and Vegetation Water content (VW) data were downloaded and processed using R. There are ascending and descending passes, with a repeat pass for each orbital direction at the equator of 2 days. Therefore to obtain daily coverage in what is already an extremely coarse resolution dataset (in comparison with ALOS-PALSAR data), with values masked where vegetation is dense or water bodies are extensive, the mean of ascending and descending values were combined. The values were cropped to the study area, comprising a small number of pixels, and compared with the backscatter values in HH and HV, as well as with the RFE 2.0 data and the derived NCR values. 3-day averages of SM and VW were also generated and tested with RFE 2.0 and RFE 2.0 derived data, and the relationships between AMSR-E metrics and RFE 2.0 products were examined for all data available (including days not within the ALOS PALSAR acquisition coverage) to increase the sample size due to the coarse resolution.

6.3.11 Software Used

This chapter used the ESA Sentinel 1 Toolbox (version 1.1.1) for pre-processing; for application of the SVM and RF classifiers, Quantum GIS (version 2.8.2) and the Orfeo Toolbox (OTB) were used within the Open Source Geospatial [foundation] for Windows (OSGeo4W) environment. R (version 2.15) with packages *rgdal* and *raster* were used to

process RFE 2.0 and AMSR-E data and create NCR figures with the HYDRO1K catchment area shapefiles.

6.4 Results

Results from this chapter are divided into nine sections: first, class separability and secondly statistics for the processed input ALOS-PALSAR imagery are reviewed and analysed; third, a brief comparison between two classification methods; fourth, the texture measures are tested to make sure they are a necessary addition to the classifier; fifth, those dealing with the differences in forest measurement and classification at multiple resolutions; finally a comparison of flooded forest area with NCR values.

6.4.1 Class Separability

To examine class separability, and to highlight the importance of the multi-temporal speckle filter used during processing, scatter plots of HH and HV ALOS-PALSAR channels were produced using the classified training dataset. Figure 6.7 shows the diminishing number of representative pixels contained within the training datasets at decreasing spatial resolution. While this is an expected result, given the use of the same training data polygons at all resolutions, class separability is shown to improve with decreasing resolution.

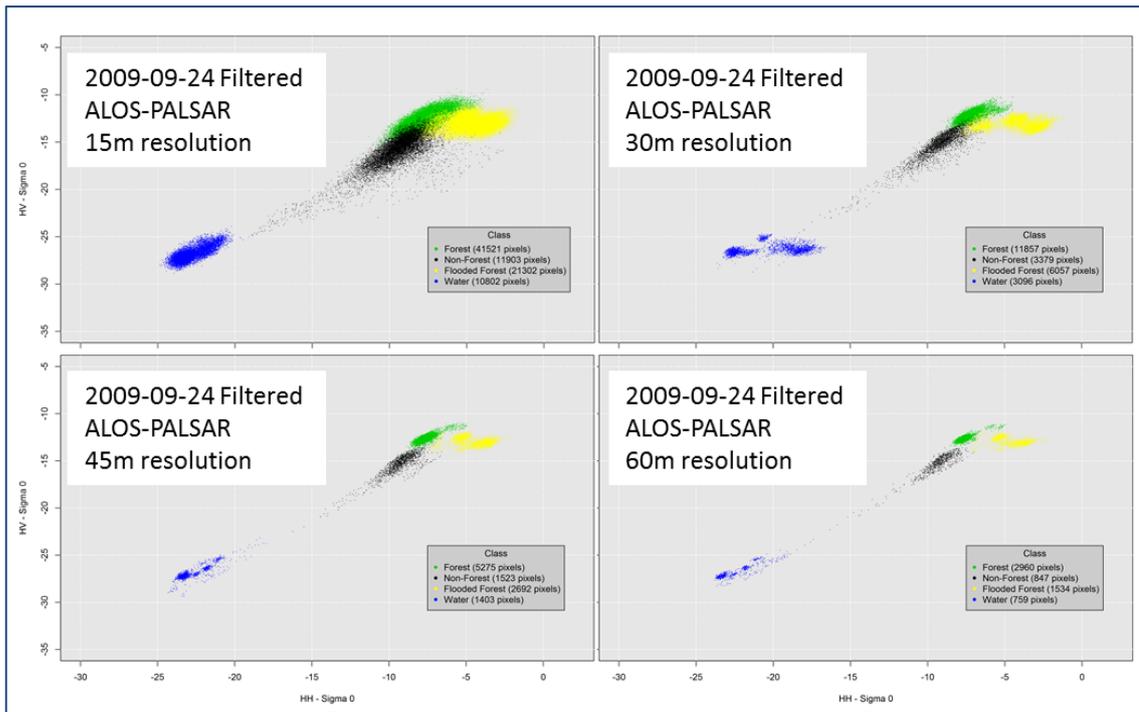


Figure 6.7: Filtered ALOS-PALSAR HV/HH classified scatterplot comparison at different spatial resolutions (15, 30, 45 and 60 m)

Figure 6.8 shows a comparison between unfiltered and speckle filtered ALOS-PALSAR classified scatter plots. In this case, with the same training polygons and at the same spatial resolution, the number of pixels is identical, but the class separability appears improved in the speckle filtered dataset. All scatterplots, both filtered and unfiltered, from all dates and at all four spatial resolutions, are available in Appendix A1.1. Similar trends of improved class separability for filtered, coarse resolution data are observed across the entire dataset.

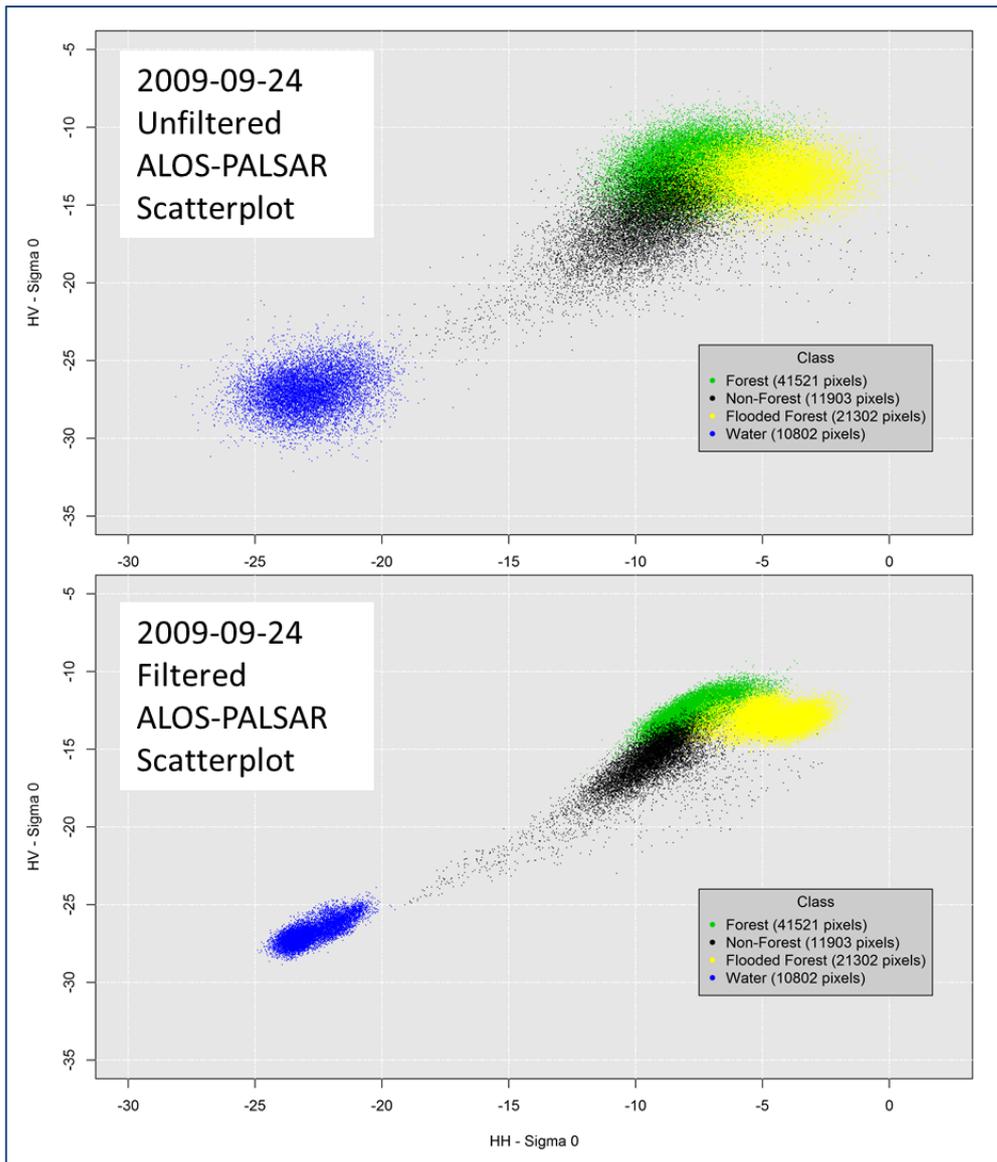


Figure 6.8: Filtered/Unfiltered ALOS-PALSAR HV/HH classified scatterplot comparison at 15m spatial resolution

Measures of class separability in the form of transformed divergence (TD) in HH and HV polarisation channels are seen in Figure 6.9 (and at all scales in Appendix B3). It is clear that in individual polarisation channels, class separability is low (except for water, which is highly separable from the other classes). The Non-Forest/Flooded Forest TD is greater than other non-water classes in HH, and the Forest/Non-Forest TD is greater in HV. Forest/Flooded-Forest has consistently well below optimal (<1.7) TD in HV. Both 2007 datasets showed low TD in all classes at this scale, although there is improvement at coarser resolutions, particularly in the Forest/Non-Forest TD in the HV channel.

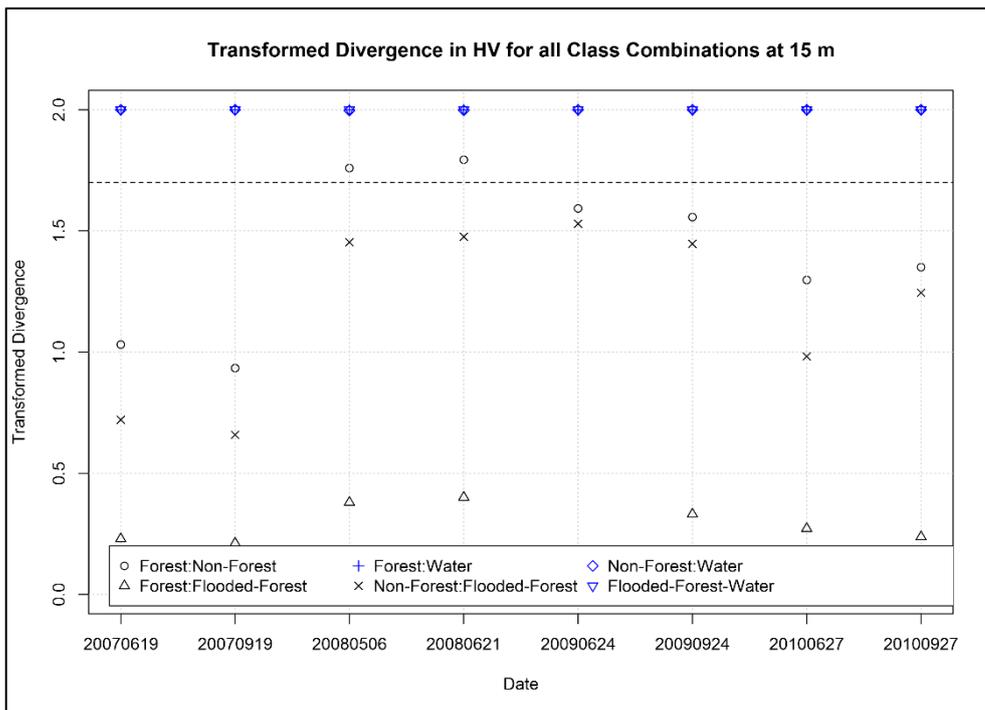
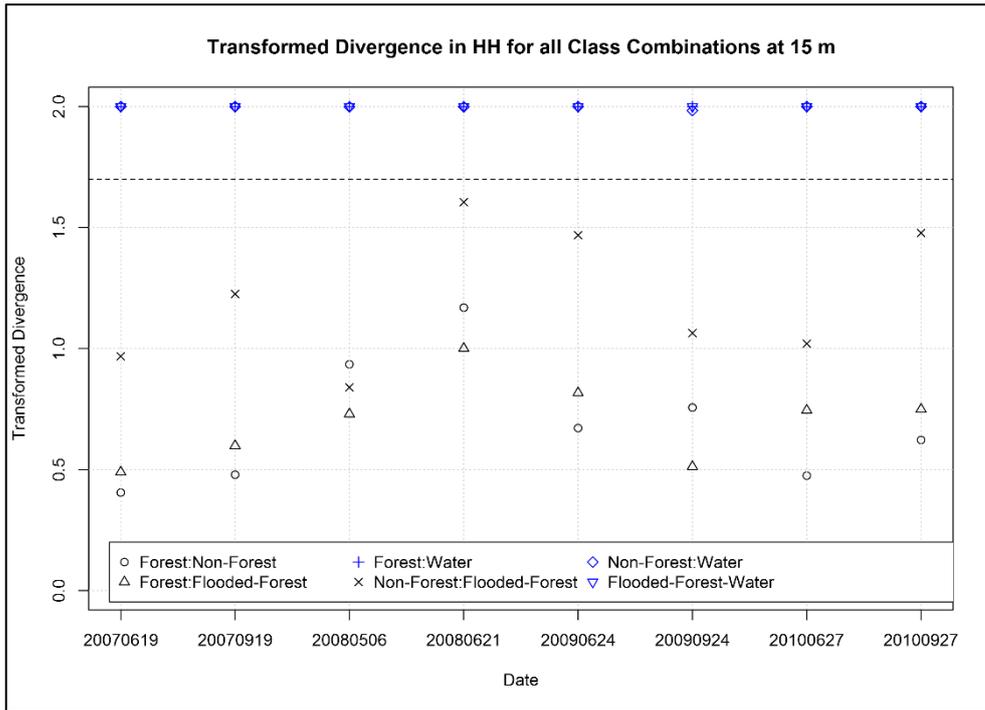


Figure 6.9: Inter class transformed divergence at 15 m in HH (top) and HV (bottom) channels; horizontal dashed line indicates 1.7 threshold of transformed divergence; other scales in Appendix B3

6.4.2 RF and SVM Classification Comparison

The RF parameter search did not yield much information to optimise the classifier above the default parameters. Figure 6.10 shows that the parameter with the greatest

influence was maximum tree depth. Lower than default values had a severe effect on overall accuracy, and higher values had a small effect in the scene analysed. Other parameters either had negligible or unpredictable effects (whether the parameter was increased or decreased) on the scene tested.

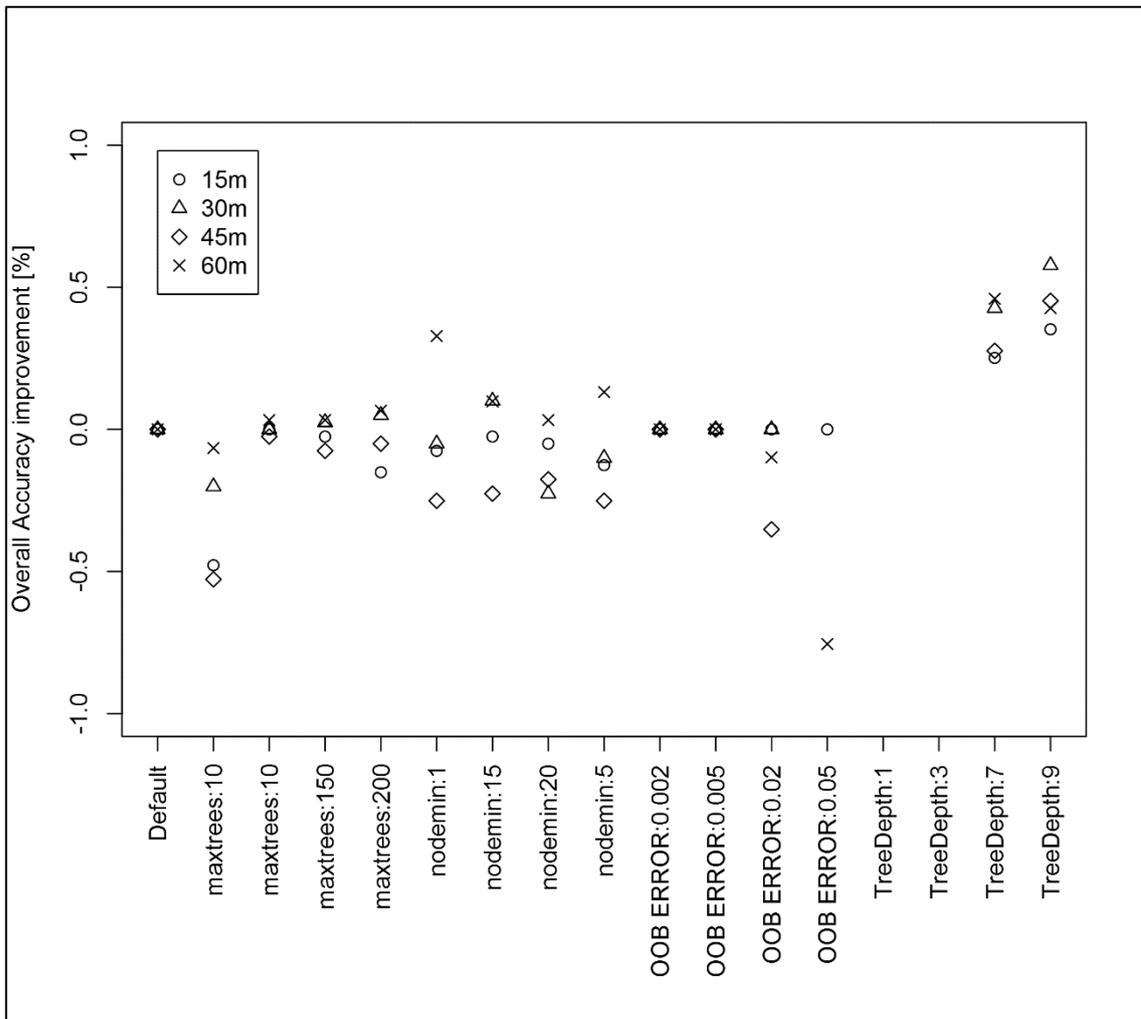


Figure 6.10: Random Forests parameter search results; parameter and value is shown on the x-axis, and the y-axis shows the difference between the overall accuracy and that of the default parameters; all four scales were tested on one scene acquisition (2010-09-27). Values for Tree Depth:1 and Tree Depth:3 were ~-50% and -4% respectively at all scales, and therefore do not appear on this chart

The two supervised computer learning classification methods were compared in terms of visual output, overall accuracy, and processing speed. The major difference between the two methods is processing speed, seen in Figure 6.11. RF processing speed is unaffected by the number of pixels per class, at around 6 seconds (rising to 9 seconds if all pixels are used at 15m spatial resolution). SVM Processing speed is directly linked to the number of training pixels used by each class, which causes limitations on the number

of training pixels that may be used by the SVM classifier. It should be noted that while parameters for both methods were optimised to a degree, there are parameters in RF that could cause an increase in processing time, in particular the maximum number of trees, which were not examined by this study as changes in this parameter offered no improvement in the parameter search. At coarser resolutions, where fewer pixels are covered by the training polygons, the number of training pixels may not need to be limited. Visually, there are no major differences between the two methods if a limit of 1000 training pixels is used (Figure 6.12), and Figure 6.13 shows that there are no significant differences between their total classification accuracies (based on an accuracy assessment from the validation dataset).

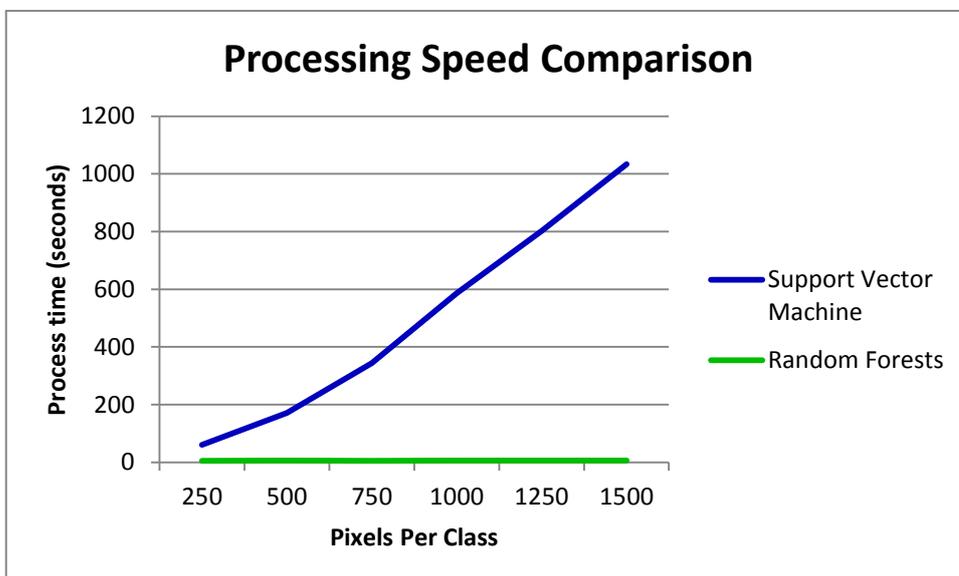


Figure 6.11: Processing speed comparison between SVM and RF for a single image classification at 15m spatial resolution

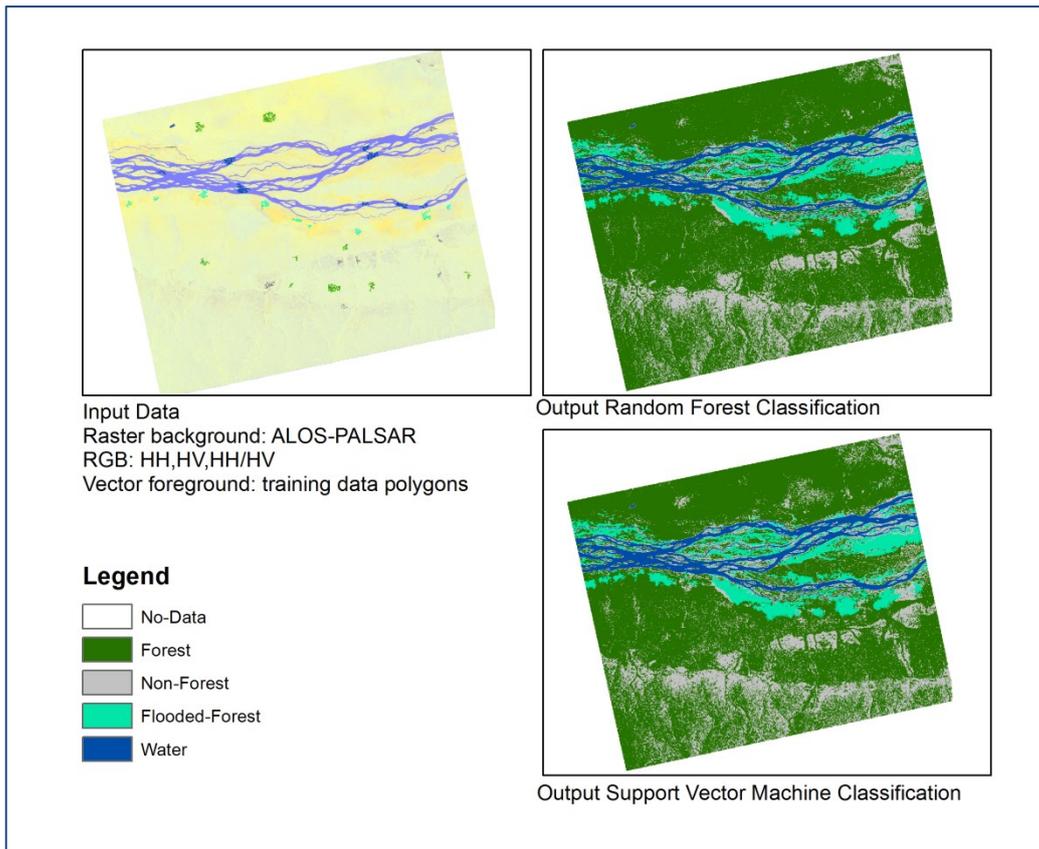


Figure 6.12: RF vs SVM classification, using similar input data and a limit of 1000 training pixels per class

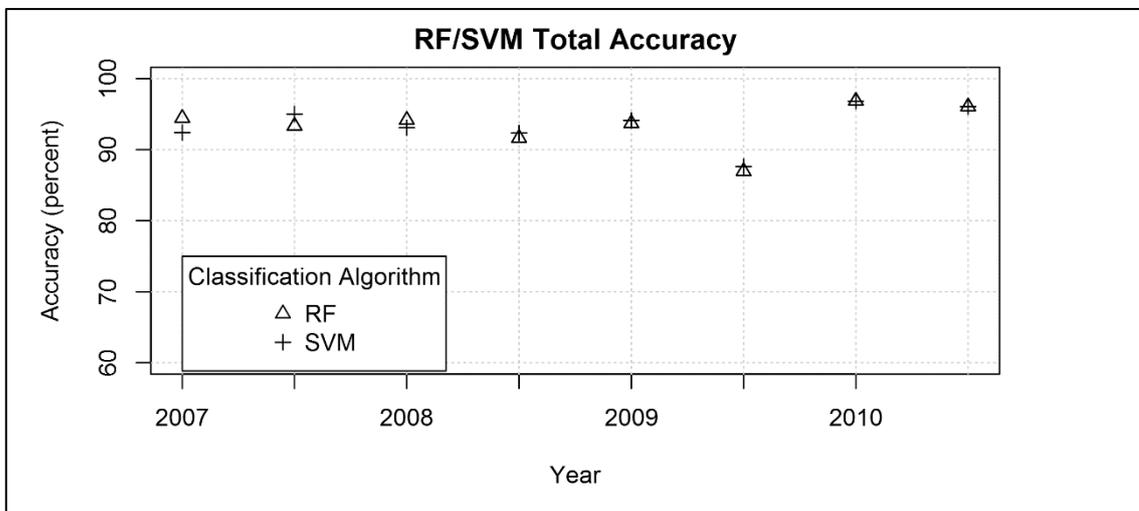


Figure 6.13: Overall accuracy assessment comparison between SVM and RF classifiers

6.4.3 Texture Information

A comparison of accuracy with and without GLCM mean and variance HH and HV texture layers was performed at all scales and dates and demonstrated the utility of the texture measure (Figure 6.14). Texture was found to improve the overall classification accuracy in all scenes and at all scales, but particularly so at the higher resolution.

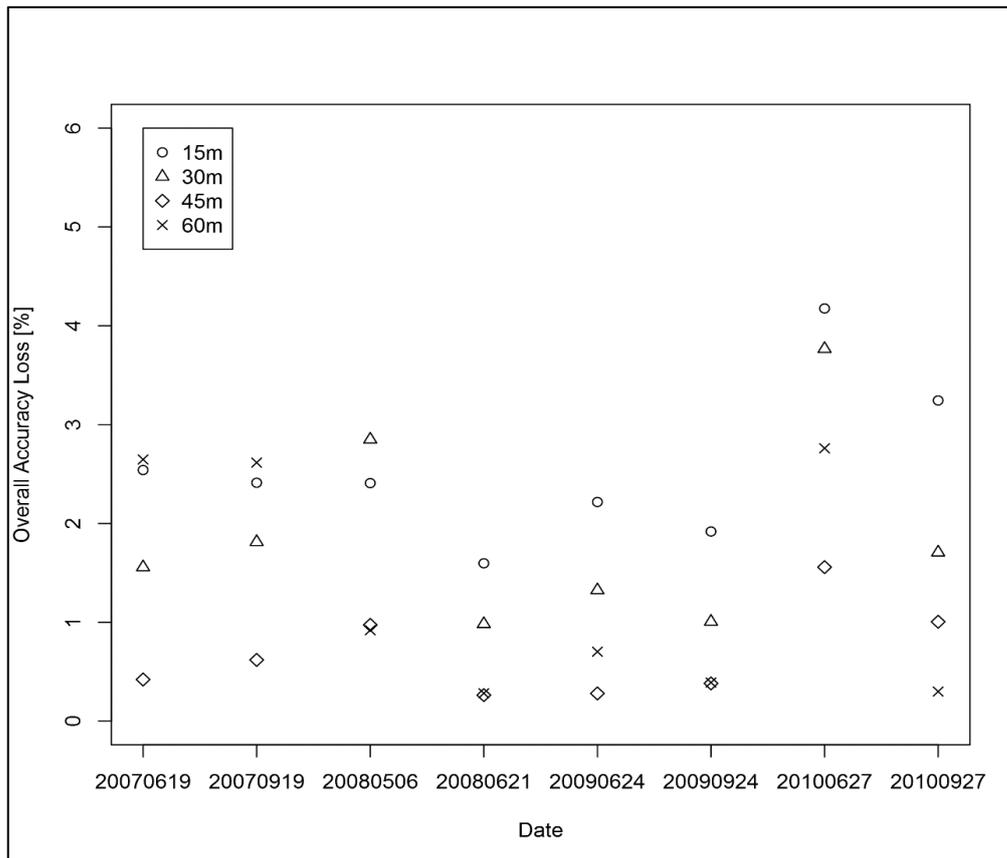


Figure 6.14: Loss of accuracy from removing mean and variance texture layers at HH and HV polarisations with scene date on the x axis and four scales represented by different symbols

6.4.4 Classification at Different Scales

The following results are based on multi-temporal speckle filtered data, classified using the RF algorithm. A comparison of four resolutions, across eight dates, is described below. In most of the eight scenes there are clear differences visible looking at the entire scene; an example is given below from the 2007-09-19 scene (Figure 6.15), and highlighted in zoomed in areas of flooded forest and non-forest (Figure 6.16 and Figure

6.17, respectively). Appendix B4.2 contains figures for all eight scenes, including details of flooded forest and non-forest. Comparing the classification with the RGB composites below in the detail zoom figures illustrates the smoothing effect of the multilooking process. At coarser resolutions the classification becomes regularised, potentially leading to the removal of erroneous labelling due to signal noise, but also overlooking smaller forest clearances within forested areas, and smaller isolated clusters of woody vegetation in largely unforested areas.

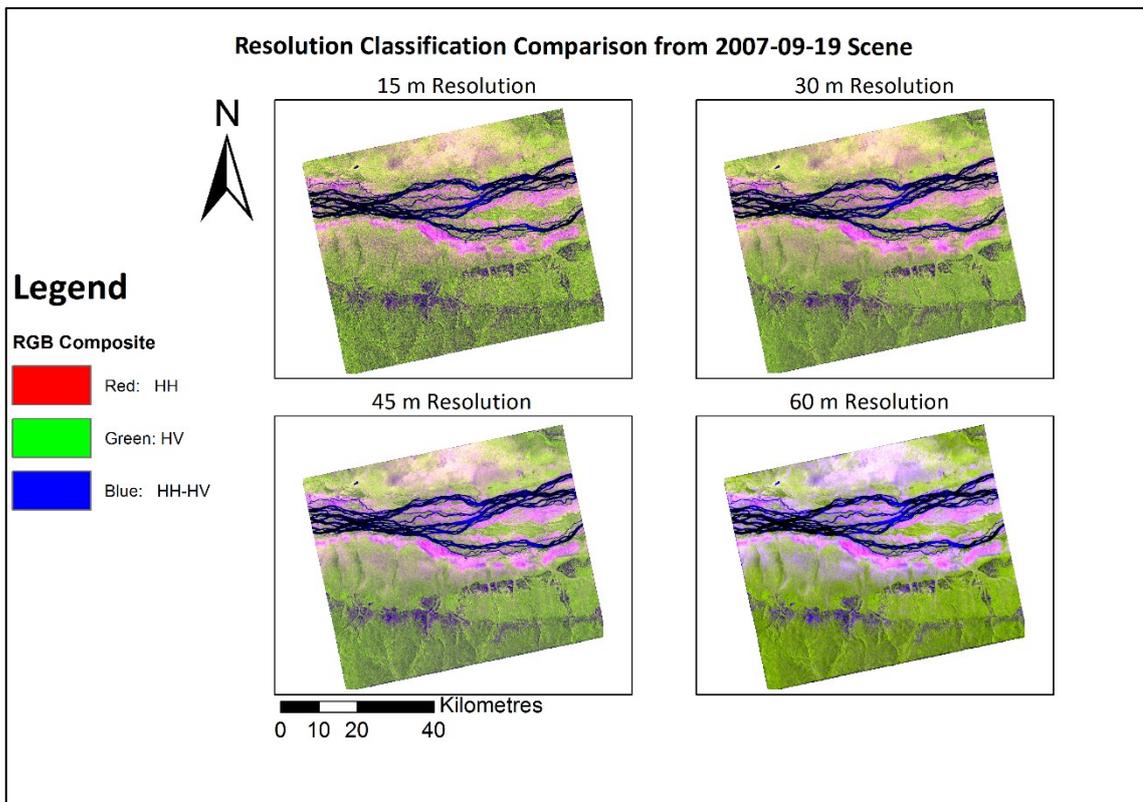
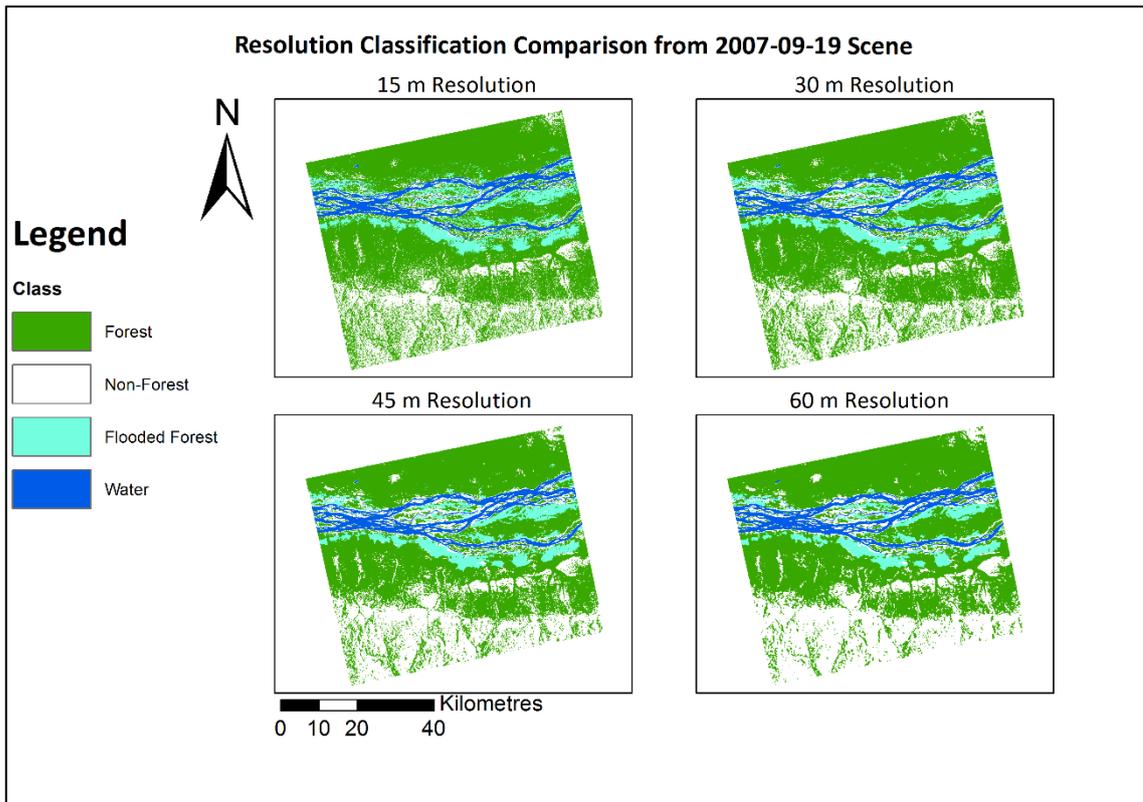


Figure 6.15: Comparison of classifications (top) at 15, 30, 45 and 60m resolutions with RGB composite of [HH:HV:HH-HV] stretched identically regardless of scale (below); full scene extent

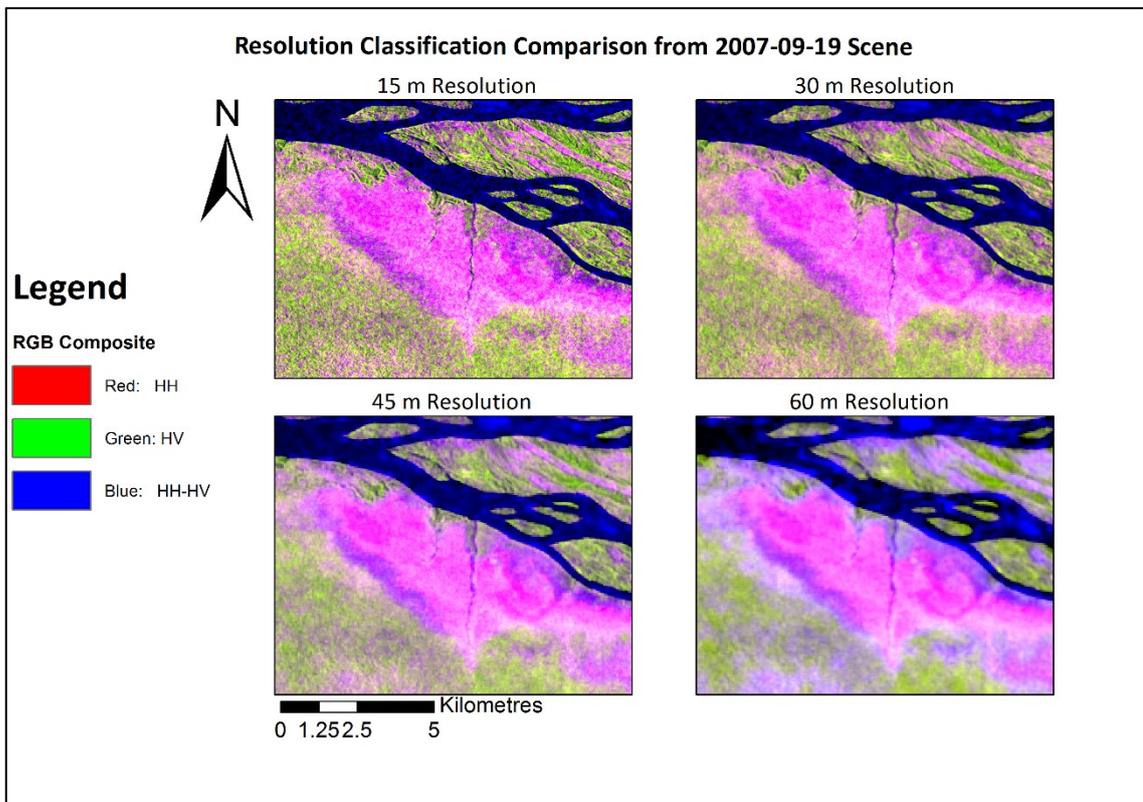
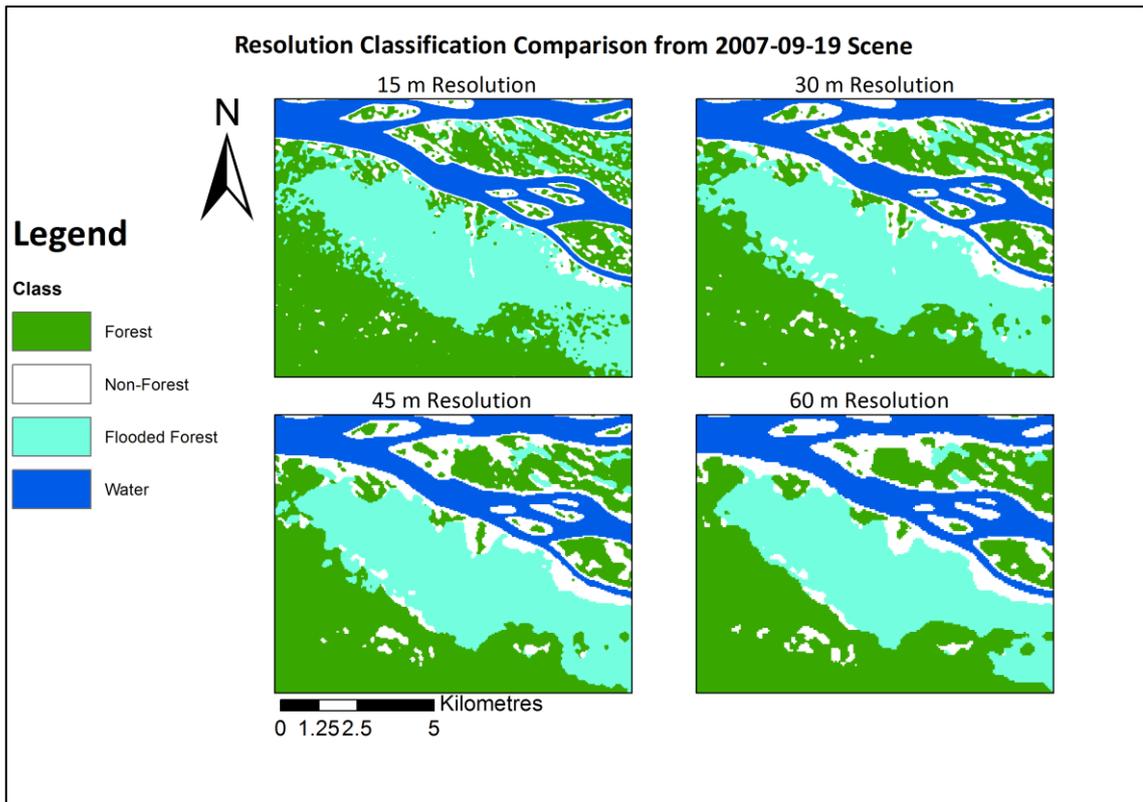


Figure 6.16: Comparison of classifications (top); flooded forest detail with RGB composite (below)

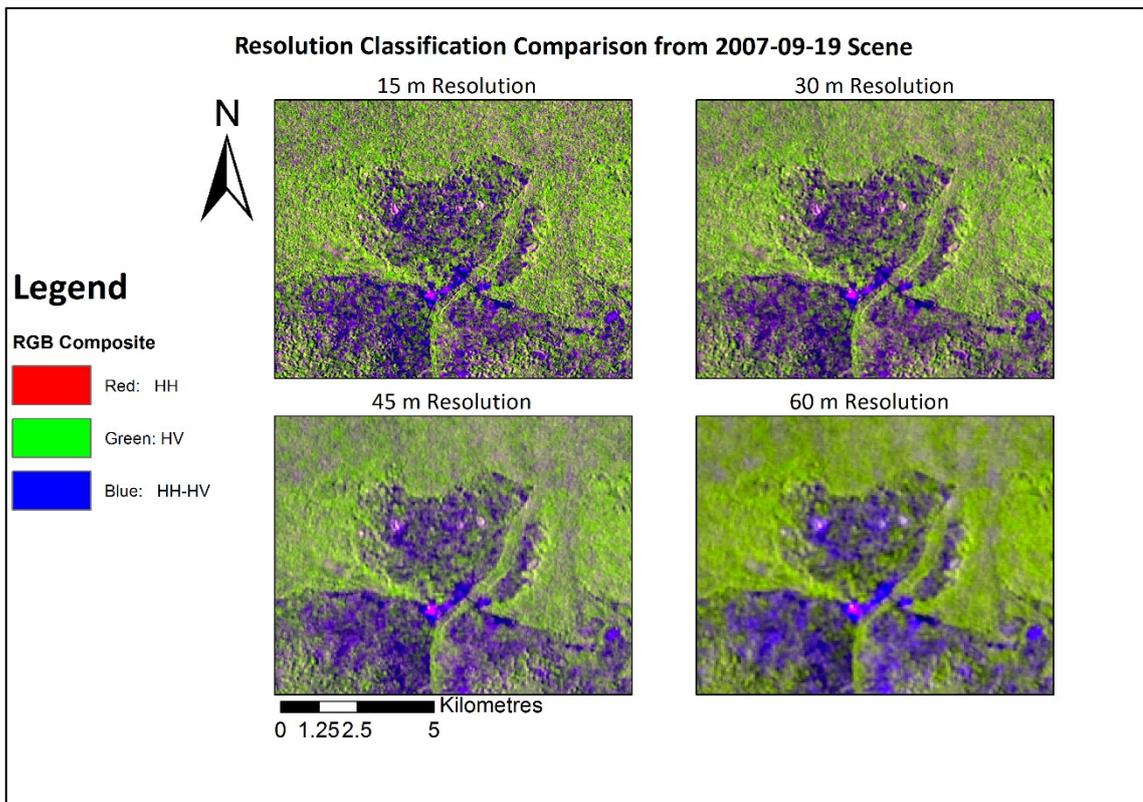
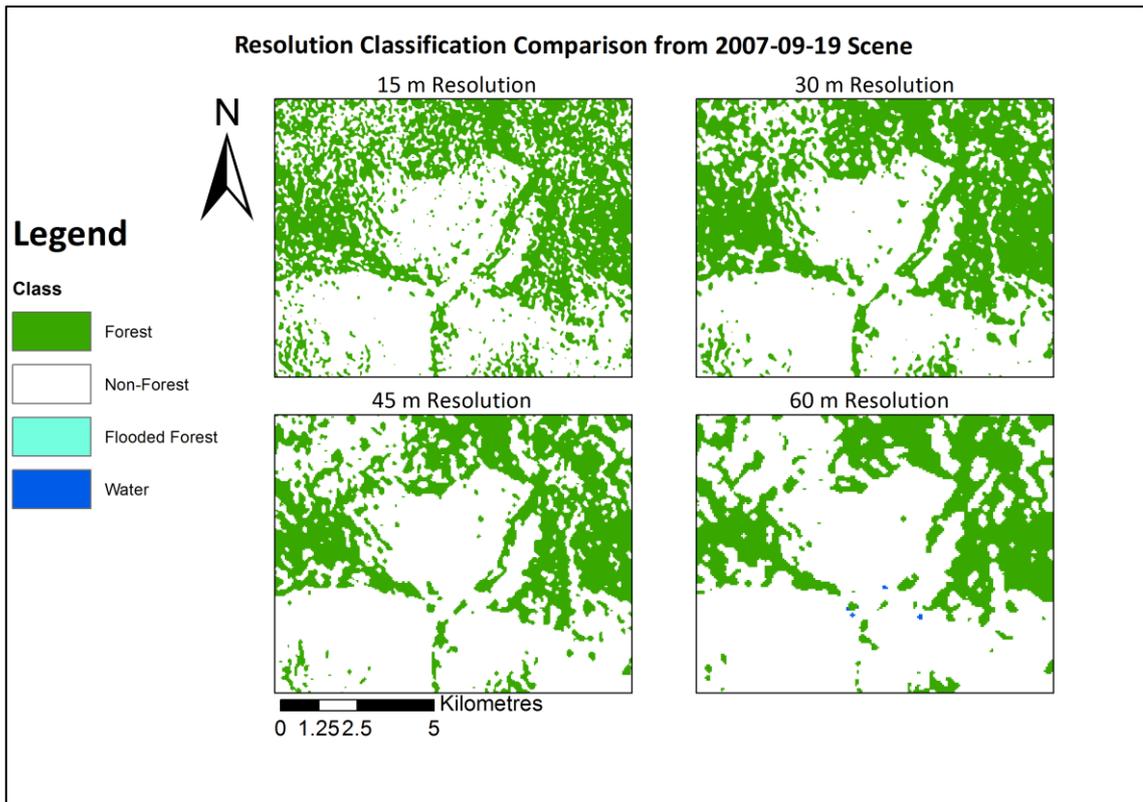


Figure 6.17: Comparison of classifications (top); non-forest detail with RGB composite (below)

Accuracy assessments show there is generally higher total accuracy (a score of all correctly classified pixels) at higher spatial resolution, seen below in Figure 6.18.

However, the performance of the classifier at each resolution is not consistently higher at finer spatial resolutions when individual classes are assessed (assessing errors of commission and omission for that class)⁴. Figure 6.18 shows fluctuating responses for the forest and flooded forest classes, with no clear dominance by any of the tested resolutions. Only the non-forest class shows a consistent advantage of higher resolution. This may be due to a combination of higher backscatter variability in the non-forest class (which covers several land cover types: settlements, crops, bare earth), a dominance of forest cover in the study area, and non-forest ground features occurring at a scale lower than that observable at coarser resolution.

The accuracies of flooded forest are low compared with the forest class, and this is due to a single validation dataset being produced for each year, as SAR backscatter signals for seasonally flooded forest are more likely to change between scenes than for any other class in the study area. Errors for flooded forest are characterised by flooded forest/forest class confusion, particularly at higher spatial resolution, although the number of scenes with greater flooded forest/non-forest class confusion increases with lower resolution images. Misclassification of flooded forest as forest rather than non-forest is anticipated since flooded forest is a subset of forest, and a positive outcome given the aims to correct misclassification of flooded forest as non-forest in the K&C Initiative forest/non-forest products.

It is clear, visibly from Figure 6.15, and in the graphs of Figure 6.18 that the classification of forest and non-forest in the first scene (from 2007-06-19) at 45 m resolution has made considerable errors. This is possibly attributable to a data calibration issue, as well as the selection of training polygons for those classes. The scatterplots (in Appendix A1.1) and low transformed divergence between classes for both 2007 scenes show poor

⁴ All pixels correctly classified as 'x' divided by all pixels that were classified as 'x' plus all pixels that should have been classified as 'x' (but were not).

class separability for forest and non-forest compared with the other years. Full confusion matrices are in Appendix AI.3.

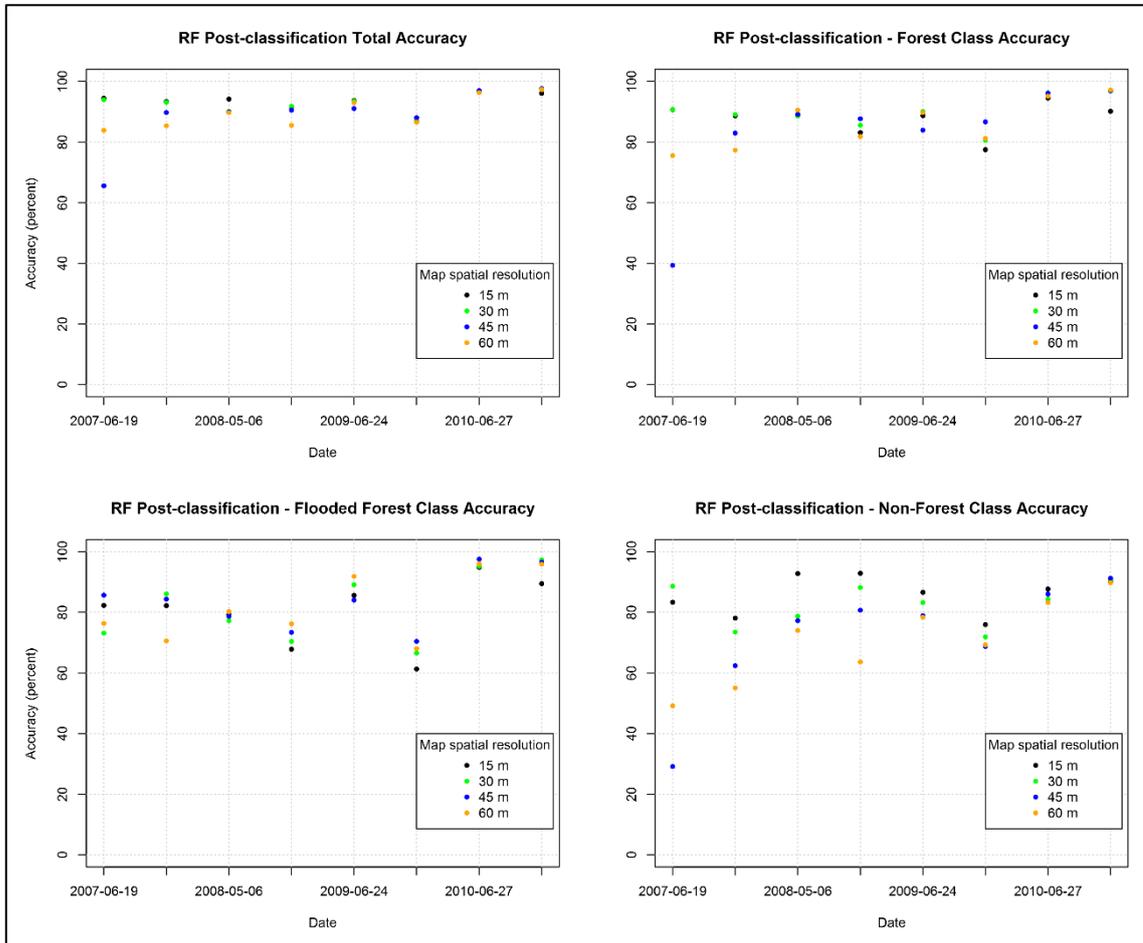


Figure 6.18: Comparison of accuracy assessments for RF classifications at multiple scales - total accuracy as well as accuracies of three classes – forest, non-forest and flooded forest

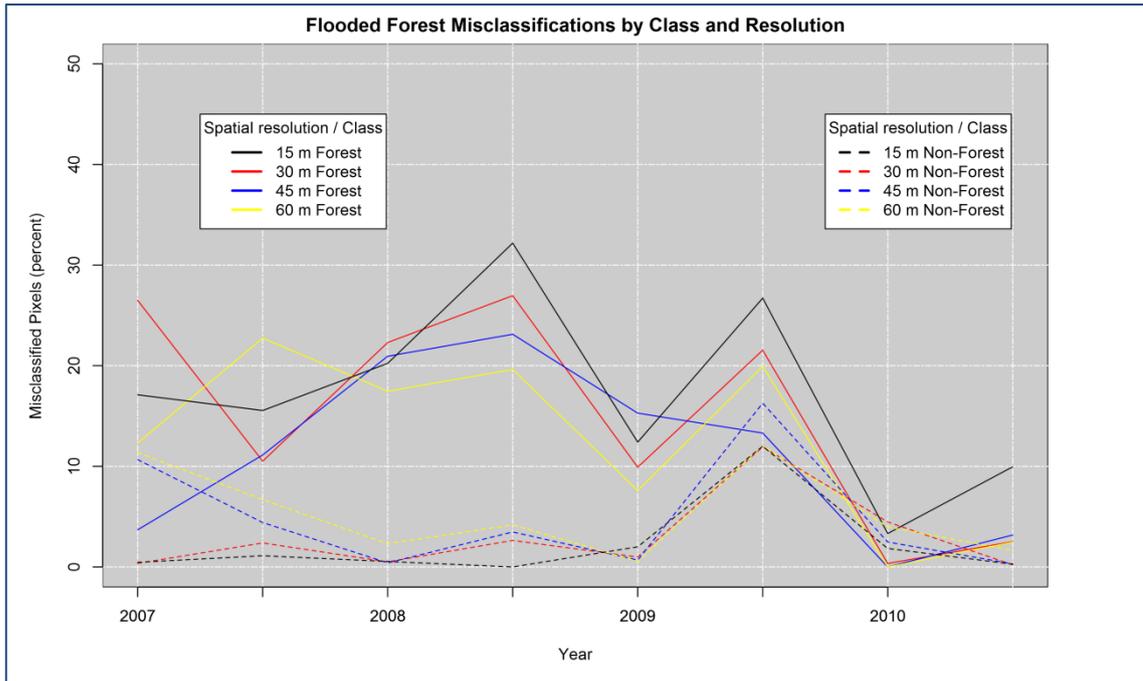


Figure 6.19: Flooded forest class confusion: forest and non-forest comparison; misclassified pixels expressed as a percentage of total number of pixels classified (correctly or incorrectly) as flooded forest

Table 6.2: Class Accuracy Statistics

15 m	Forest	Non-Forest	Flooded Forest	Water	Total
Range	0.1701	0.0638	0.3356	0.0715	0.0990
Standard Deviation	0.0521	0.1690	0.1102	0.0246	0.0306
Mean	0.8782	0.8594	0.8034	0.9610	0.9340
Coefficient of Variation	5.94	7.42	13.71	2.56	3.28
30 m	Forest	Non-Forest	Flooded Forest	Water	Total
Range	0.1632	0.1891	0.3075	0.1950	0.1068
Standard Deviation	0.0530	0.0707	0.1165	0.0680	0.0342
Mean	0.8970	0.8239	0.8187	0.9148	0.9296
Coefficient of Variation	5.91	8.58	14.22	7.44	3.68
45 m	Forest	Non-Forest	Flooded Forest	Water	Total
Range	0.5755	0.6207	0.2710	0.2352	0.3202
Standard Deviation	0.1830	0.1947	0.0976	0.0823	0.0996
Mean	0.8285	0.7180	0.8382	0.9015	0.8868
Coefficient of Variation	22.09	27.12	11.65	9.12	11.23
60 m	Forest	Non-Forest	Flooded Forest	Water	Total
Range	0.2166	0.4059	0.2799	0.2314	0.1346
Standard Deviation	0.0817	0.1388	0.1122	0.0989	0.0525
Mean	0.8606	0.7031	0.8189	0.8737	0.8971
Coefficient of Variation	9.50	19.74	13.70	11.32	5.85

6.4.5 Comparison to Thematic Maps Resampling

Thematic maps were resampled from the highest resolution to the coarsest, post classification, and the results compared both visually and statistically in terms of the differences in reported class areas. The following three figures show the visual results (see Figures 6.17-6.19 for comparison with ALOS-PALSAR RGB composites). The 15 m products are identical as it is the source of the post-classification resampling in each case. There is a marked difference between both methods visible in the study area at the full scene level but more noticeably in the close-up figures. Post classification resampling to a coarser resolution appears to have minimal effect in this study area, and indeed this is borne out by the reported class areas, where there is almost no difference in the reported area in any of the scales, while there is much more variation in the multilooked (pre-classification) reported areas (Figure 6.18).

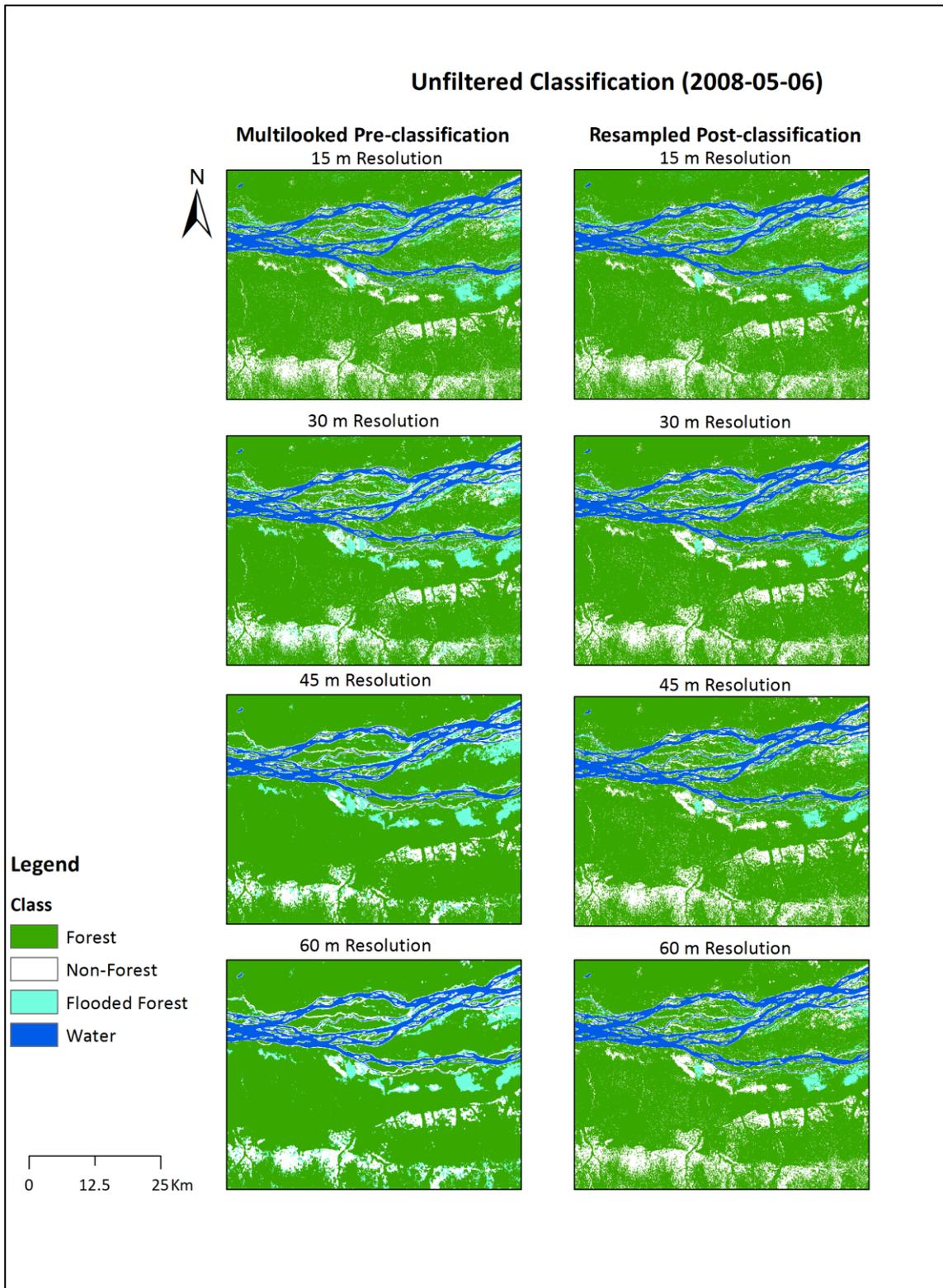


Figure 6.20: Class comparison between pre- (multilooking) and post-classification (thematic resampling) scale adjustments; full scene extent

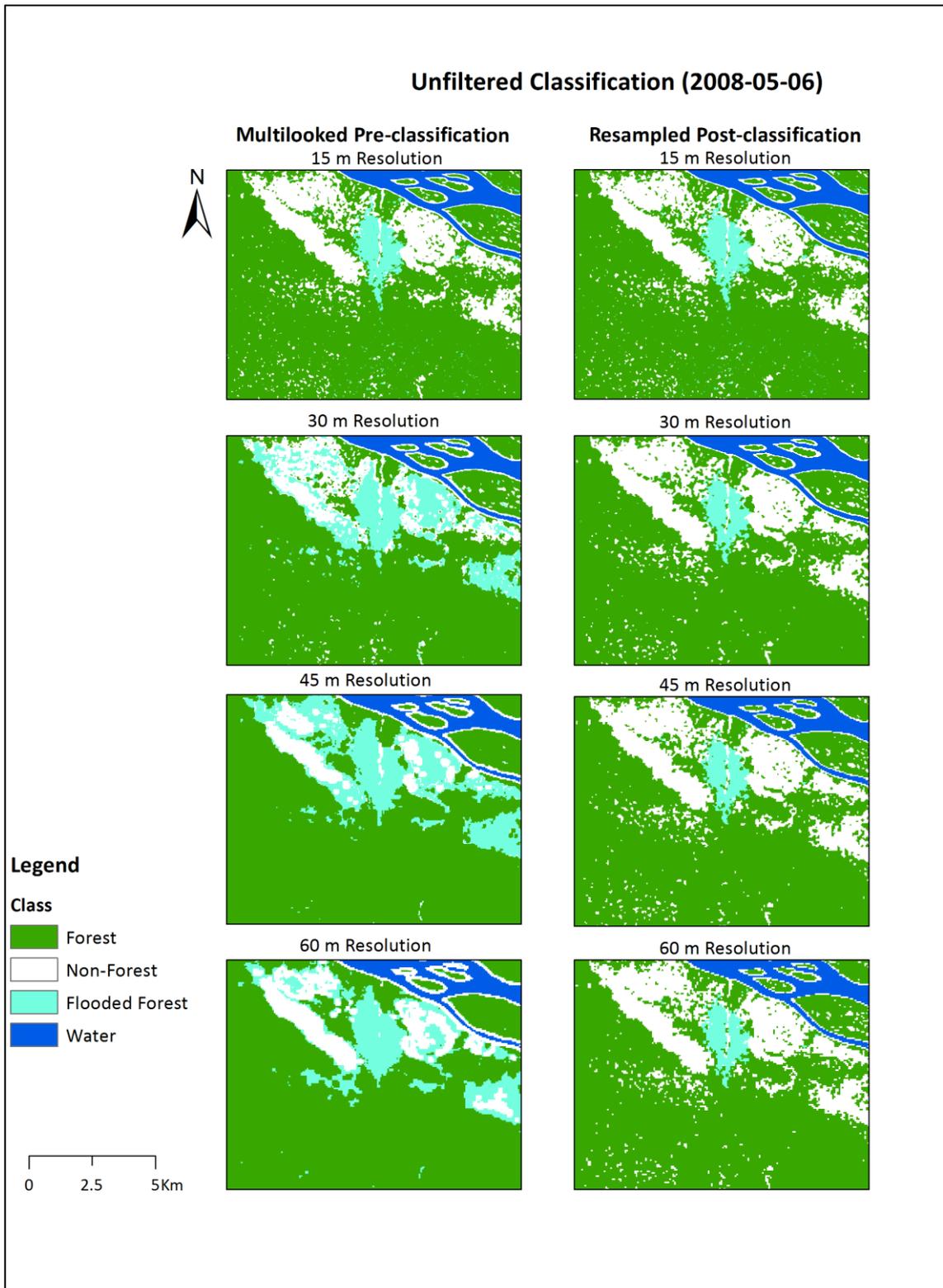


Figure 6.21: Class comparison between pre- (multilooking) and post-classification (thematic resampling) scale adjustments; flooded forest detail

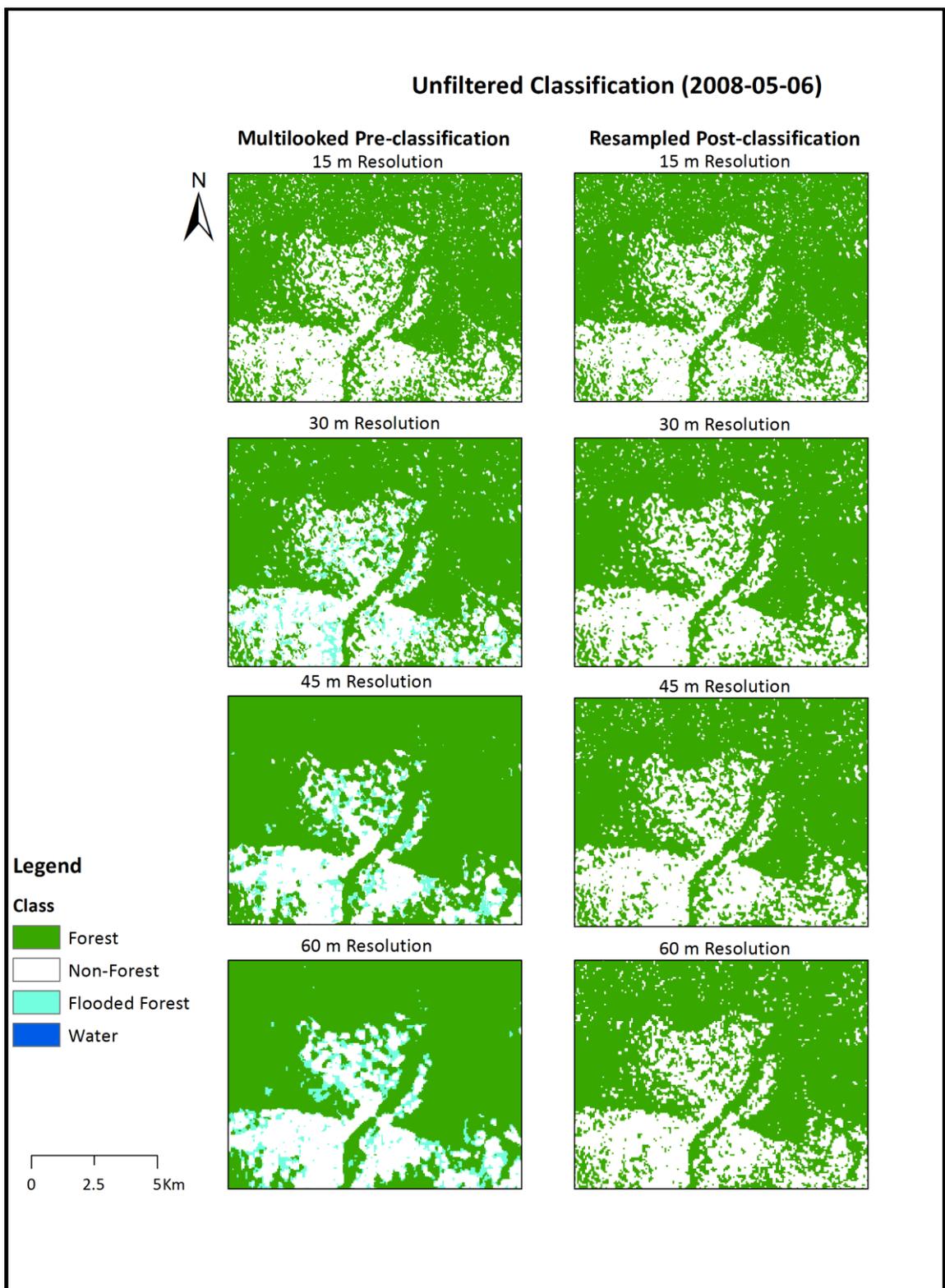


Figure 6.22: Class comparison between pre- (multilooking) and post-classification (thematic resampling) scale adjustments; non-forest detail (flooded forest appears in this data as the water body buffer has not been applied)

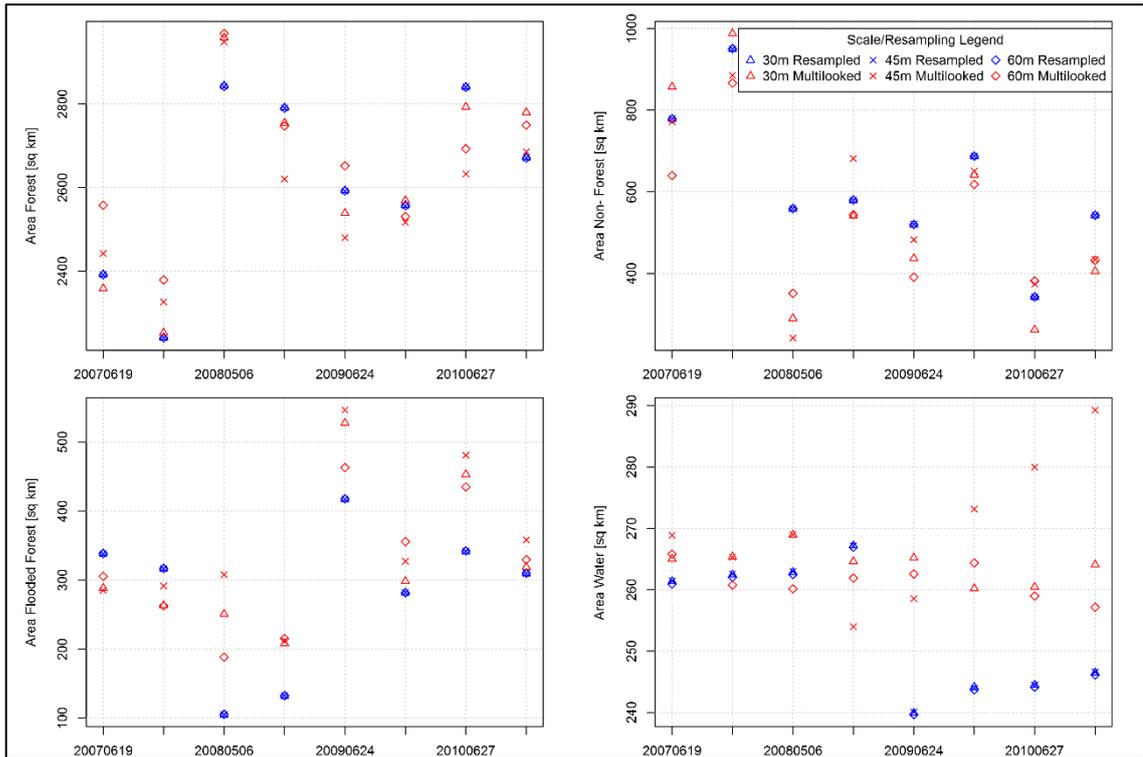


Figure 6.23: Comparison between class areas in km^2 reported by thematic map resampling and pre-classification multilooking

6.4.6 Class Area Differences

Forest, flooded forest and non-forest area were computed at each resolution, as well as a 'combined forest' area of forest and flooded forest classes, and are displayed in Figure 6.24 below. Stable values for non-forest are assumed to be more reliable, as neither considerable deforestation nor afforestation has occurred in the period covered by the imagery. Due to seasonality and the influence of rainfall on backscatter in the flooded forest class, more variability is expected, although the aim of the study is to map areas of forest that regularly flood, not just areas that are flooded at the exact time of the image acquisition. When the forest and flooded forest classes are combined, the coefficient of variation (standard deviation as a percent of the mean) in reported area decreases compared with either individual class, which appears to confirm the misclassification results in Figure 6.19.

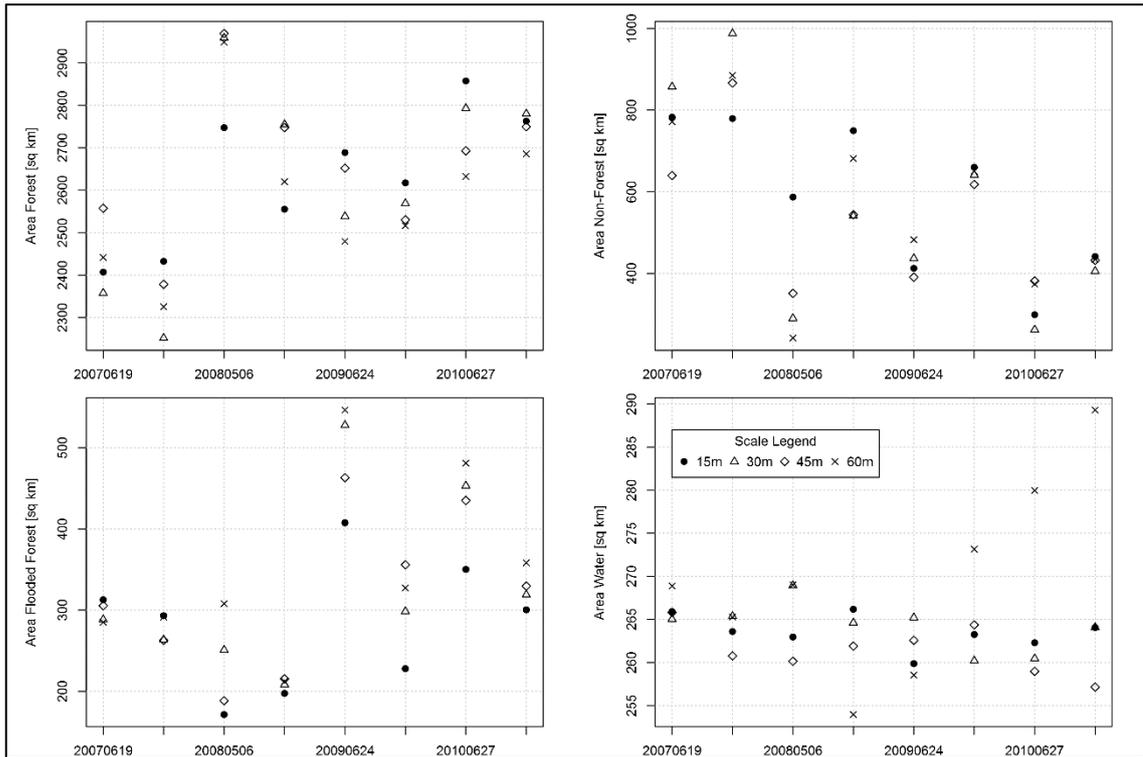


Figure 6.24: Comparison of reported class areas at different scales (clockwise from top left: Forest, Non-Forest, Flooded Forest, Water)

Across forest, non-forest and flooded forest, there is greater stability at higher resolutions (15 and 30 m), and more erratic reporting of area at lower resolutions (45 and 60 m). The range and standard deviation of reported area (Figure 6.25) show this more clearly, particularly for the forest and non-forest classes. Flooded forest area has lower range and standard deviation, because it covers a smaller area than forest and non-forest, although its coefficient of variation at all resolutions is more comparable to non-forest. The water class is included for comparison; it is separable from the other classes (Figure 6.7) and not likely to change considerably.

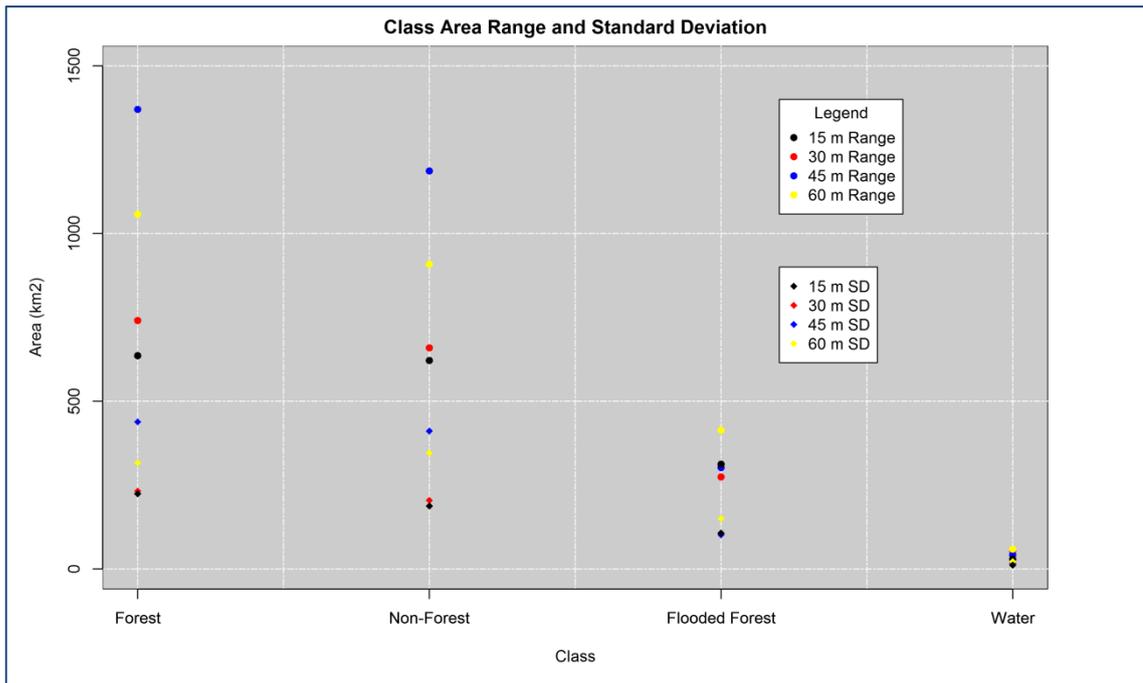


Figure 6.25: Area range and standard deviation at all spatial resolutions across the eight classified images

Table 6.3: Class area statistics from RF classification for the 8 classified images at each resolution

15 m	Forest	Non-Forest	Flooded Forest	Water	Combined Forest
Range	635.92	621.45	312.20	28.87	639.85
Standard Deviation	224.14	187.49	106.83	11.31	192.82
Mean	2668.11	575.73	271.25	253.26	2939.36
Coefficient of Variation	8.40	32.57	39.38	4.47	6.56
30 m	Forest	Non-Forest	Flooded Forest	Water	Combined Forest
Range	740.58	659.17	274.12	52.99	682.18
Standard Deviation	232.19	204.37	102.66	18.61	210.44
Mean	2679.40	556.64	281.79	250.50	2961.19
Coefficient of Variation	8.67	36.72	36.43	7.43	7.11
45 m	Forest	Non-Forest	Flooded Forest	Water	Combined Forest
Range	1370.15	1186.49	302.21	41.39	1215.94
Standard Deviation	438.41	411.06	101.74	15.54	425.20
Mean	2528.94	709.08	285.33	244.88	2814.26
Coefficient of Variation	17.34	57.97	35.66	6.34	15.11
60 m	Forest	Non-Forest	Flooded Forest	Water	Combined Forest
Range	1057.36	908.56	413.40	58.94	964.21
Standard Deviation	315.67	345.86	150.17	21.65	366.13
Mean	2529.82	672.73	307.59	256.97	2837.41
Coefficient of Variation	12.48	51.41	48.82	8.42	12.90

The K&C Initiative product and this study's RF classifications (the four scenes which correspond to the dates used in the four annual mosaic products at this location) are shown in Figure 6.29, and are directly comparable as they are both derived from the same original data. All four resolutions of the RF classifications were used, and the K&C Initiative forest/non-forest product is presented at both resolutions (25 and 50 m). Immediately noticeable are the misclassifications between forest/non-forest in the K&C Initiative product in the southern half of the scenes compared with this study's RF classifications. These are consistent misclassifications though, and are likely to be due to the application of a wider regional threshold for forest in the algorithm employed by JAXA.

There are no clear differences in resolution between the 25 and 50 m versions of the K&C Initiative product in the study area. This is likely the result of a dominance of the forest class in this area (based on the K&C Initiative classification).

6.4.7 NCR Results

Figure 6.26 shows a side by side comparison of the ALOS-PALSAR HH polarisation backscatter with the ten-day NCR scale described in section 6.3.9 (a value for each basin catchment normalised as a percentage using the maximum and minimum values for ten-day cumulative rainfall in that basin catchment between 2006-02-02 and 2011-06-30). Processing was carried out over five basin catchment areas that contribute closely to the area covered by the eight ALOS-PALSAR scenes acquired. Due to the higher sensitivity of HH polarisation to below-canopy moisture, it was compared rather than HV.

Two dates stand out as having comparatively high rainfall, 2008-05-06 and particularly 2009-06-24. Looking at the detailed area of interest close to the river on the right hand side of the figure, which have been stretched to show areas of strong HH backscatter over -9 dB in blue, there is a reduction in HH backscatter during rainy periods compared with the other drier scenes. The 8-scene range of HH values was used during training of the flooded forest class, in expectation of this HH disparity between drier and wetter scenes. Since a higher HH backscatter appears to characterise areas of forest that are

seasonally inundated, yet is indicative of relative dryness, the apparent reductions of flooded forest areas on the two dates in question seen in Figure 6.27 were predictable. Indeed, the peaks of NCR correspond to drops in the scene mean HH backscatter, evidenced by Figure 6.28. The misclassification of flooded forest as non-forest is a less expected result, particularly in the 2008-05-06 scene. This was more apparent in the K&C Initiative forest/non-forest map products in Figure 6.29, and was a reason for this study.

From the data available, and from the perspective of a binary forest/non-forest map, it is likely that the 'flooded forest' landcover type is 'forest' falsely exhibiting (from the SAR backscatter signature) as 'non-forest' during periods of increased rainfall, rather than 'non-forest' falsely exhibiting as 'forest' during periods of decreased rainfall.

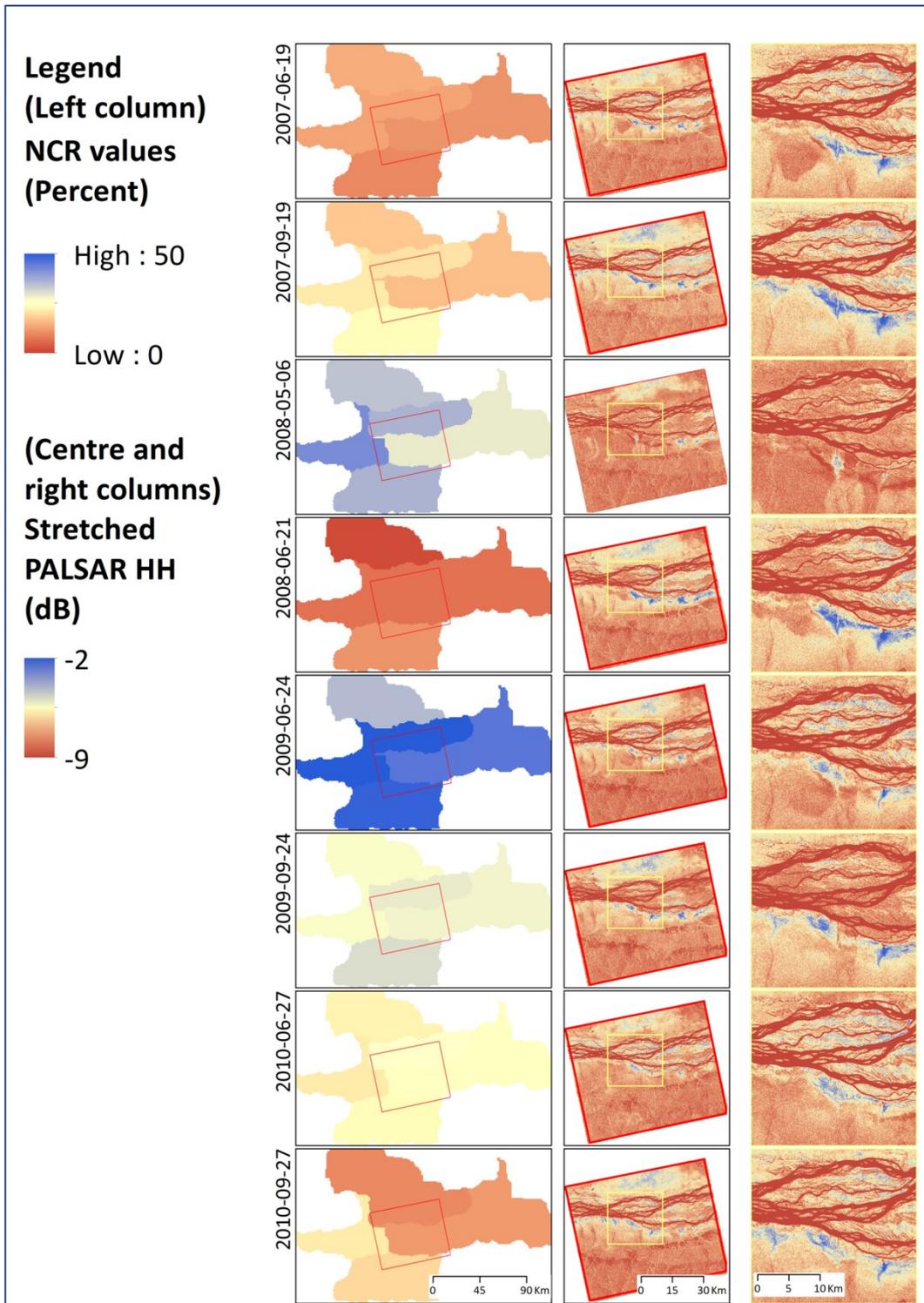


Figure 6.26: Comparison of NCR values with ALOS PALSAR HH σ^0 backscatter (at 15 m spatial resolution)

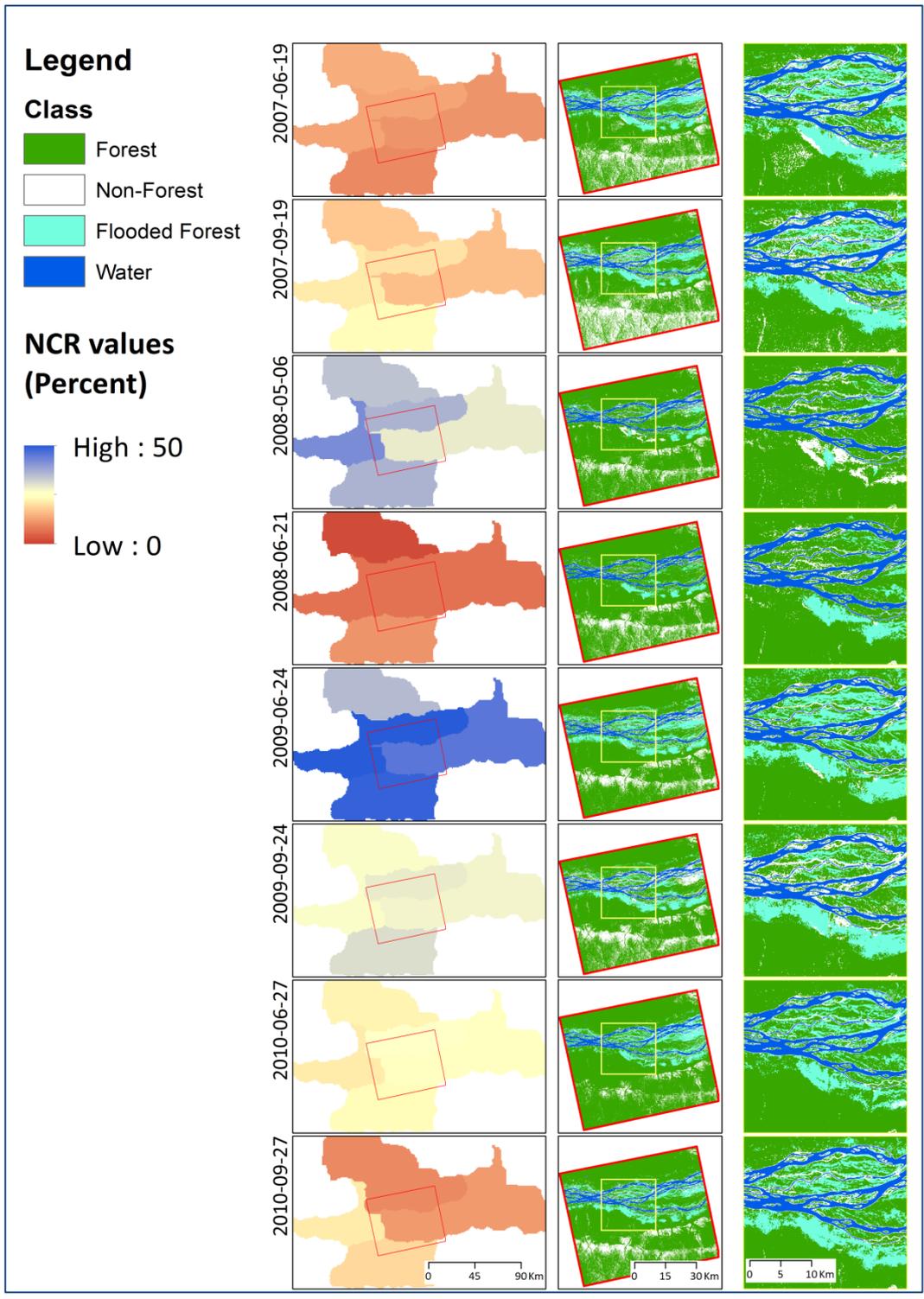


Figure 6.27: Comparison of NCR values with RF classification (at 15 m spatial resolution)

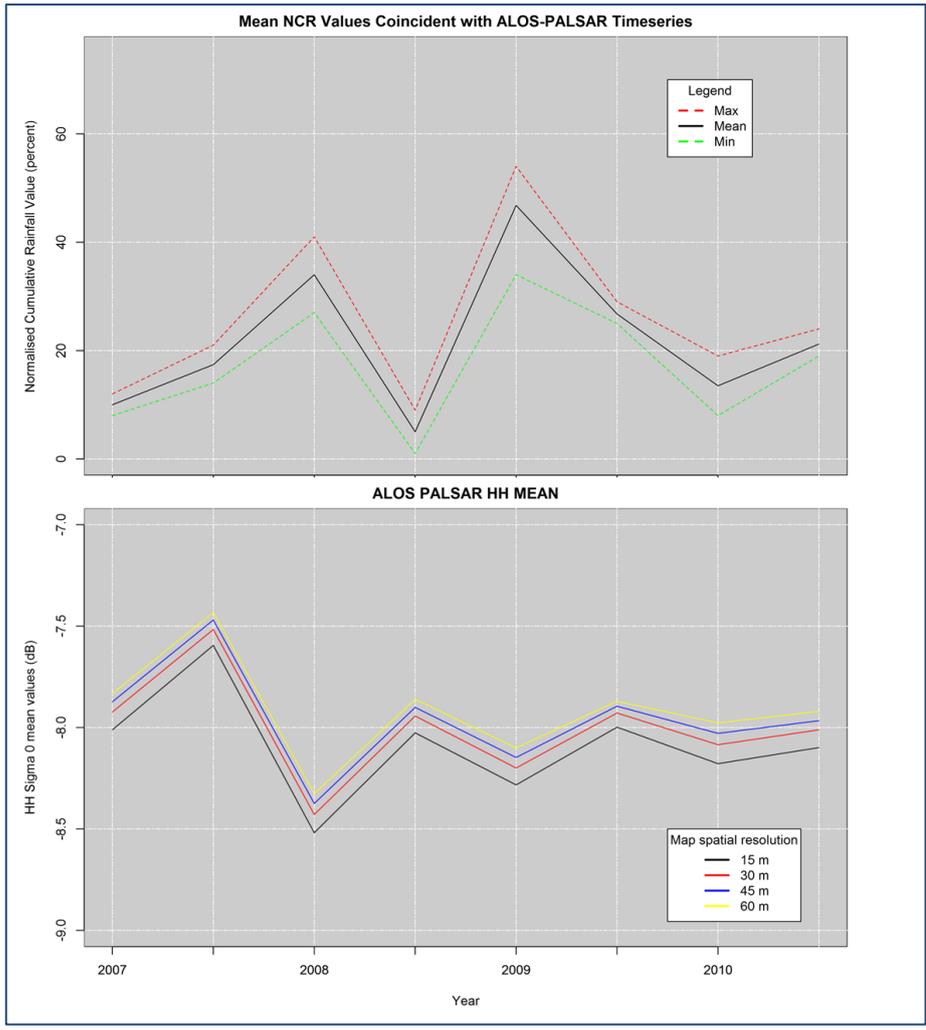


Figure 6.28: Comparison of mean NCR values with mean ALOS-PALSAR HH backscatter at multiple resolutions

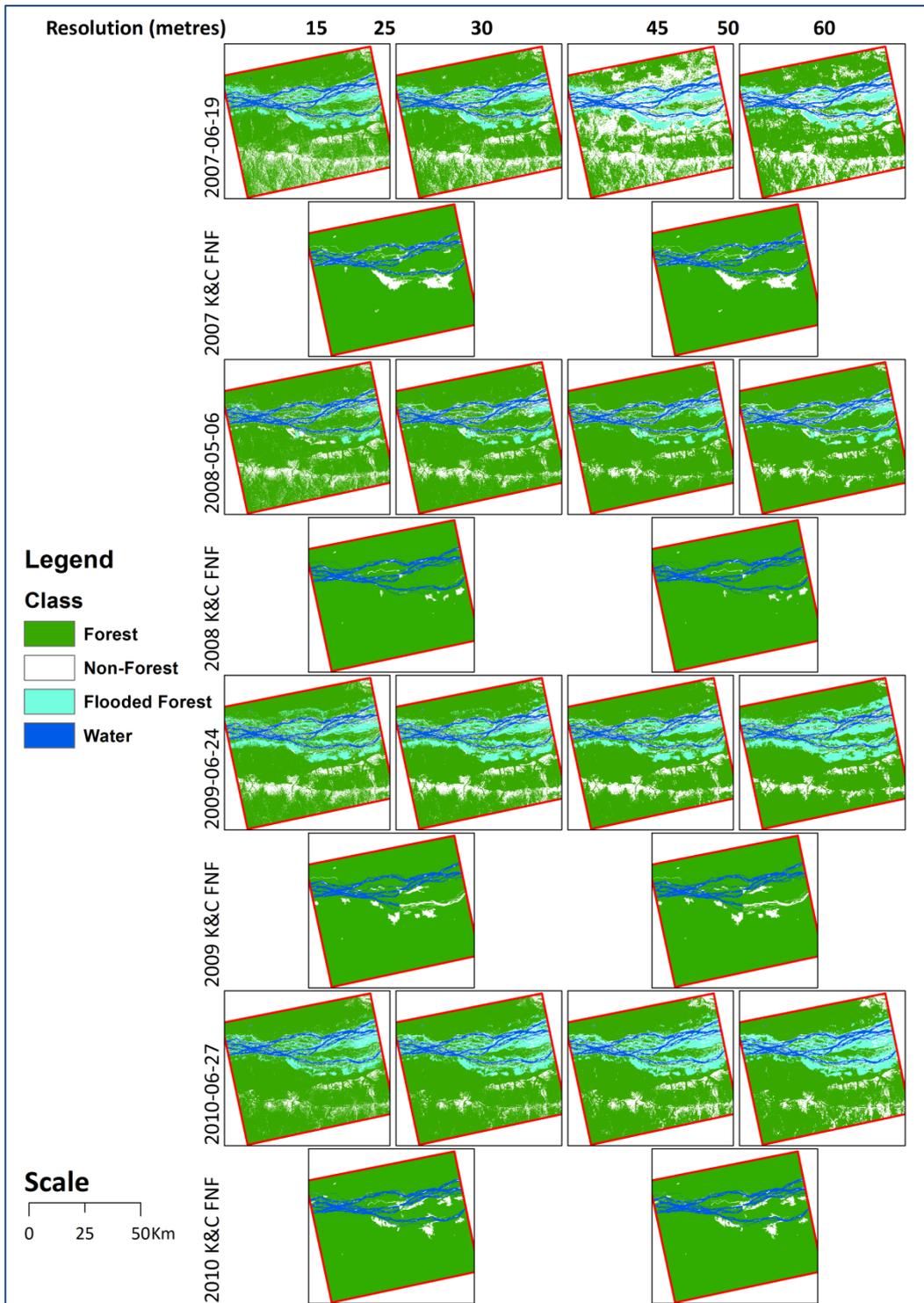


Figure 6.29: Comparison of RF classified scenes with the corresponding K&C Initiative Forest/Non-Forest product at multiple resolutions

6.4.8 AMSR-E Results

Due to the coarse resolution, there were few overlapping pixels of AMSR-E SM and VW in the study area, as seen in Figure 6.30 below. Ultimately, the comparison between AMSR-E and RFE 2.0 data, including the NCR, did not yield favourable results. This is in part due to the low number of data points available, but also due to the study area, as discussed in section 2.2.7. A comparison of NCR and dekadal sums over the whole study area, for the duration of the ALOS-PALSAR mission, was undertaken and the graph in Figure 6.32 demonstrates that the relationship was poor even over a large timeline.

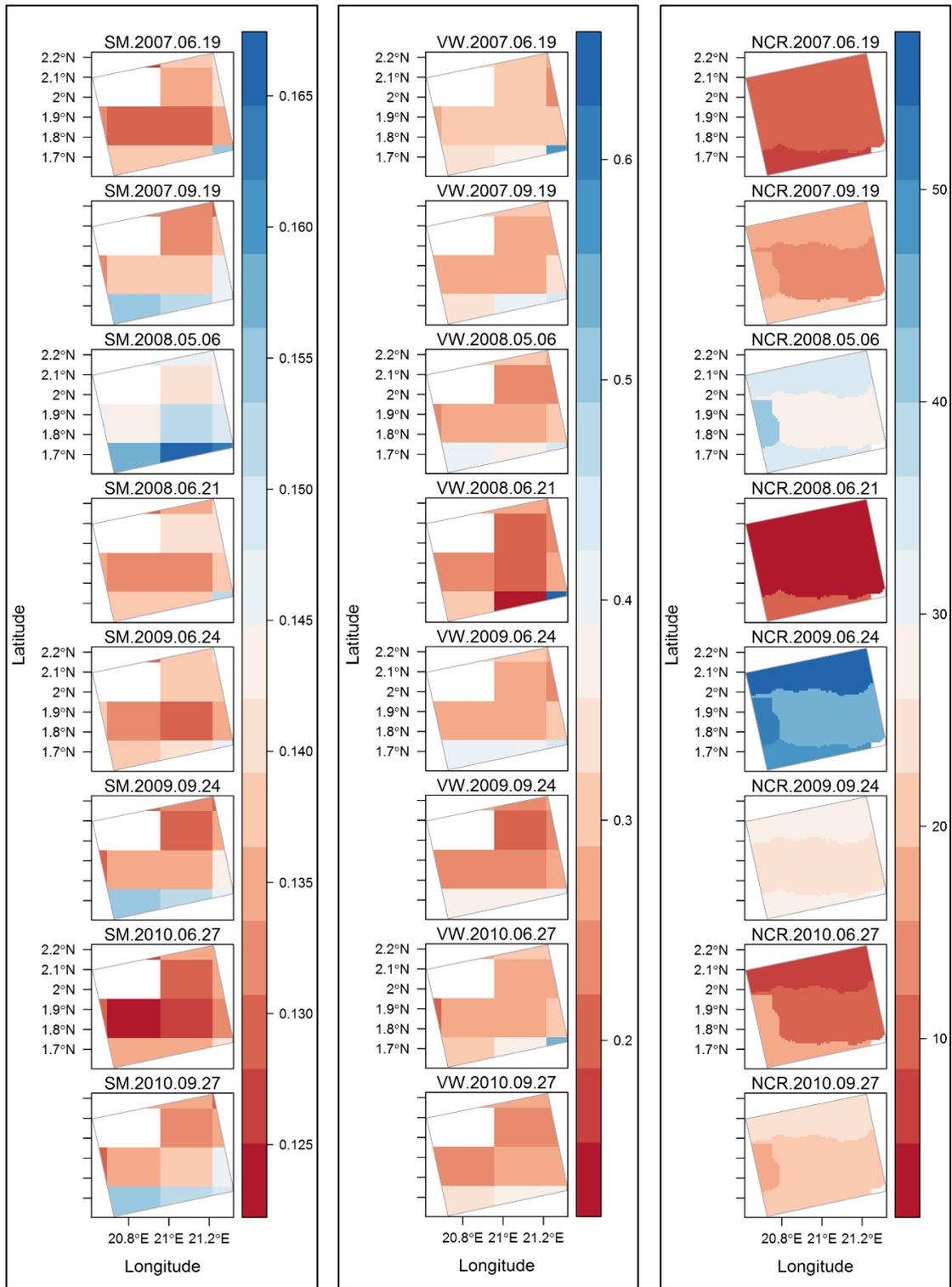


Figure 6.30: AMSR-E Soil Moisture [gcm^{-3}] and Vegetation Water Content [kgm^{-3}] compared with the NCR [%] in the study area

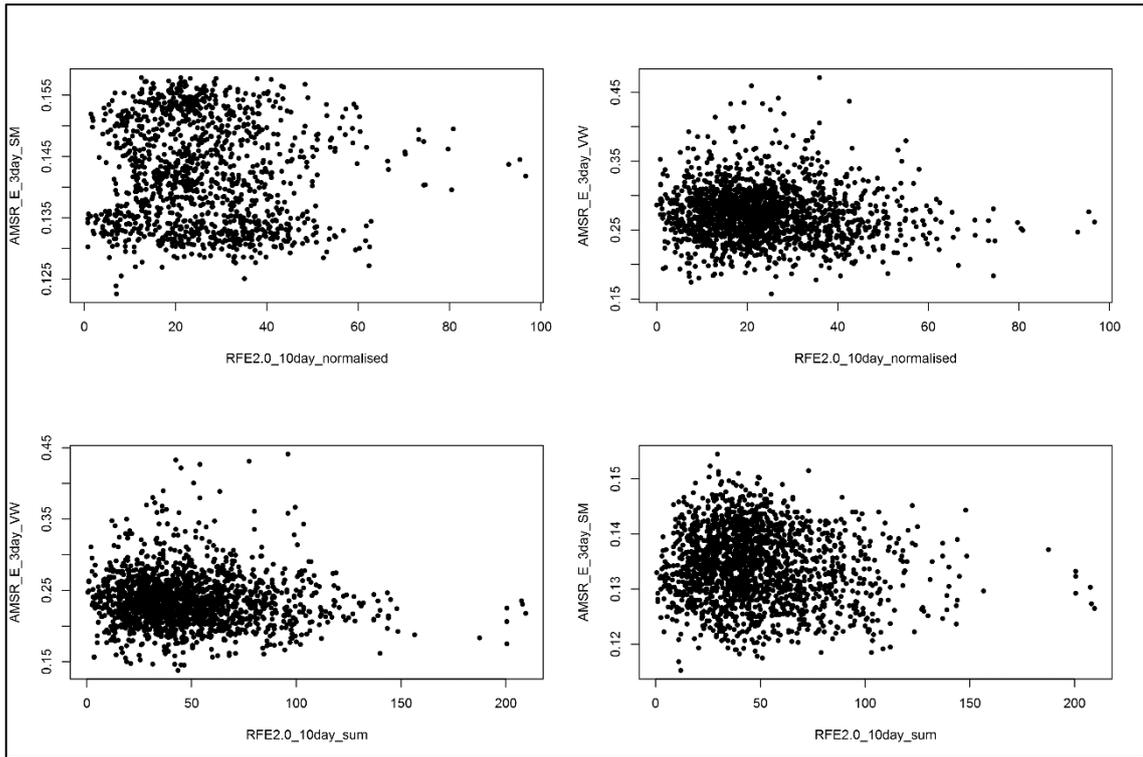


Figure 6.31: Comparison of AMSR-E VW (left) and SM (right) with NCR (top) and RFE 2.0 10 dekadal sum (bottom) in the study area, data points taken from the ALOS-PALSAR lifespan (January 26th 2006 until June 6th 2011)

6.5 Discussion

In general, the results obtained are in line with other similar studies using L-band SAR data to measure forest parameters in tropical rainforests. The use of texture information to improve the classification accuracy (by up to 4%) is seen in this study as it was in Thomas et al. (2015), Walker et al. (2010), and de Grandi et al. (2000a). Like the other studies which also used random forests or similar classification tree bagging algorithms (Ferreira-Ferreira et al., 2015; Bwangoy et al., 2010; Walker et al., 2010), this study performed tests to determine the optimal parameters to be used. Bwangoy et al. (2010) also showed the relative contribution of each layer. This was not a necessary step in this study as there were relatively few input features compared with the optical layers

used by Bwangoy et al., and as stated above the importance of the additional texture layers had already been established.

6.5.1 Reference Dataset

The high Moran's I values indicating clustering of the reference and training datasets is unfortunate, but unavoidable in a relatively small study site with few and obvious examples of certain classes (such as the river). It is possible that this may have caused bias in the classification, according to Legendre (1993), regardless of any rigorous visual and contextual quality control.

An approach that may at least decrease the clustering, while maintaining the speed of reference dataset creation, could be to produce a random or stratified random sample of points that act as seeds for the region growing ROI algorithm, which could then be limited to low spectral distance thresholds and minimum pixels per ROI. They could then be manually classified or discarded if they do not pass visual and contextual criteria.

6.5.2 Image Statistics and Multi-temporal Speckle Filtering

The first two results of this study, namely the increased class separability from multi-temporal speckle filtering of the SAR data and the increased speed of the RF classifier compared with the SVM classifier, were more expected than the other Chapter 6 results. The steps taken to obtain them and the outputs produced were necessary for the latter results: the scatter plots, class separability and histograms informed the training and validation process, as well as interpretation of classification results; the selection of RF over SVM enabled the use of a larger training dataset. The use of multi-temporal speckle filtering was required for obtaining meaningful results from the higher resolution SAR scenes. Speckle filtering is a standard procedure and was performed in several of the other SAR studies discussed in section 2.3 (de Grandi et al., 2000; Sgrenzaroli et al., 2004; Betbeder et al., 2014; Thomas et al., 2015), although without multitemporal speckle filtering due to the use of short time series data. Several studies used the K&C Initiative mosaic SAR data that is already somewhat filtered by resampling to 50m (Walker et al., 2010; Qin et al., 2015). SVM processing speed was not a considerable obstacle for processing a single ALOS-PALSAR scene, even at the higher 15 m spatial

resolution; multiple scenes with increased feature layers (such as additional texture metrics) could be processed in parallel using a high performance computing cluster. However, to expand this to a wider area with the increase in training pixels necessitated by that, use of the SVM classifier in its current form becomes untenable. For that reason, the remainder of the results used the RF classifier.

6.5.3 Scale Effects

Accuracy assessments produced using sampled validation data by definition are not a perfectly precise measure of accuracy. For this study, three factors contributed to a more robust accuracy assessment. Firstly, a single ALOS-PALSAR scene was used with validation points created using a semi-automated region growing algorithm, so the number of validation points is relatively high. The greater density of validation points makes it more representative of the image as a whole. This may not be feasible if the method is applied to a wider area, but the semi-automated process does facilitate an overall larger number of points, both for training and validation. Secondly, the validation points were specifically selected to be near class borders, a punitive (in terms of accuracy) measure designed specifically to identify likely errors (misclassifications have increased probability at class borders, particularly at coarser spatial resolutions). While this probably resulted in reduced accuracies, especially at coarser resolutions, it allowed a truer insight into class accuracy in the locations that are important for forest change detection. Thirdly, a whole scene statistical analysis of reported area for each class was undertaken. Unless an accuracy assessment produces a falsely low score, reduced accuracy should correspond to a greater variability with whole scene class areas. The below comparison of the coefficient of variability of class area with mean class accuracy at each resolution shows an inverse relationship between the two metrics across forest and non-forest classes.

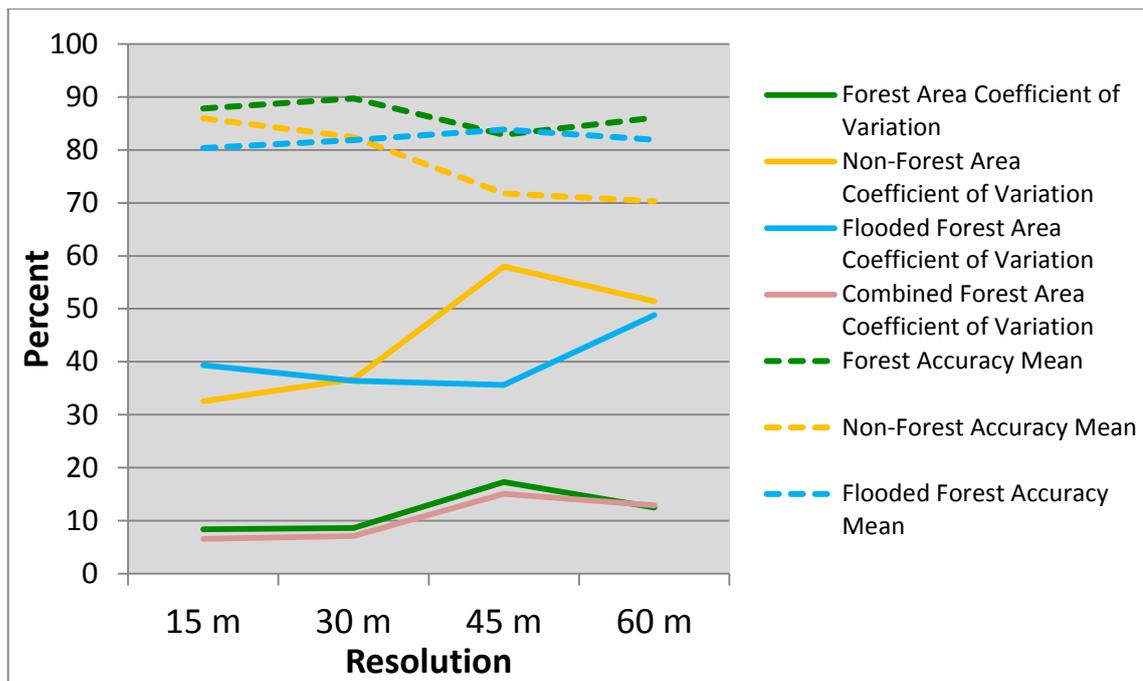


Figure 6.32: Comparison of 8-scene coefficient of variation of class area with 8 scene mean class accuracy at multiple resolutions

The combined forest area coefficient of variation in Figure 6.32 is included to show the reduction in variability when the two classes are combined.

The improvements in accuracy of the classification at higher spatial resolution, as well as the more stable class area measurements particularly of non-forest, answer the question of the influence of scale on forest measurements in the Congo basin. This result could inform efforts to both establish a reliable map of baseline forest cover in the Congo. In addition, there is no significant difference between the accuracy and class area reporting between the 15 m and 30 m resolutions, with the drops in accuracy and variability of area reporting occurring at 45 m and 60 m. This is a positive result given the global availability of ALOS-PALSAR mosaics at 25 m resolution from JAXA's K&C Initiative, as well as hoped for ALOS-2 data in the coming years. One potential drawback is the reduced number of dates (four compared with eight used in this study) which may reduce the effectiveness of the multi-temporal speckle filtering process. The forest/non-forest products produced by JAXA do not appear, at least in this scene, to show any improvements in classification with the change in resolution, and the reasons for this requires further investigation, given the results of this study. One possible reason may be the OBIA approach used, which classified based on mean pixel statistics from

segments (Shimada et al., 2014) – if the same segment polygons were used from the 50 m classification rather than a completely new (and computationally expensive) segmentation performed at 25 m, then no major changes could be expected. Another explanation could be the multi-temporal speckle filtering which was not used in the JAXA product. For other studies described in section 2.3 using L-band SAR data in the same region of the Congo Basin (Betbeder et al., 2014; Bwangoy et al., 2010; de Grandi et al., 2000a), the scale used has been relatively coarse ($\geq 100\text{m}$ resolution). The results of the scale analysis in this study suggest that applying the same methods to higher resolution data could yield improved results.

6.5.4 NCR Effects

Establishing a link between NCR values and HH backscatter in the study area, and in the flooded forest class, had several potential challenges. The first was the possibility that none, or all, of the eight available scenes experienced relatively high rainfall in the ten days leading to image acquisition. Figure 6.28 shows some variability in mean NCR from the five contributing sub-basin catchments over the eight dates, although not as much when compared with the full range of values (Figure 6.6). The next possibility was that the local rainfall had less of an influence than regional rainfall in the wider area, particularly upstream, and the NCR of the five selected sub-basins would not correlate with either of the targeted values. Conversely, one of the local sub-basin catchments could have had a disproportionately influential effect on the scene. The minimum and maximum NCR values from Figure 6.28 show that even if this were the case, there is not enough variability to measure the effect of any individual sub-basin catchment.

There is a weak inverse correlation between NCR mean values and mean HH backscatter, with a coefficient of determination (R^2) value of 0.2757 (Figure 6.33), rising to 0.4283 if a single outlier (the second scene) is removed. This is not high enough to establish a definite link, and may be caused by there being too few scenes with too little variability of NCR values. The result of the weak correlation is that it is difficult to ascertain from the available data whether NCR is a robust metric to use for determining extremity of flood events in ALOS-PALSAR scenes. Further data would be needed to either establish or disprove a link, perhaps from scenes acquired during known flood events in the region. It would be possible to apply the NCR method to the whole Congo

Basin and test it against the K&C Initiative mosaics. This would be computationally slow, but more likely to yield a definitive result.

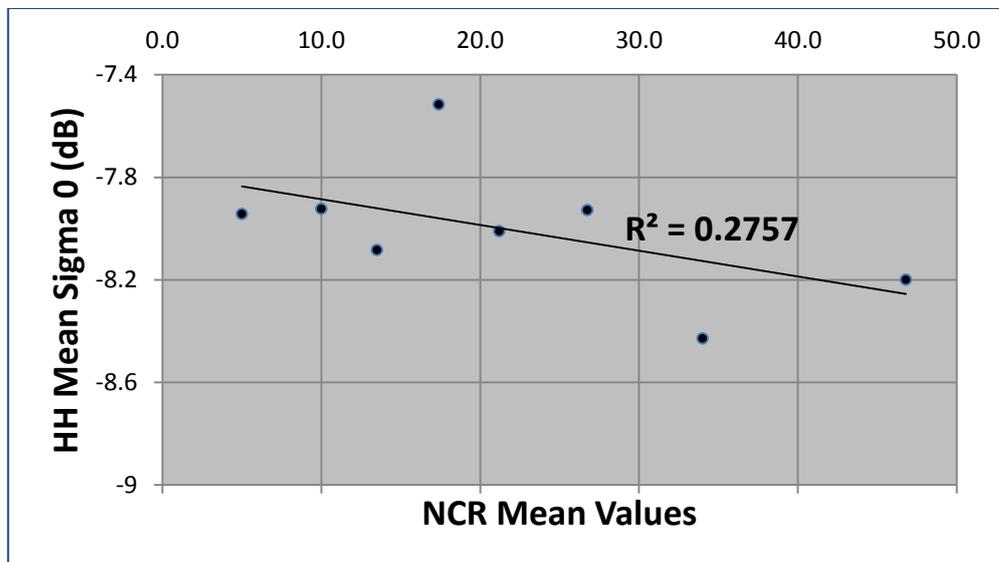


Figure 6.33: Relationship between mean ALOS-PALSAR HH σ^0 backscatter and mean NCR values

While it is difficult to link NCR with the areas classified as flooded forest, the HH backscatter values and range and the proximity to the Congo River suggest that they are likely to be affected by local hydrology. In addition, from the high resolution optical imagery used in generating training and validation datasets, and the HV backscatter values, the flooded forest areas are demonstrably tree covered. The value of analysing HH range in proximity to major waterways and in flood plains remains, and the creation of a separate flooded forest class is strongly recommended by this study. This is in line with previous research, including that by Betbeder et al. (2014), and Bwangoy et al. (2010), which used HH channel L-band SAR data to define wetland classes.

6.6 Conclusions

Chapter 6 aimed to provide answers to the final two research questions. First of all:

To what extent, and how, can spaceborne SAR data in the Congo Basin, combined with rainfall data metrics, account for seasonally inundated forest in SAR-derived forest maps?

An existing rainfall data metric (BERM) was adapted and modified to generate a simple and novel metric (NCR) with potential for wider applicability to SAR interpretation in the Congo Basin, the rest of continental Africa, and other locations worldwide with available consistent daily rainfall information. In the scene selected, the range of HH values across a timeseries was shown to be a reliable indicator of flooded forest, supporting the work of Betbeder et al. (2014), once water bodies were masked out. It was shown that areas of flooded forest exhibit HH backscatter signatures in drier periods that are higher than normal forest HH backscatter signatures, and lower backscatter signatures in wetter periods, which is unexpected given established relationships between vegetation water content and radar backscatter. The two peak mean NCR values appear to correlate inversely with mean HH values (Figure 6.28), but displayed a poor R^2 value, perhaps due to variable backscatter values in the six drier scenes (out of a total of eight). Despite certain areas occasionally being classified as non-forest, the flooded forest class was demonstrably shown to be a subset of the forest class.

Secondly:

What are the effects of scale on Central African forest maps generated from SAR data, and how does this impact the accuracy of maps generated from coarser resolution wide coverage SAR datasets?

Using a robust validation method, this study showed an inverse correlation between variability of class area reporting and mean class accuracy across four resolutions tested. The results of this showed that a decrease in accuracy of the dominant classes, and especially non-forest, accompanied a decrease in spatial resolution, and was also related to greater variability in class area reporting. The coefficient of variability was used to attempt to balance the dominance of the forest class throughout the scene. The differences between the two higher resolutions were not significant, whereas there was a divide between the two higher and the two lower resolutions. The impact this could have on coarser resolution wide coverage SAR datasets is that greater confidence can be placed in the 25 m resolution K&C Initiative mosaic data than the previously released 50 m data, with a caveat that changes are implemented in the classification approach used, and a multi-temporal speckle filter is applied, as it was shown to greatly improve

class separability. This feeds into the recommendations for research question 1, and the conclusions of chapter 5, demonstrating that a higher resolution mosaic product, if processed properly and in conjunction with appropriate ancillary datasets, can improve mapping accuracy and stability across forest and non-forest classes.

Chapter 7: Discussion and Conclusions

7.1 Discussion in the Context of Literature

As stated in Chapter 3, the top-down approach followed across Chapters 4-6 aimed to refine the study area selection while improving the classification. This was achieved, and the results obtained at all stages inform the research questions and contribute to existing knowledge in this domain.

Chapter 4 showed that across a broad area of the Congo Basin, existing optical based forest cover products contradict SAR classification, and vice versa. Misclassification of large areas of cropland as forest by the LTC was identified through a relatively simple decision tree classifier. Weaknesses of the available L-band SAR product, such as high HV backscatter in urban areas causing misclassification, added weight to the use of the HH channel to characterise non-forest, confirmed by the methods of JAXA in their forest/non-forest product generation (Shimada et al., 2014). The area analysed in Chapter 4 completely covered the Congo Basin and much further beyond. The effort expended in interpreting and preparing the initial release of the K&C Initiative dataset for comparison with the optical dataset was useful in understanding the difficulties of SAR image calibration and geocoding when creating such a mosaic, even though the more manageable 50 m release, with two additional years of data, was made available by JAXA in January 2014.

Chapter 5 covered a smaller, but still considerable area of the Congo rainforest. The potential danger of an urban non-forest class derived from automated HH thresholds was identified, particularly in a semi-transparent mode that does not distinguish it from other non-forest types. The use of a global object-oriented approach was shown to have lost the ability to identify finer scale forest changes that would otherwise be picked up by the resolution of the imagery available. While the validation showed a clear advantage of the trained SVM classifier over JAXA's K&C Initiative forest/non-forest approach, and obvious misclassifications of moisture-affected forest were avoided, the accuracy assessment was nevertheless not reliable enough to give confidence in the reported class areas.

Results from all chapters, in particular chapters 4 and 5, demonstrate clear improvements in the current state of forest monitoring abilities in the regions studied, in terms of **spatial resolution** (DeFries et al., 2002; Mayaux et al., 2004; Rosenqvist et al., 2000; Hansen et al., 2003; Arino et al., 2007; De Grandi et al., 2000a; Bwangoy et al., 2010; Betbeder et al., 2014), **recency of acquired data** (Rosenqvist et al., 2000; De Grandi et al., 1999; DeFries et al., 2000; Lindquist & D’Annunzio, 2016; Mayaux et al., 2002; Hansen et al., 2003), **completeness of coverage** (Lindquist & D’Annunzio, 2016; Sexton et al., 2013; Thomas et al., 2015; Betbeder et al., 2014), **relevant and comparable measures of accuracy** (Shimada et al., 2014; Sgrenzaroli et al., 2004; Bwangoy et al., 2010; Ferreira-Ferreira et al., 2015; Mayaux et al., 2004; De Grandi et al., 2000a), and **user replicability**, either from transparency of algorithms, data availability (respectively, Hansen et al., 2013; Bueso-Bello et al., 2017), or in terms of **clear forest definitions** used (Walker et al., 2010; Betbeder et al., 2014; Ferreira-Ferreira et al., 2015; Thomas et al., 2015; Bueso-Bello et al., 2017).

The development of the NCR index in this study provides a useful way to continue research into managing the effects of moisture on SAR-based land cover classifications in the tropics, and with the advent of higher temporal resolution datasets, in particular Sentinel-1 C-band SAR, as well as further K&C Initiative L-band SAR mosaics, it could provide a simple and quantifiable method to establish the relationship of rainfall on SAR signals over wide areas in the tropics.

7.2 Conclusions in the Context of Research Questions

7.2.1 Question 1: Ideal SAR Imagery and Methods for Tropical Forest Mapping

Chapters 4 and 5 mainly addressed the first research question ‘*What are the ideal image acquisition parameters and classification techniques for an annual forest map in the Republic of Congo from spaceborne SAR data, given currently available data and regional seasonal effects on image quality?*’ through a combination of demonstrable improvements over existing products and a process of elimination. In summary, it was found that the first criterion of ideal image acquisition should be accurately geocoded

SAR imagery, or a technique to accurately geocode it, which is not a trivial task to produce on a continental scale; the JAXA release was then fortunately made available. Segmentation, while a global solution, either needs to be discarded in favour of high resolution exploiting pixel-based classification methods, or improved with additional region specific decision nodes and a segmentation algorithm that is tuned to the regional backscatter signature class separations, particularly in areas vulnerable to deforestation and forest degradation. Both chapters acknowledged the need for the acquisition of adjacent images in as short a time frame as possible when producing a wide area mosaic in the Congo basin. This reduces the likelihood of seasonal backscatter differences, which can be difficult to correct for especially without metadata relating to the scenes comprising the mosaic. Metadata was largely unavailable for the data used in Chapter 4, but was present in Chapter 5 (Figure 5.1). This included contributing scene acquisition date, crucial in order to work with concurrent ancillary datasets, and local incidence angle, allowing terrain corrections to be made (although the dataset by this stage arrived pre-calibrated for terrain differences).

While the SVM classifier in Chapter 5 was shown to be an improvement over the compared method, it was slow to process with available computing equipment and, in the form used, difficult to batch process. These two factors negated any considerable tuning of the available parameters to improve on the original classification. The ability to script and therefore automate the classification process was identified, as well as the use of a classifier that could use a greater number of training points without suffering from a considerable reduction in processing speed. Chapter 5 also showed that more work was necessary to improve the training dataset quality, particularly for non-forest areas. The application of water and urban masks, as well as the inclusion of a separate class for flooded forest could have removed a source of error, as carried out in Chapter 6. It was proposed that daily rainfall estimate data could be useful for validating areas of flooded forest in the mosaics, as well as selecting scenes from the mosaics with more uniform acquisition conditions, using a method similar to that of (Lucas et al., 2010). (Watanabe & Shimada, 2006)

Chapter 6 continued Chapter 4 and 5's findings for research question 1 in several ways. First of all, the random forest classifier was proved to be a much more efficient, if not

more accurate predictor of the classes. The flooded forest class was introduced, and trained using a combination of HH Range and optical imagery. The stratified sampling method for selection of training data used in Chapter 5 was eschewed in favour of a more efficient, semi-automated method that allowed selection of relatively even numbers of class training samples in a scene dominated by the forest class, which was also used for a similar and more robust validation dataset.

7.2.2 Question 2: HH Range and NCR

As previously described in section 6.6, Chapter 6 aimed to answer the final two research questions. First of all:

To what extent, and how, can spaceborne SAR data in the Congo Basin, combined with rainfall data metrics, account for seasonally inundated forest in SAR-derived forest maps?

Question 2 was answered, and in the process an existing rainfall data metric (BERM) was adapted and modified to generate a simple and novel metric (NCR) with potential for wider applicability to SAR interpretation in the Congo Basin, the rest of continental Africa, and most locations worldwide with available consistent daily rainfall information. In the scene selected, the range of HH values across a timeseries was shown to be a reliable indicator of flooded forest, once water bodies were masked out. It was shown that areas of flooded forest exhibit HH backscatter signatures in drier periods that are higher than normal forest HH backscatter signatures, and lower backscatter signatures in wetter periods. Despite certain areas occasionally being classified as non-forest, the flooded forest class was demonstrably shown to be a subset of the forest class.

7.2.3 Question 3: Scale effects

What are the effects of scale on Central African forest maps generated from SAR data, and how does this impact the accuracy of maps generated from coarser resolution wide coverage SAR datasets?

Question 3 was also addressed by Chapter 6, as concluded in section 6.6, and answered first using a more robust validation method, which also showed an inverse correlation between variability of class area reporting and mean class accuracy across four resolutions tested. The results of this showed that a decrease in accuracy of the dominant classes, and especially non-forest, accompanied a decrease in spatial resolution, and was also related to greater variability in class area reporting. The coefficient of variability was used to attempt to balance the dominance of the forest class throughout the scene. The differences between the two higher resolutions were not significant, whereas there was a divide between the two higher and the two lower resolutions. The impact this could have on coarser resolution wide coverage SAR datasets is that greater confidence can be placed in the 25 m resolution K&C Initiative mosaic data than the previous 50 m data, with a caveat that changes are implemented in the classification approach used, and a multi-temporal speckle filter is applied, as it was shown to greatly improve class separability.

7.3 Future Direction of SAR Remote Sensing in the Congo Basin

The number of spaceborne SAR sensors is set to increase, particularly in the domain of longer wavelength L- and P-band sensors. Forest remote sensing has not been the primary objective of any spaceborne SAR sensors to date, so the modes of acquisition and acquisition strategy have not always been optimal for this application. However, this will change with the BIOMASS mission, which will carry the first spaceborne P-band sensor and is dedicated to the monitoring of forest *AGB* (Le Toan et al., 2011). The Copernicus Sentinel 1 constellation, of which one satellite was launched in 2014 and a second launched in 2016, provides a potential six-day repeat acquisition cycle. While the C-band SAR sensor is not as suited to monitoring of small changes in *AGB* or precise measuring of carbon stocks, it will be a useful tool for rapid monitoring of forest disturbance (ESA, 2014) and the short repeat pass will lend itself to analysis of interferometric coherence measures, as mentioned in section 2.2.5. More recent studies are being encouraged (Reiche et al., 2016) to use Sentinel-1 to overcome the

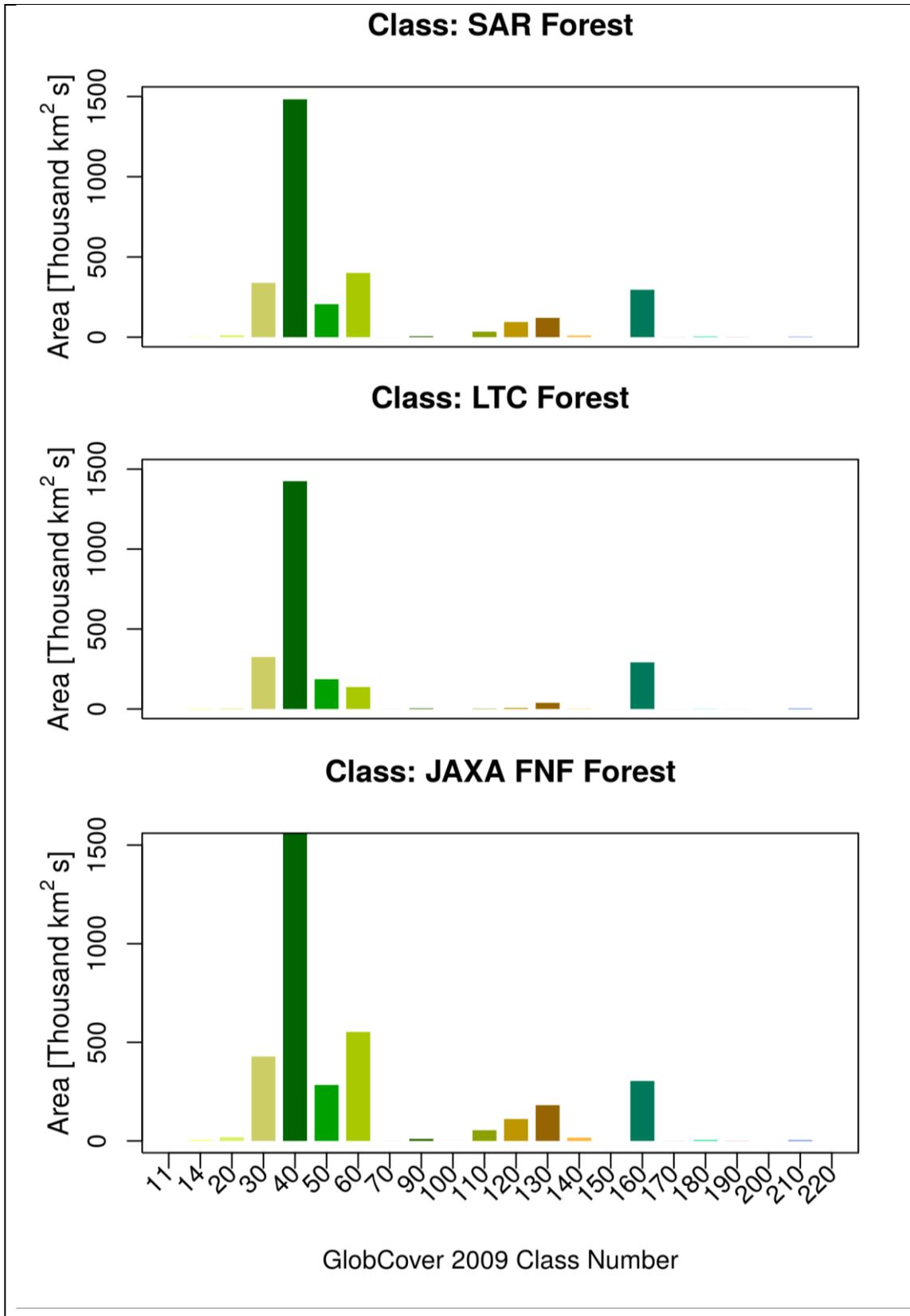
limitations of C-band SAR for tropical forest monitoring, through the analysis of very high temporal resolution data 'cubes'; this is only very recently becoming possible in the pan-tropical zone due to limited persistent coverage outside Europe. Optimal temporal coverage from Sentinel 1 (6-day or higher repeat visit frequency, from overlapping tiles and use of ascending and descending modes) has centred on Europe, with sporadic and incomplete coverage of the tropics, due to the physical power limitations of the sensor.

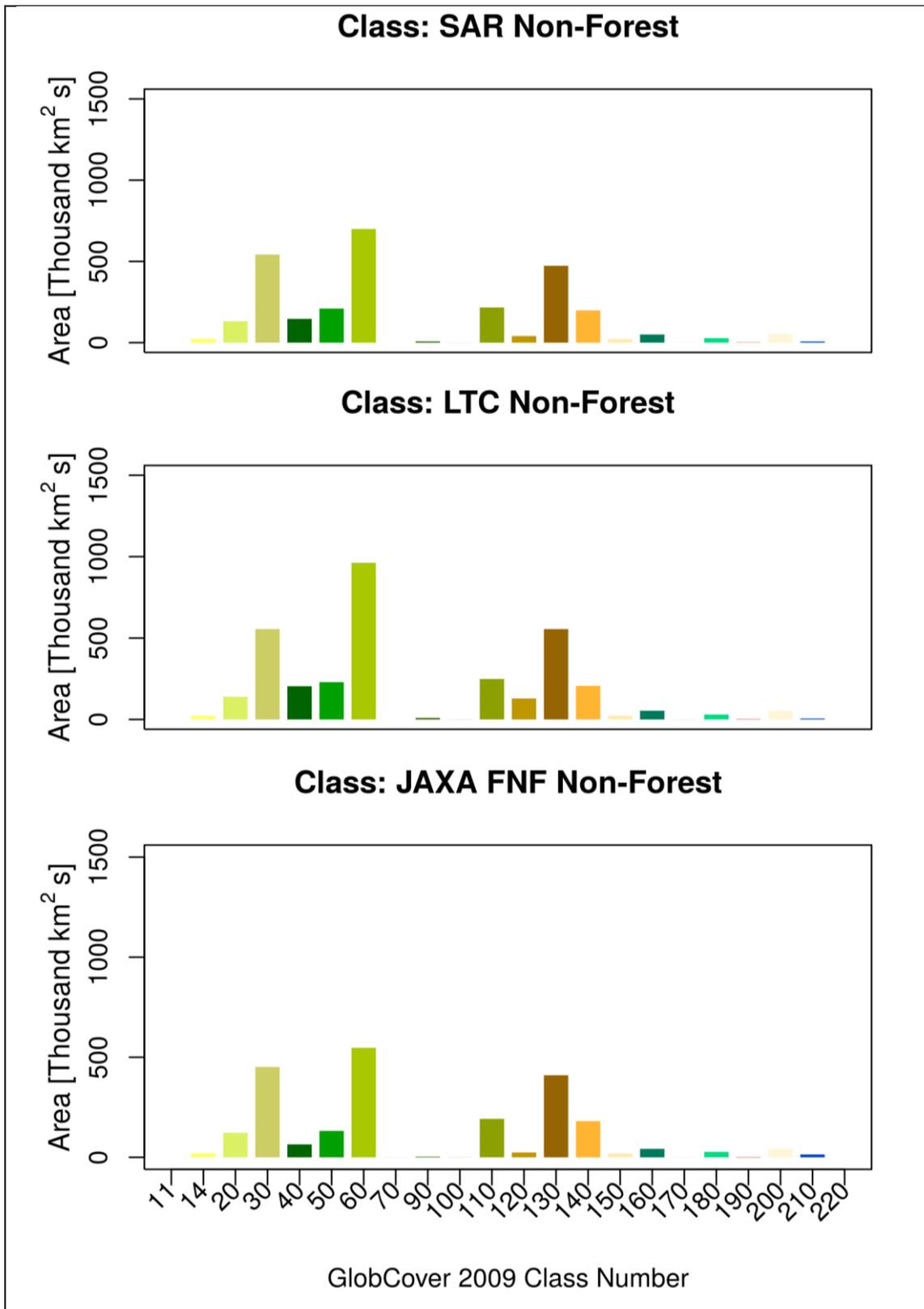
The launch of ALOS-2 with its continued contribution towards the K&C Initiative (Rosenqvist et al., 2014), and the planned launch of SAOCOM-1A and 1B in 2018 (CONAE, 2014) will continue the role of spaceborne L-band SAR for forest remote sensing.

It is generally the case that newer SAR sensors (since 2000) have the capability for acquisition in any of the four polarisation configurations, although acquisition of simultaneous fully polarimetric scenes remains limited to experimental applications. More advanced techniques of SAR that exploit the interferometric analysis of two or more similarly acquired images of the same scene can offer further potential for SAR, and this has been demonstrated in large-scale studies of boreal forest (Wagner et al., 2003; Tansey et al., 2004; Thiel et al., 2009). This will be enhanced by shorter repeat pass acquisitions and longer wavelengths of SAR to overcome the difficulties this technique faces in wetter, higher AGB-containing tropical forests (Le Toan et al., 2011).

Appendix A - Chapter 4 additional figures

A1 GlobCover 2009 distribution comparisons





Appendix B – Chapter 6 additional figures

B1 Haralick GLCM Texture Class Histograms

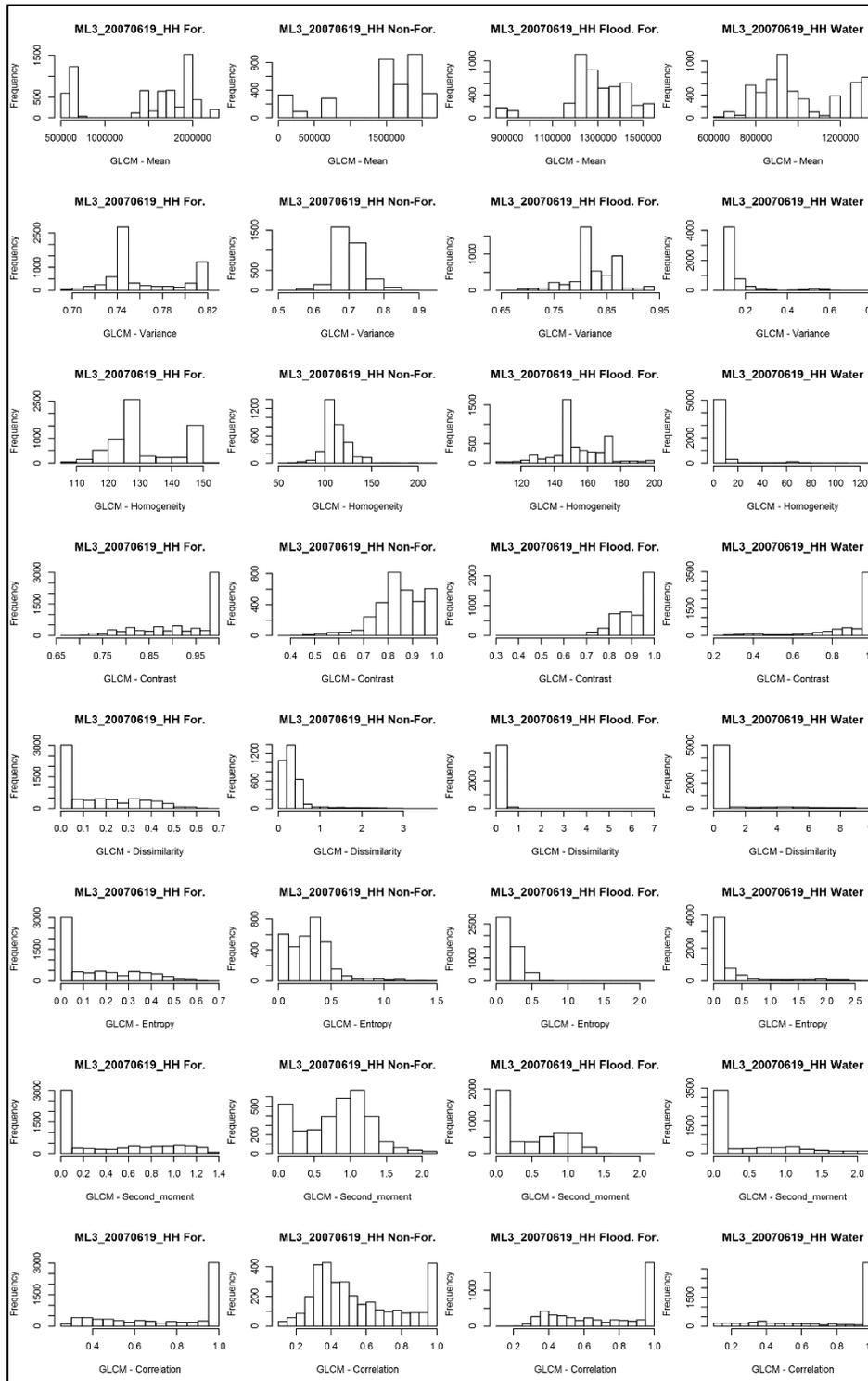
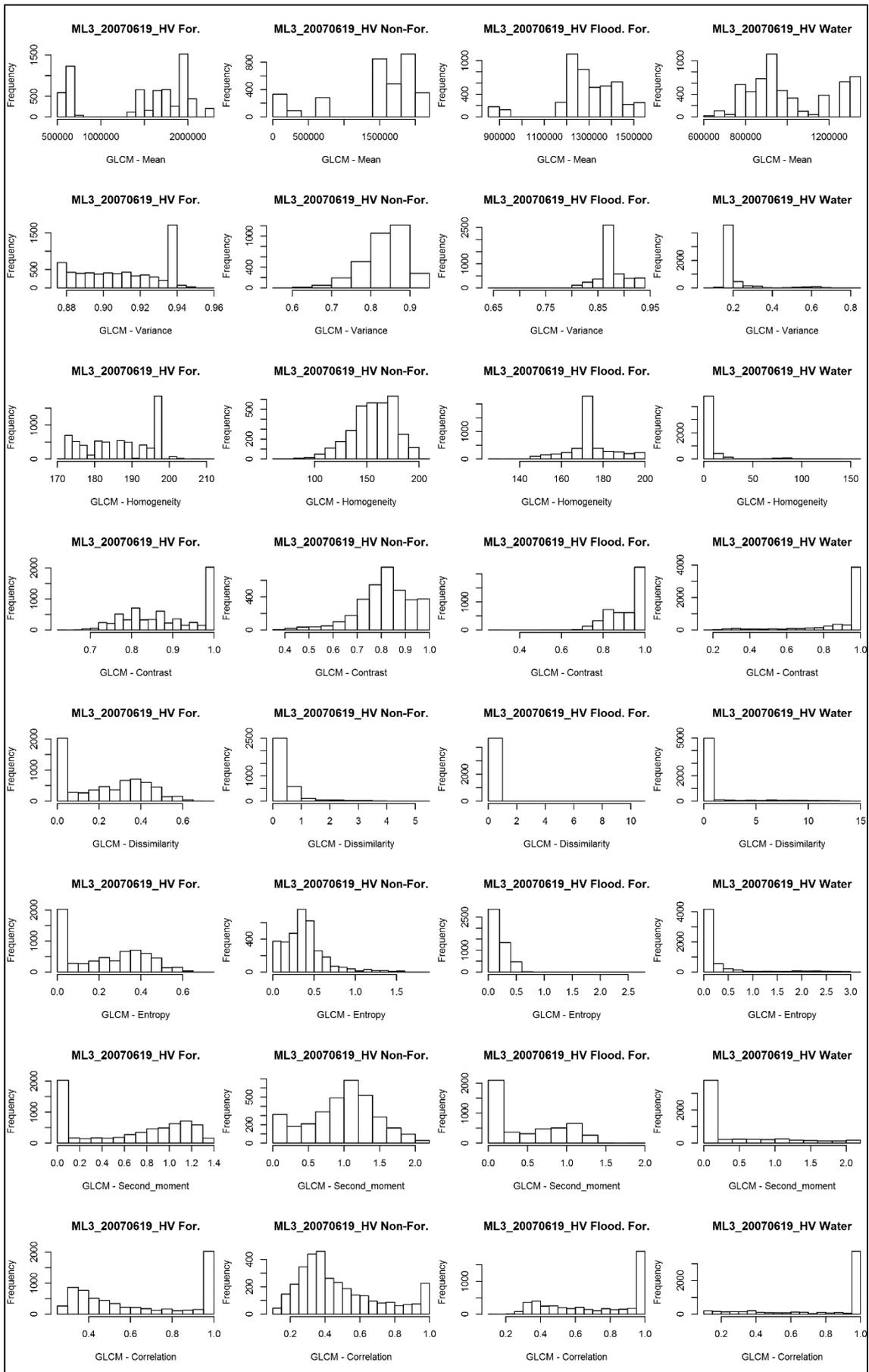
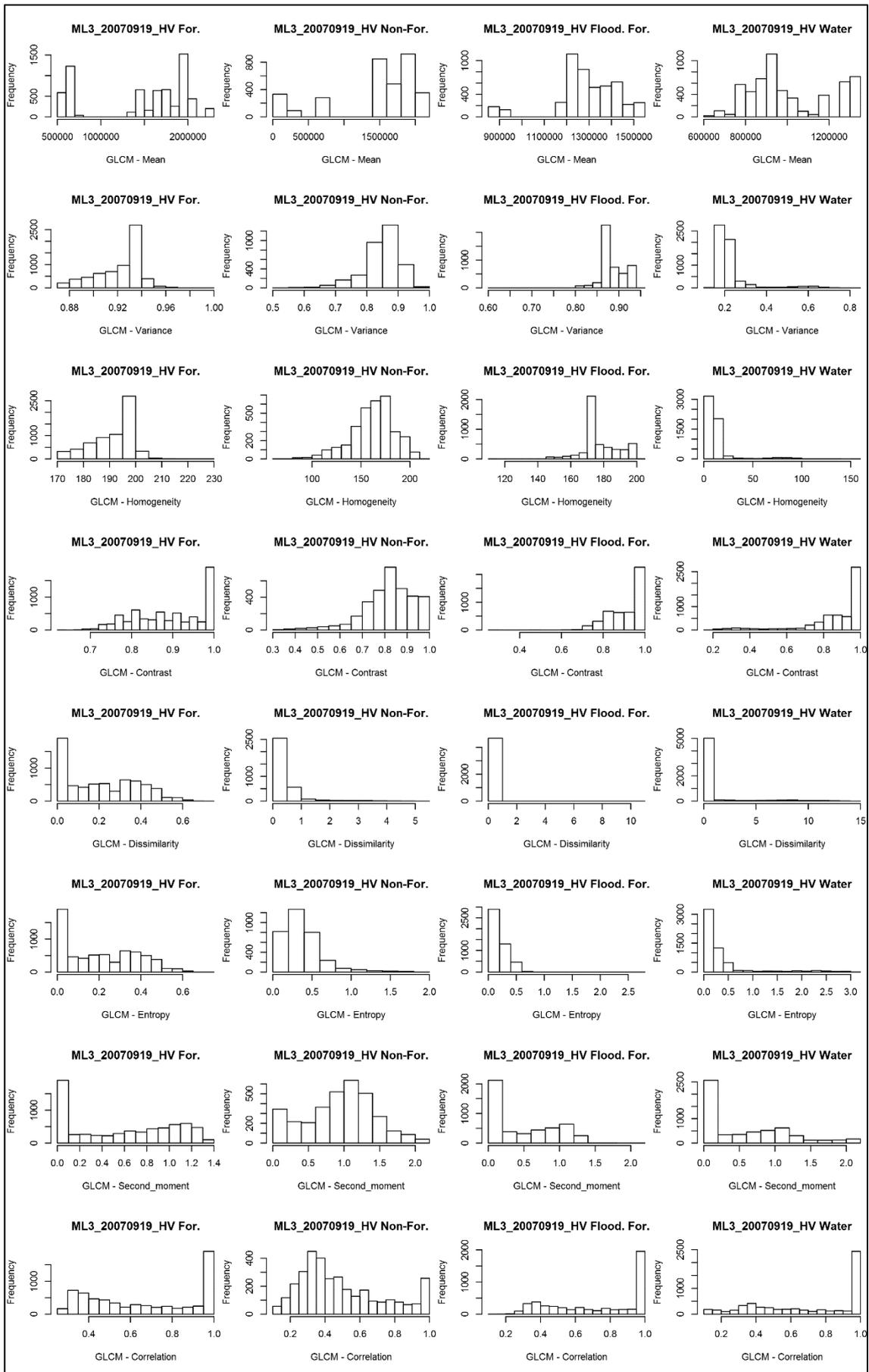
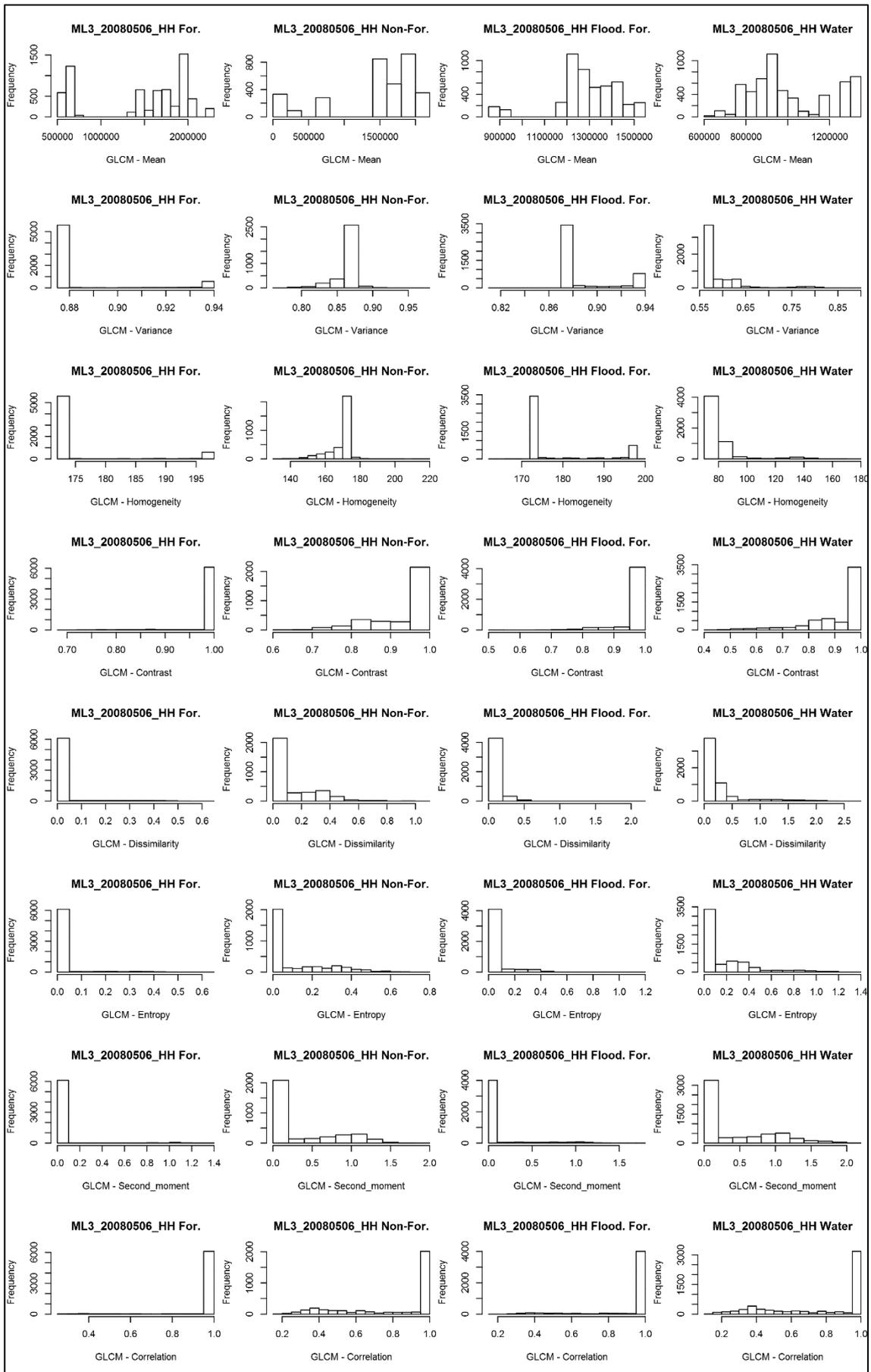
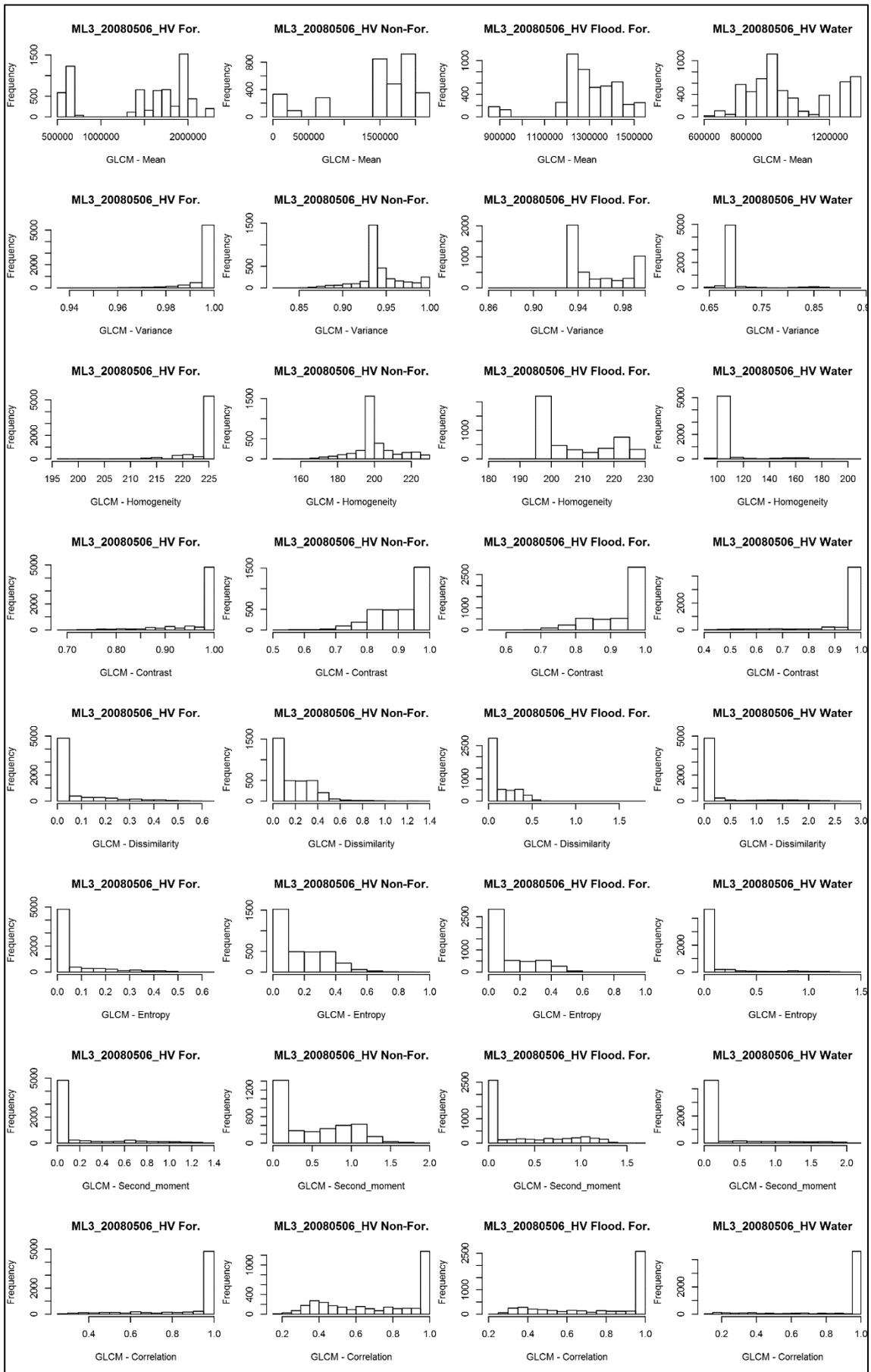


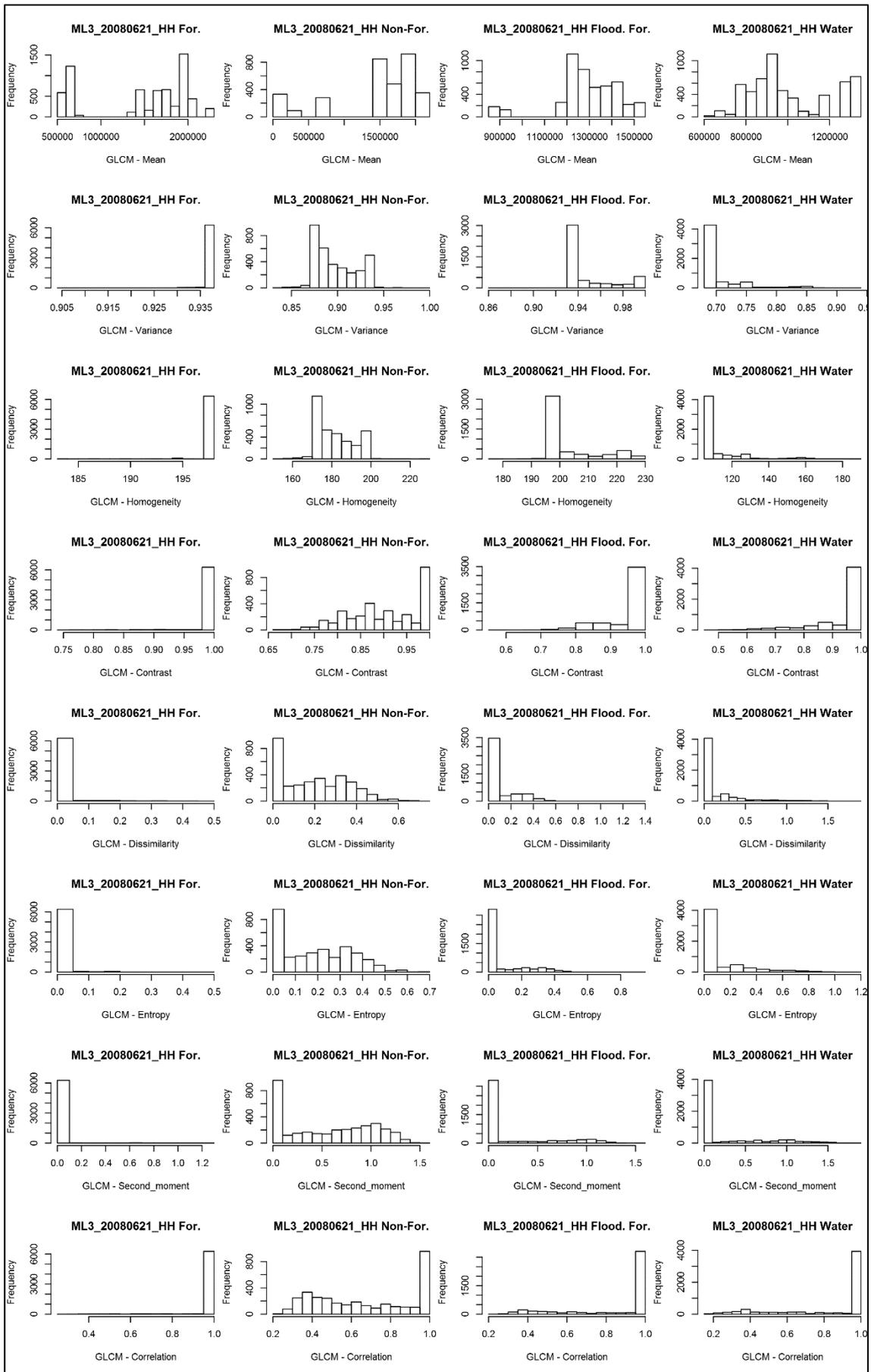
Figure B1: Histograms of the eight Haralick GLCM texture measures from Training and Validation sample areas; labels indicate the date, polarisation channel and class; frequency scales are different in each histogram to better illustrate range of values in more homogeneous texture measures. Texture generated from 45 m ALOS-PALSAR data; continued in following 15 pages

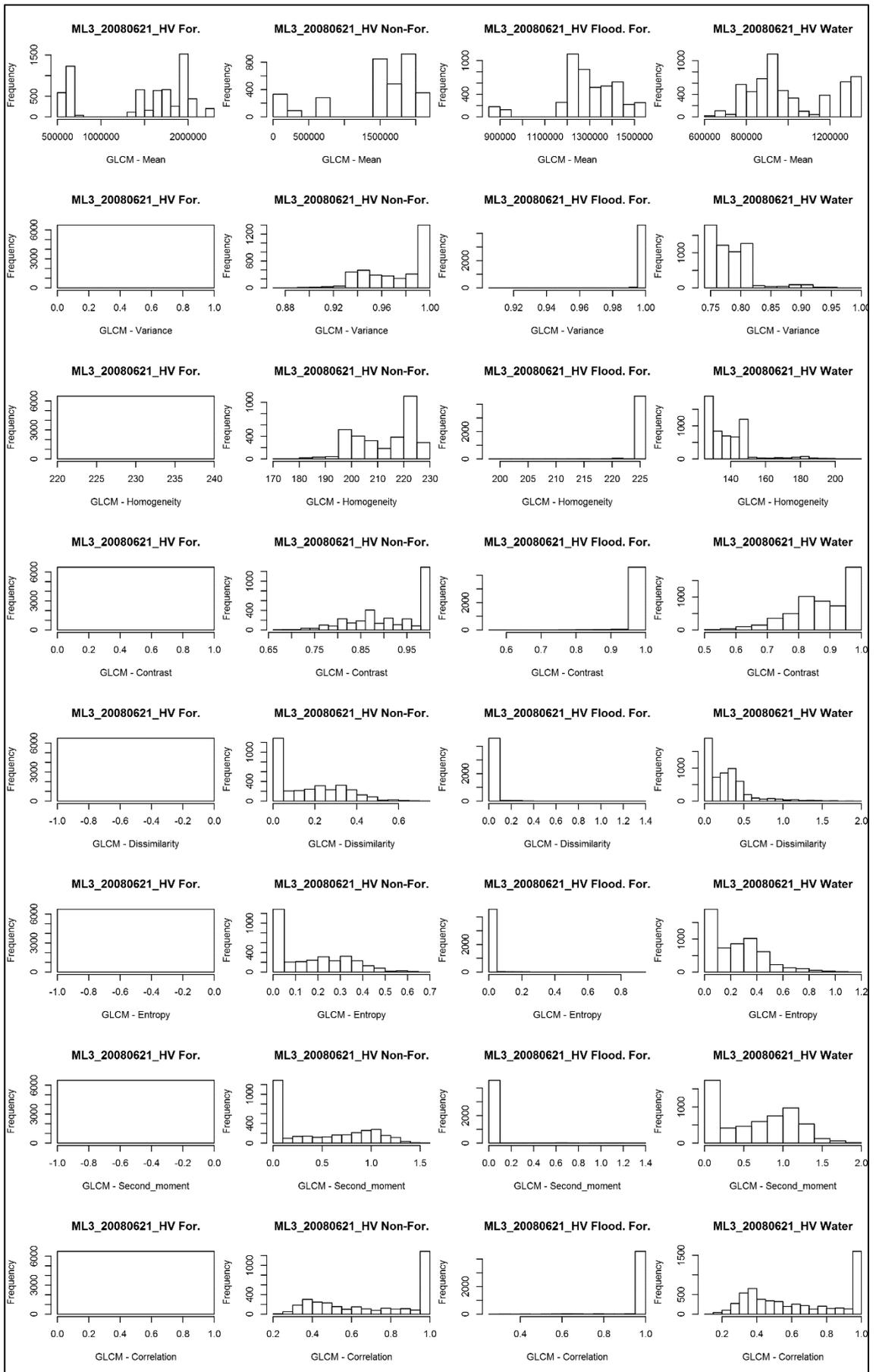


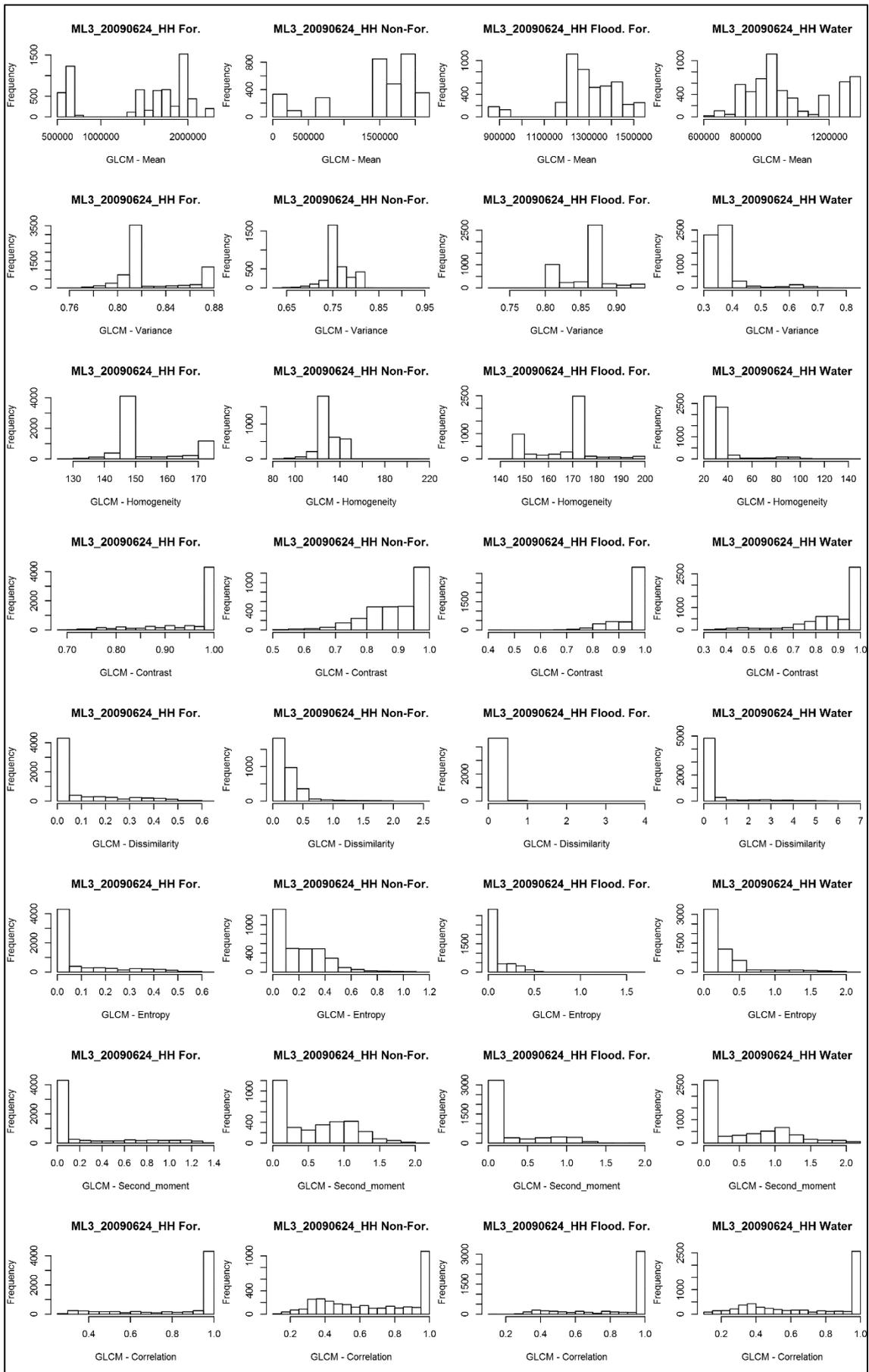


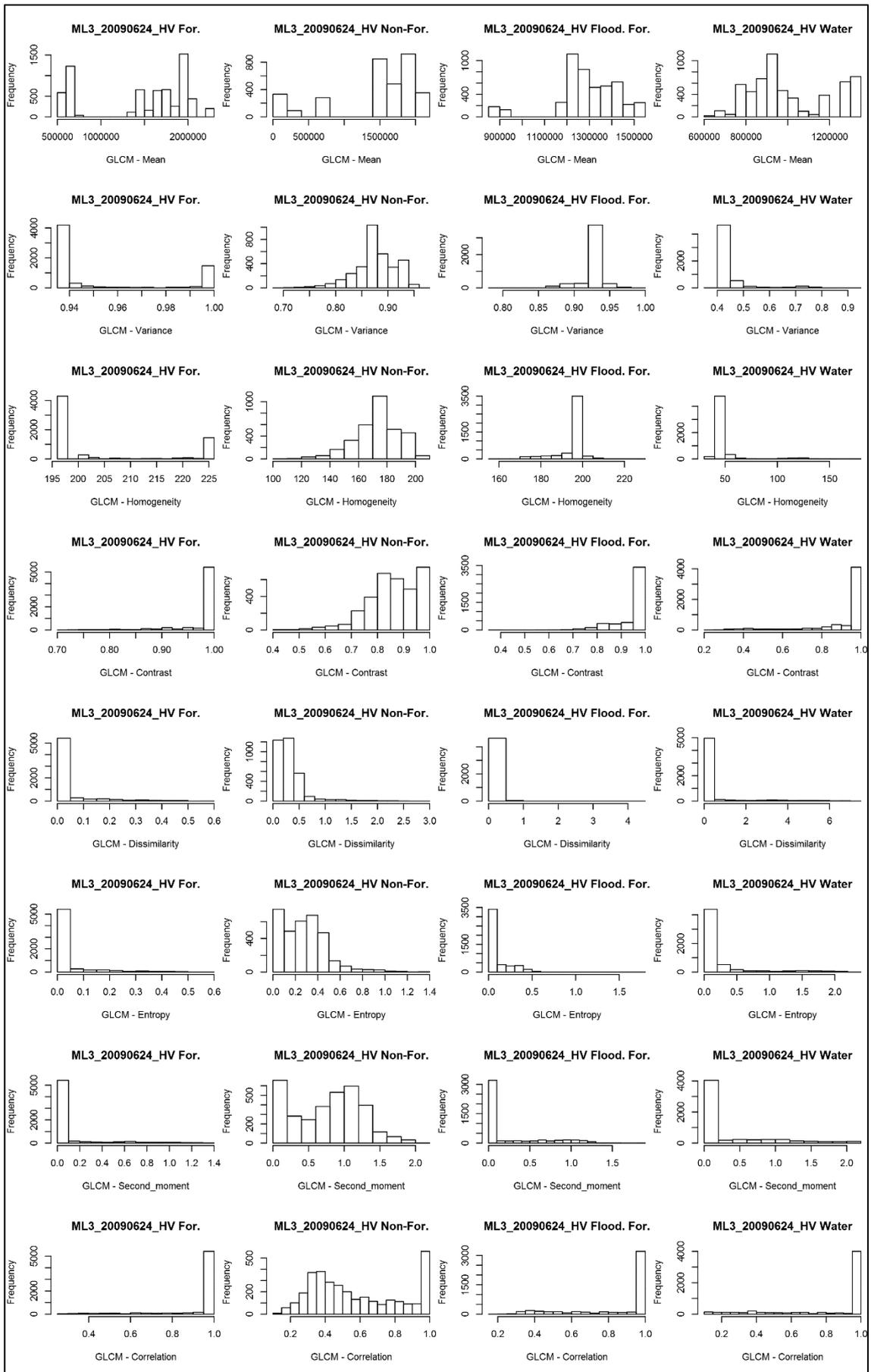


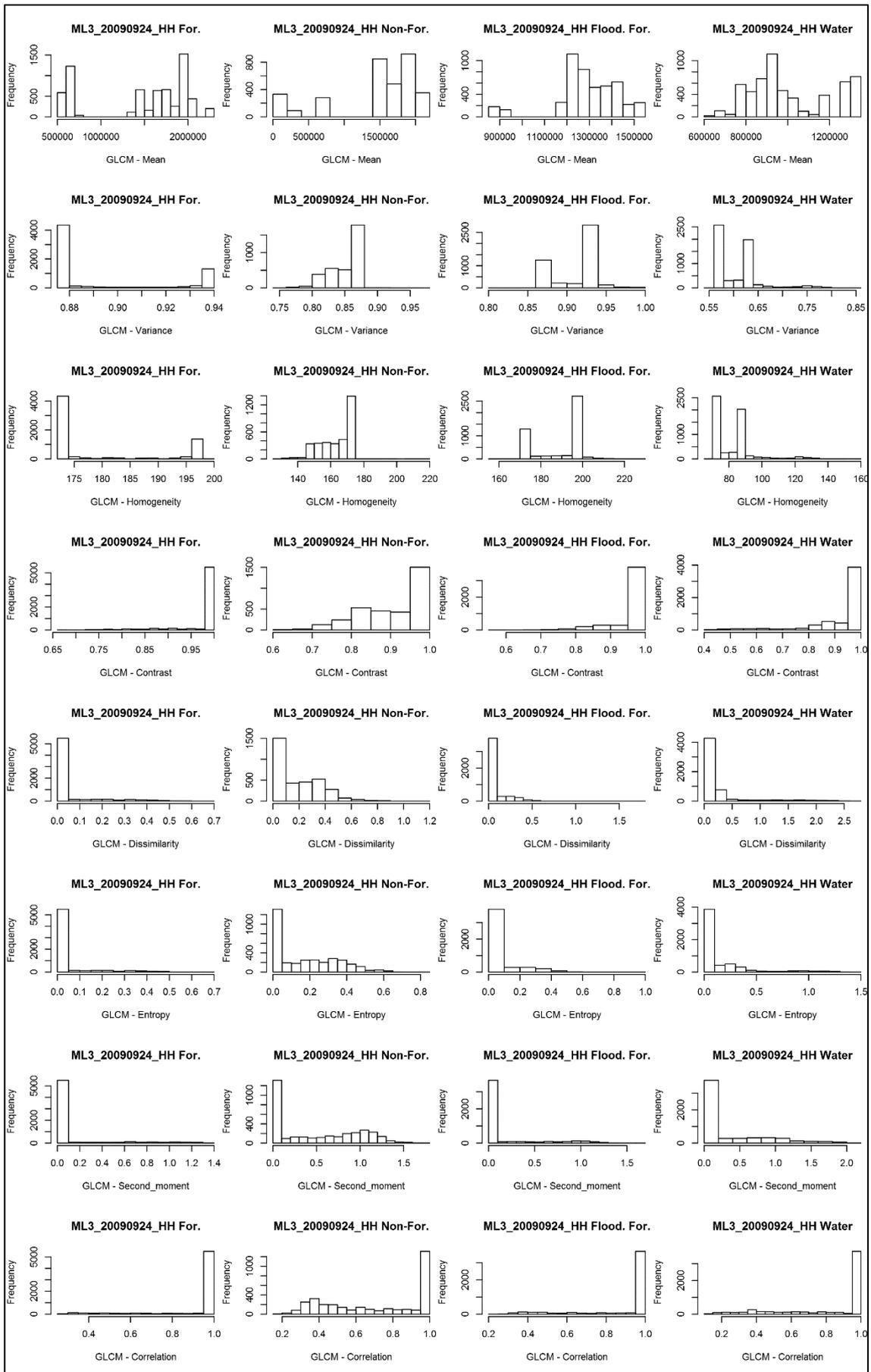


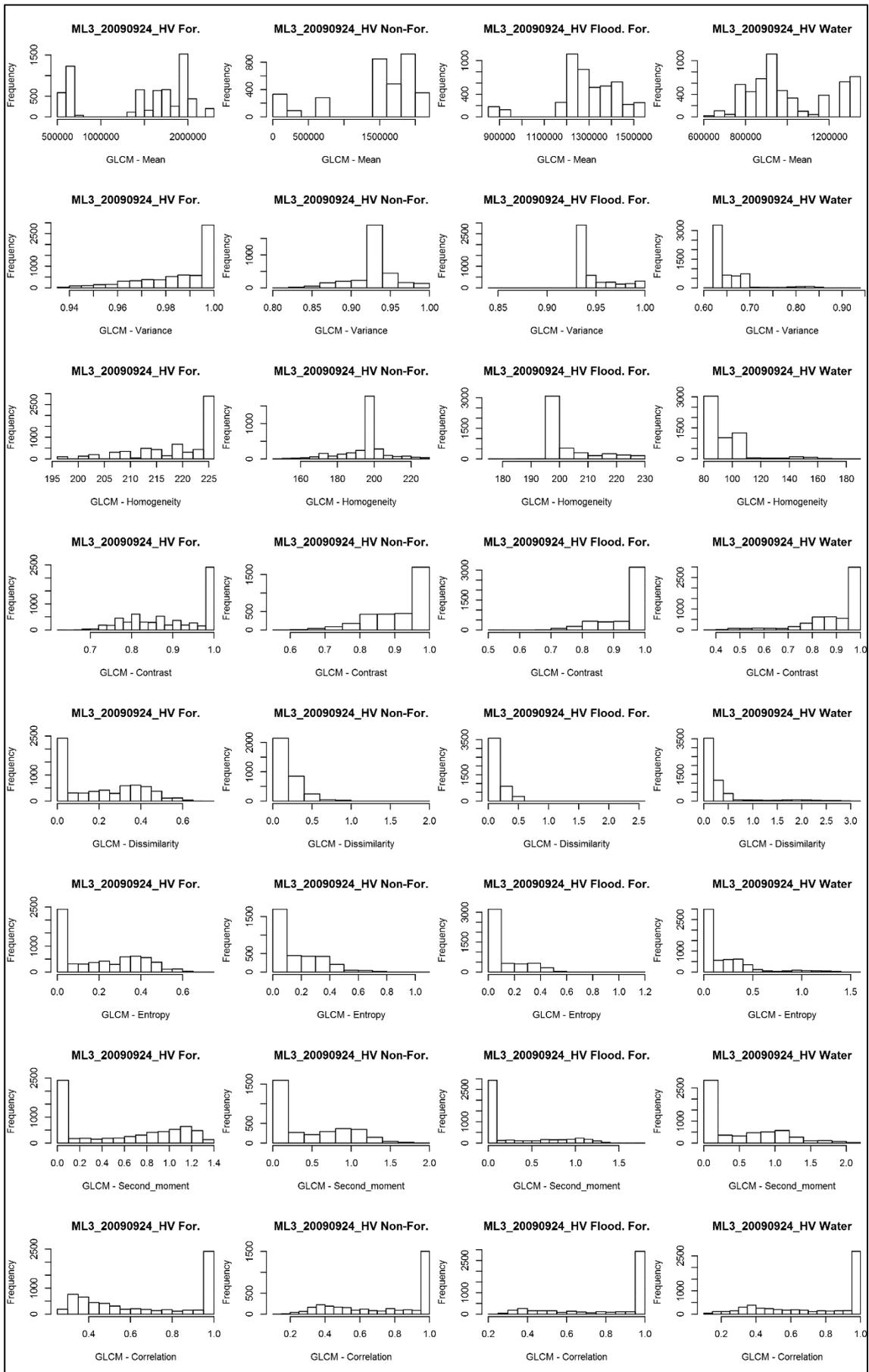


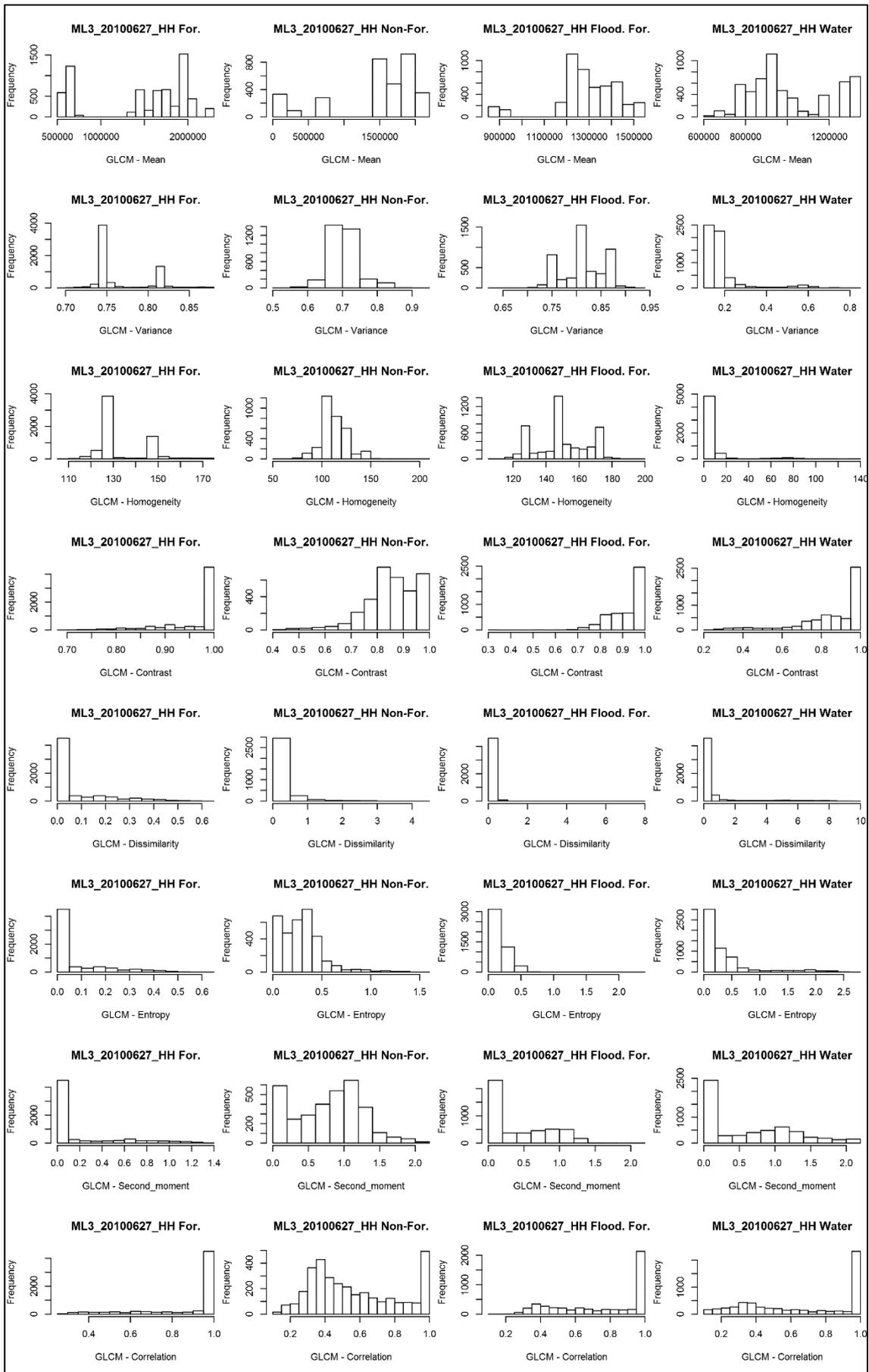


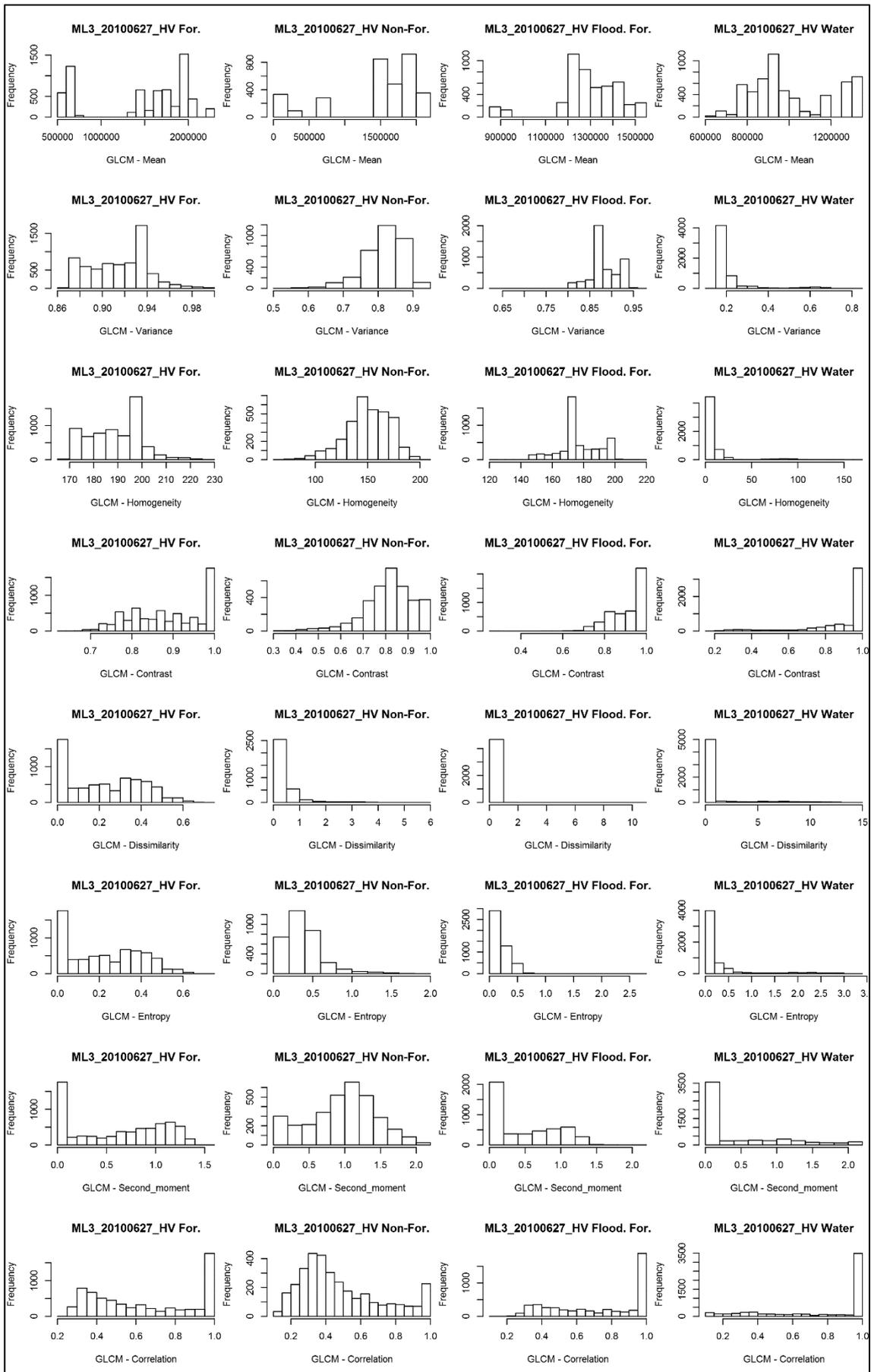


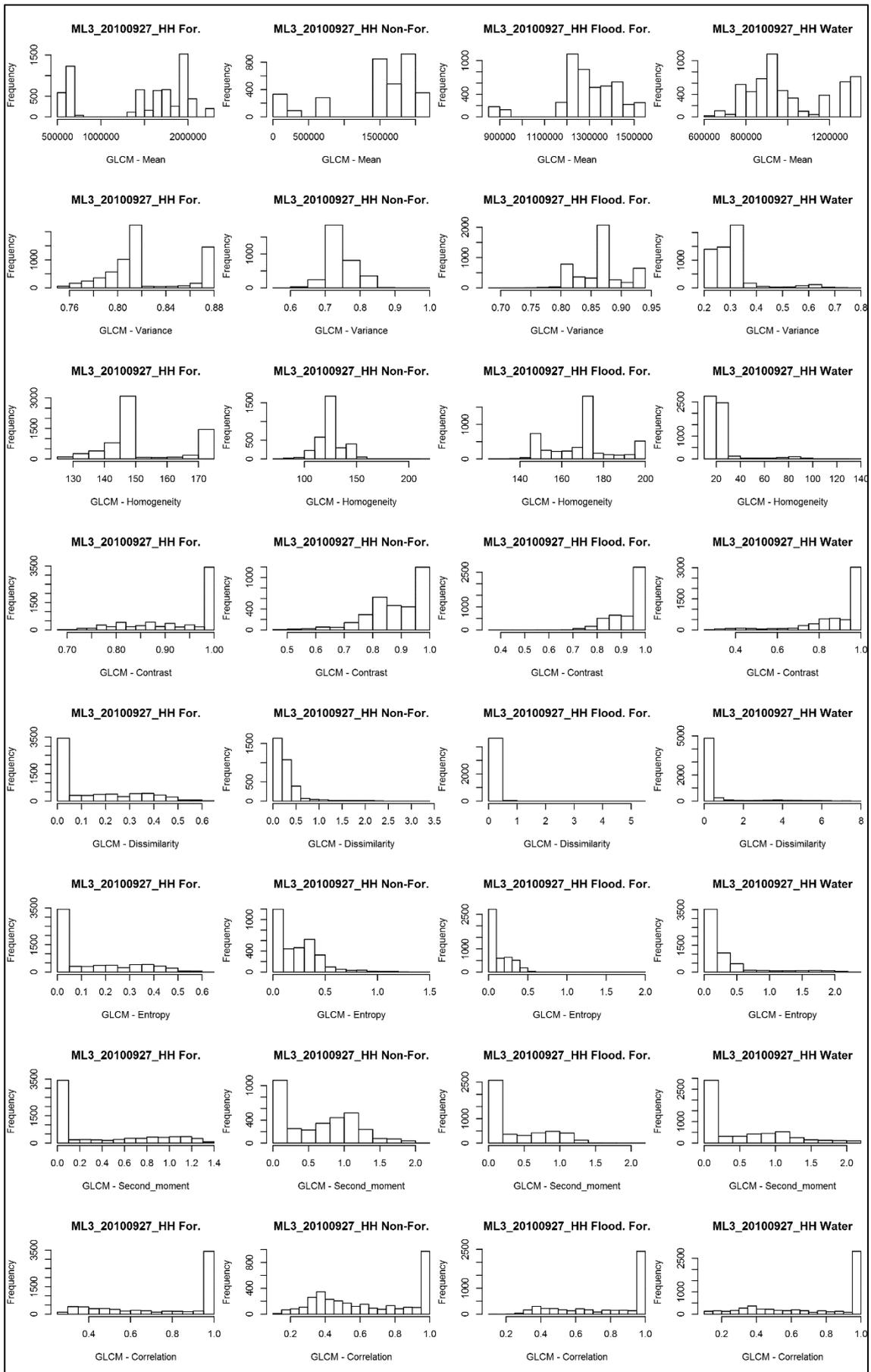


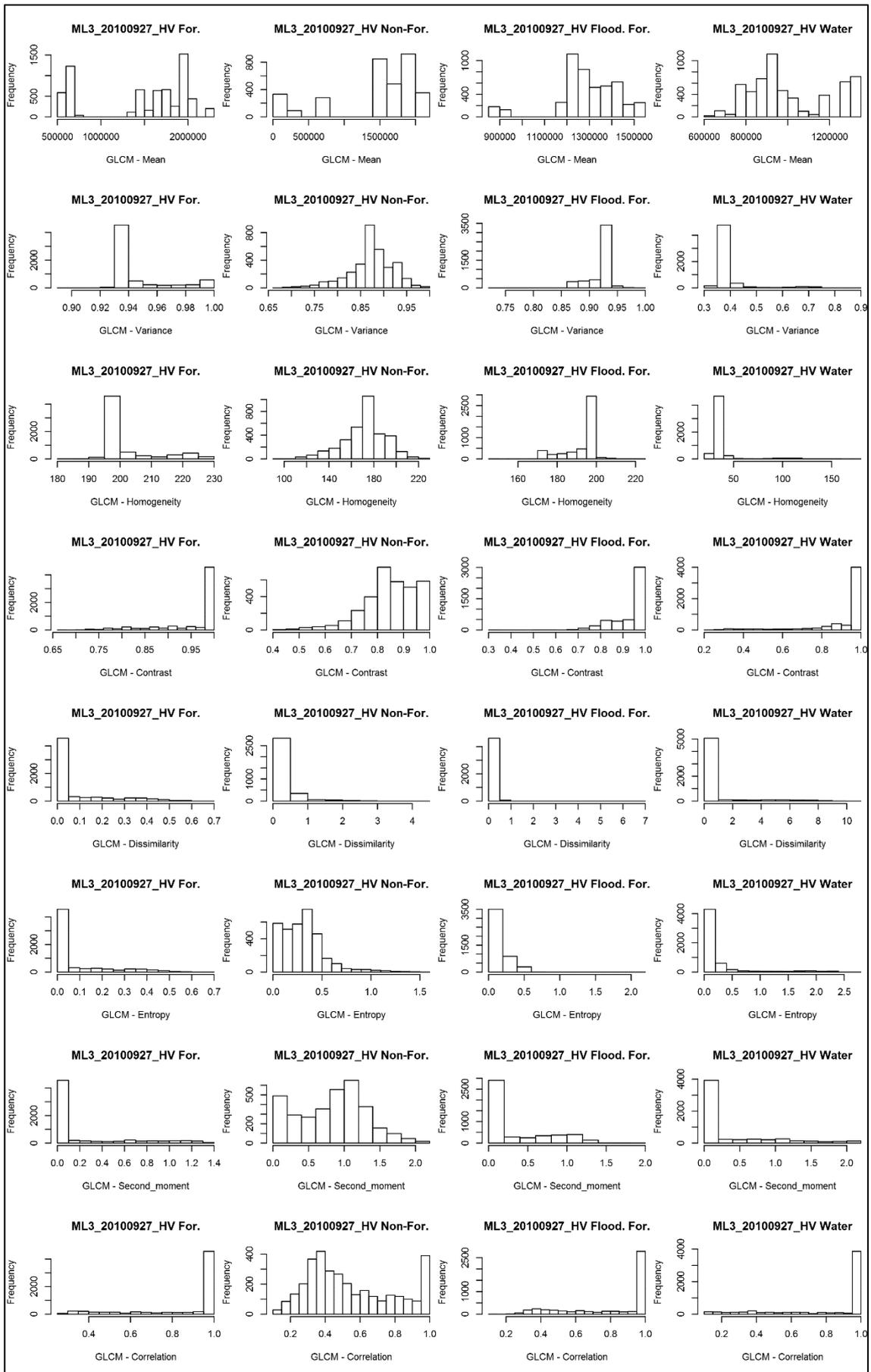












B2 Filtered and Unfiltered Scatterplot comparisons

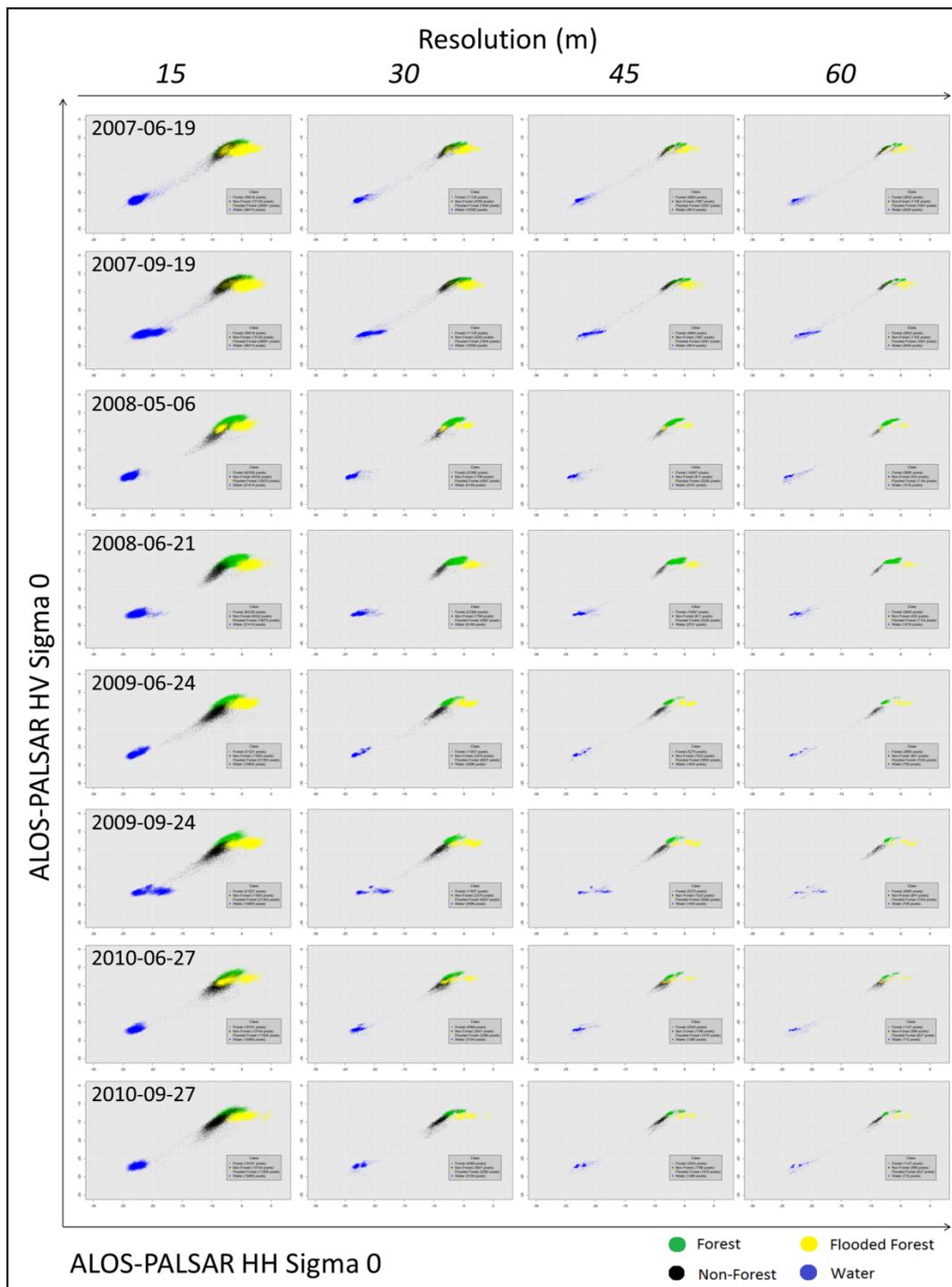


Figure B2.0.1: Classified scatterplots from filtered ALOS-PALSAR HV/HH time series from 2007-10 at decreasing spatial resolutions from left to right

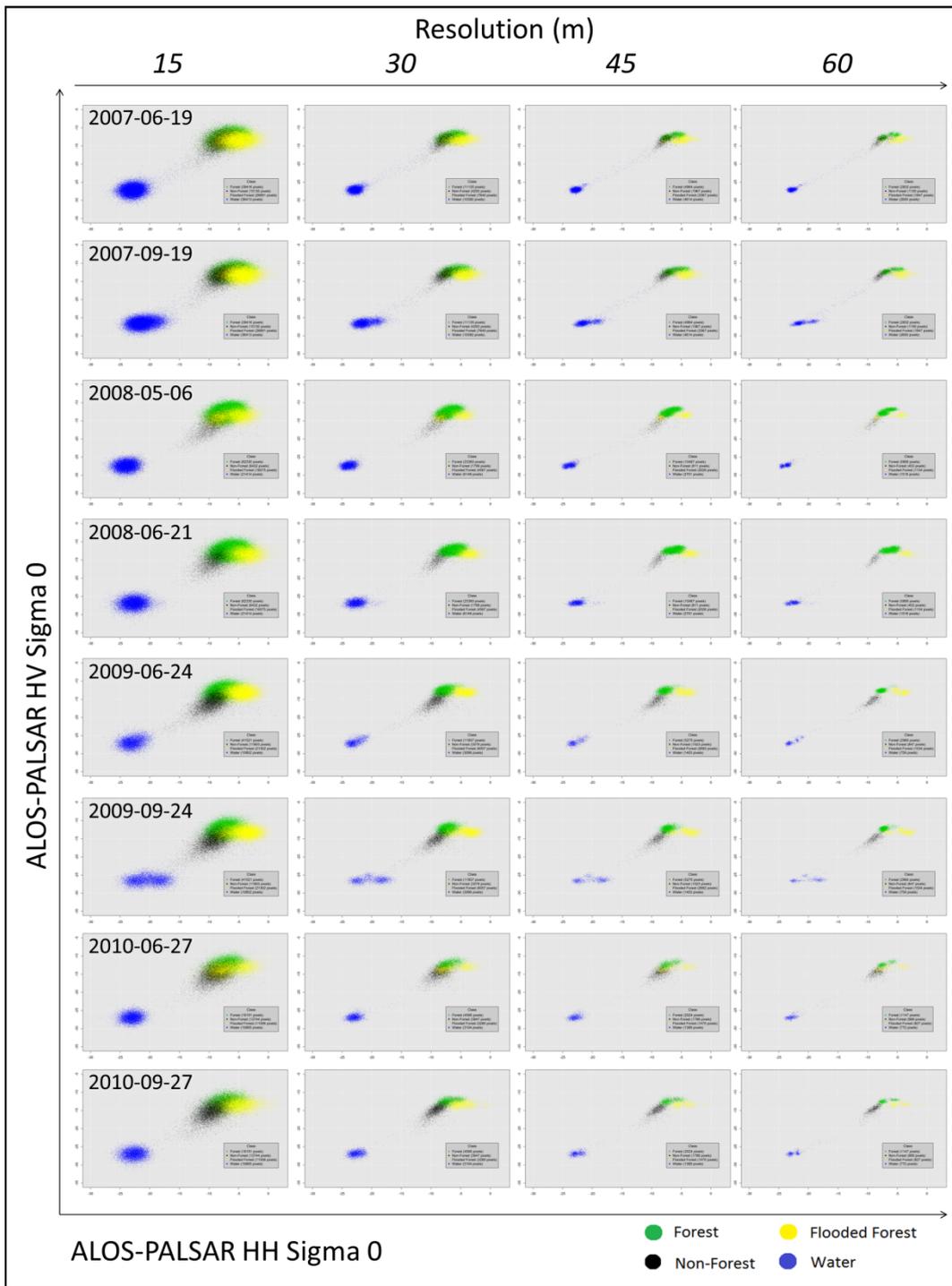
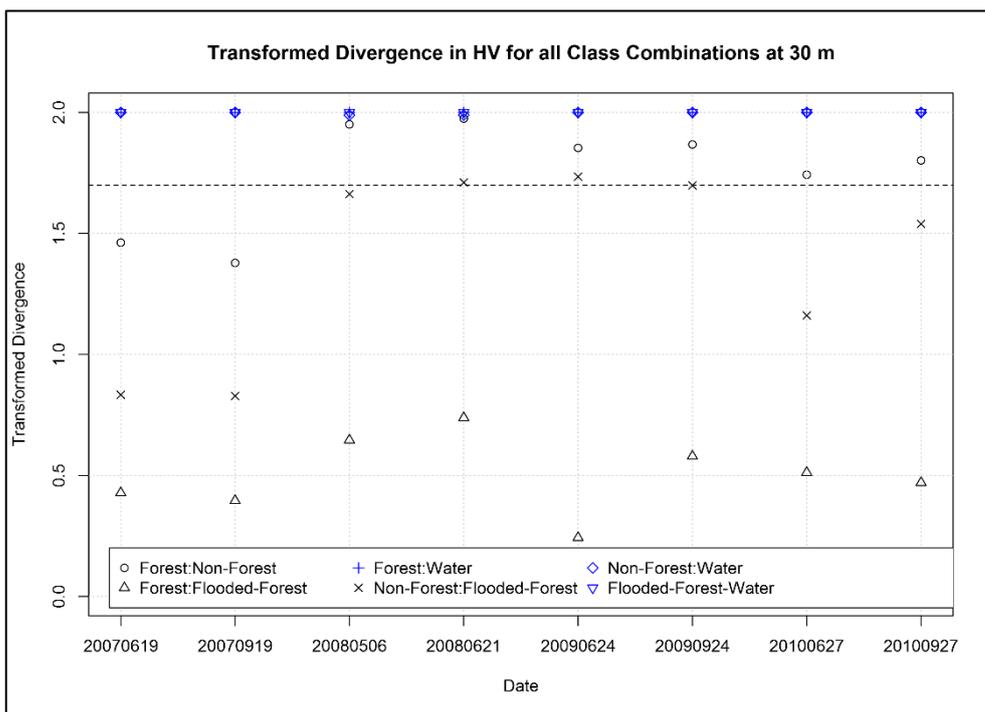
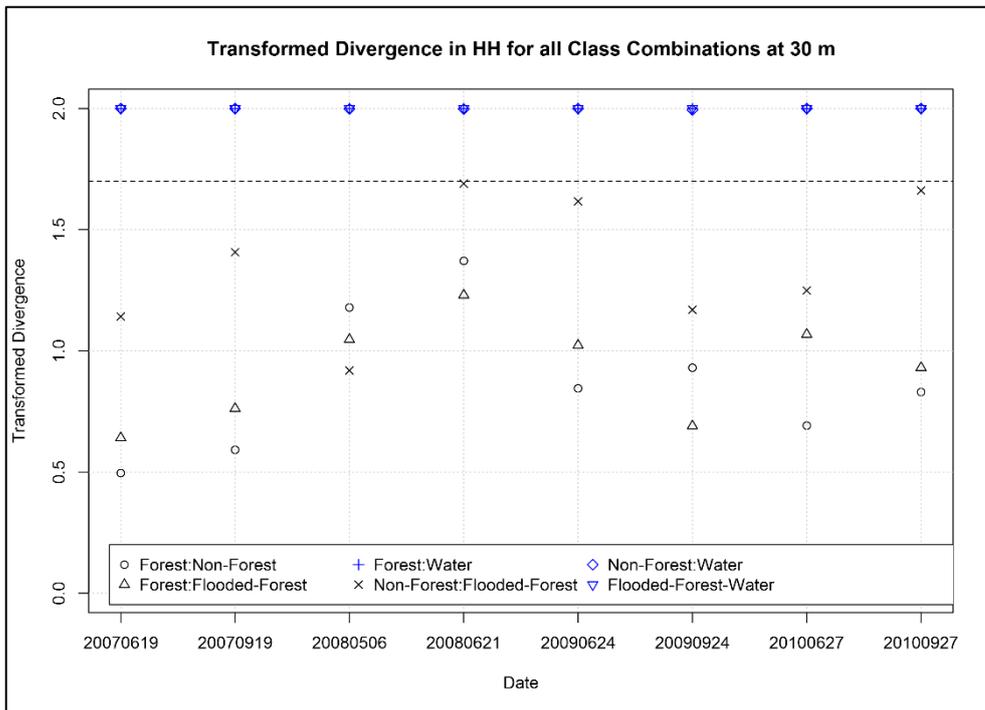
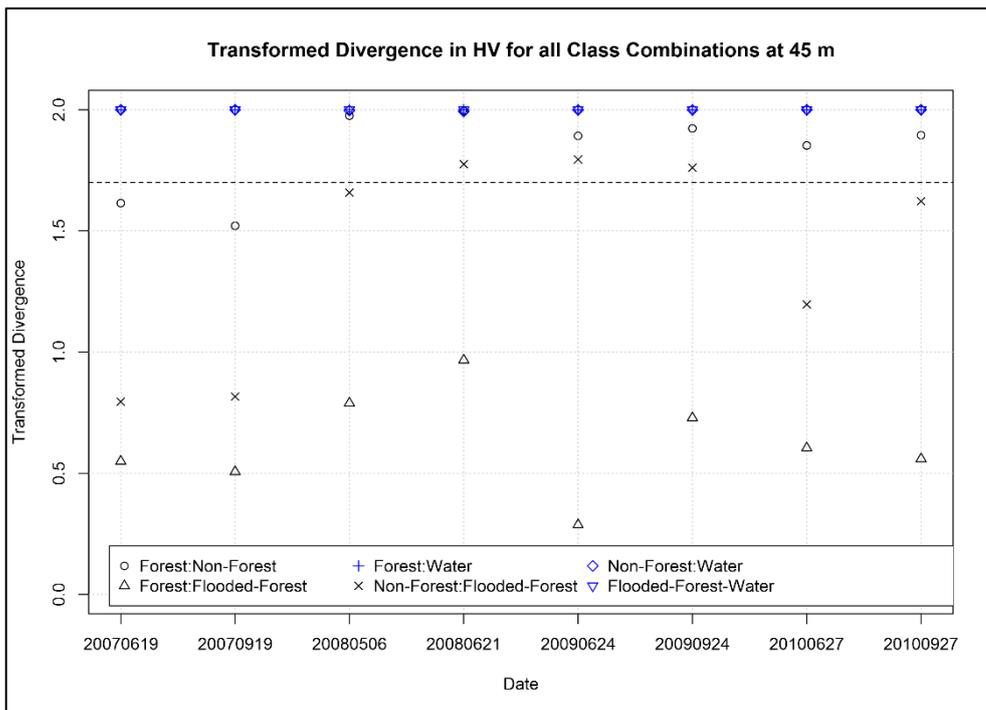
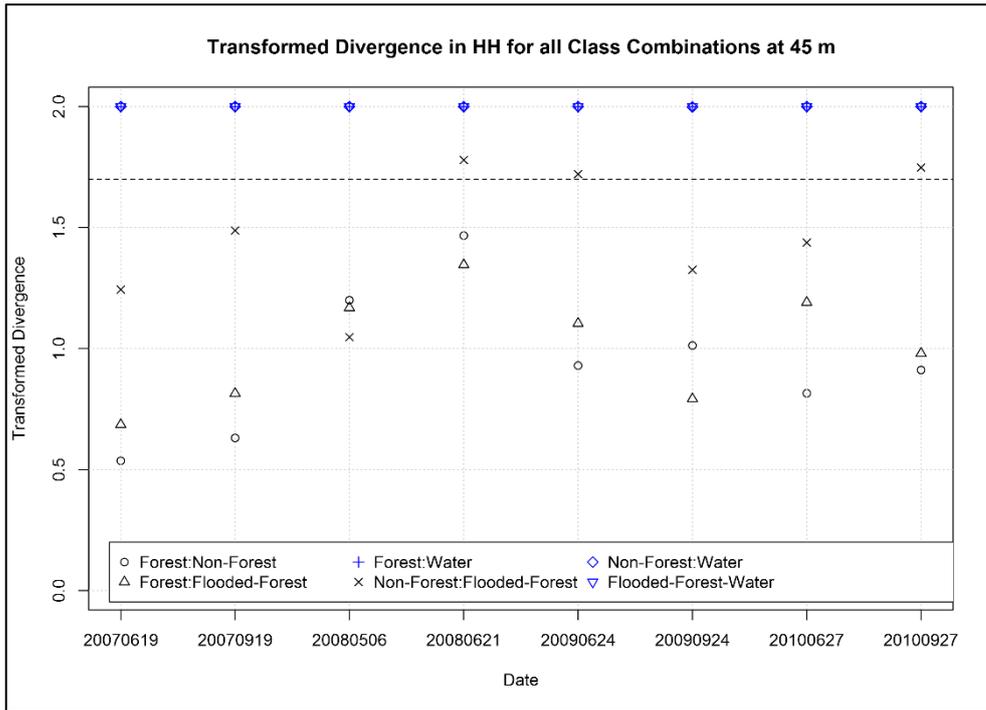
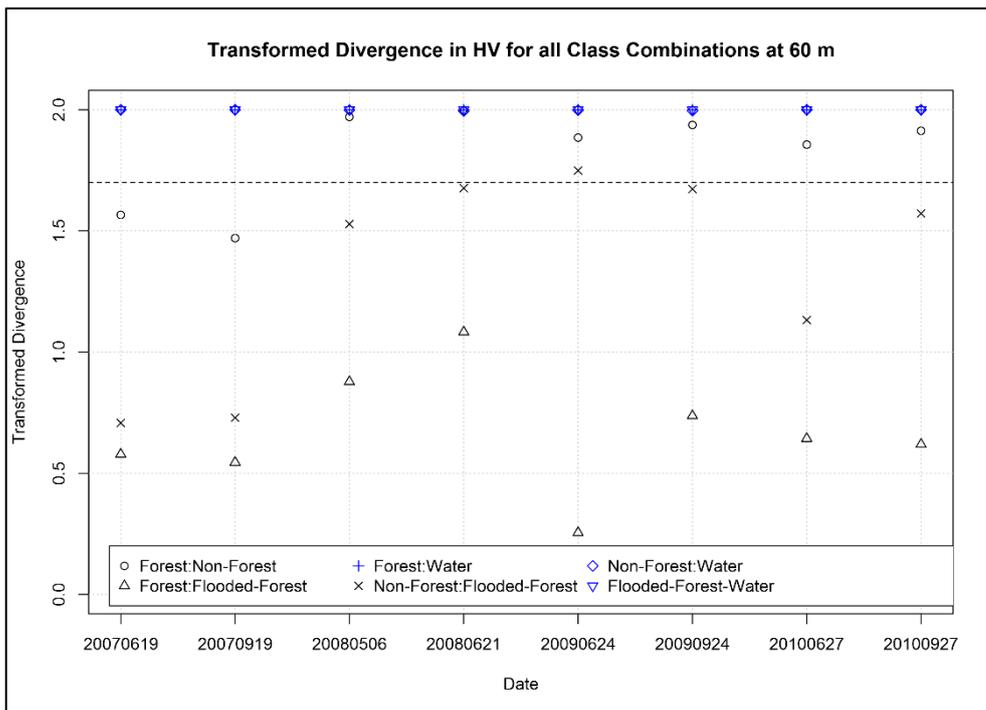
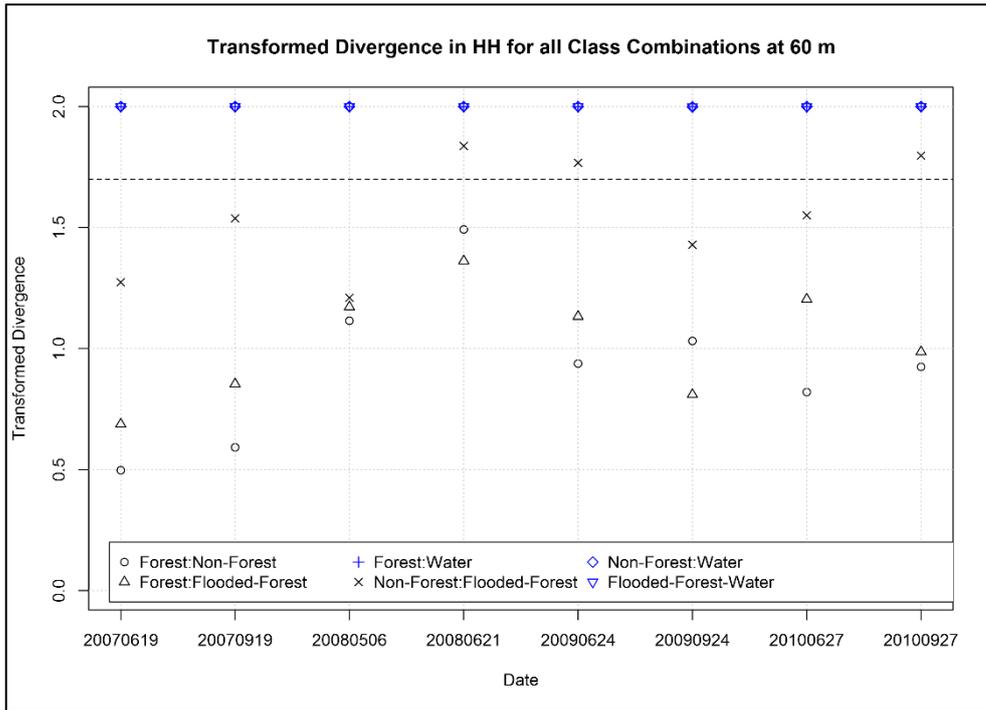


Figure B2.0.2: Classified scatterplots from unfiltered ALOS-PALSAR HV/HH time series from 2007-10 at decreasing spatial resolutions from left to right

B3 Inter Class Transform Divergence in HH and HV channels







B4 Scale differences in forest classification

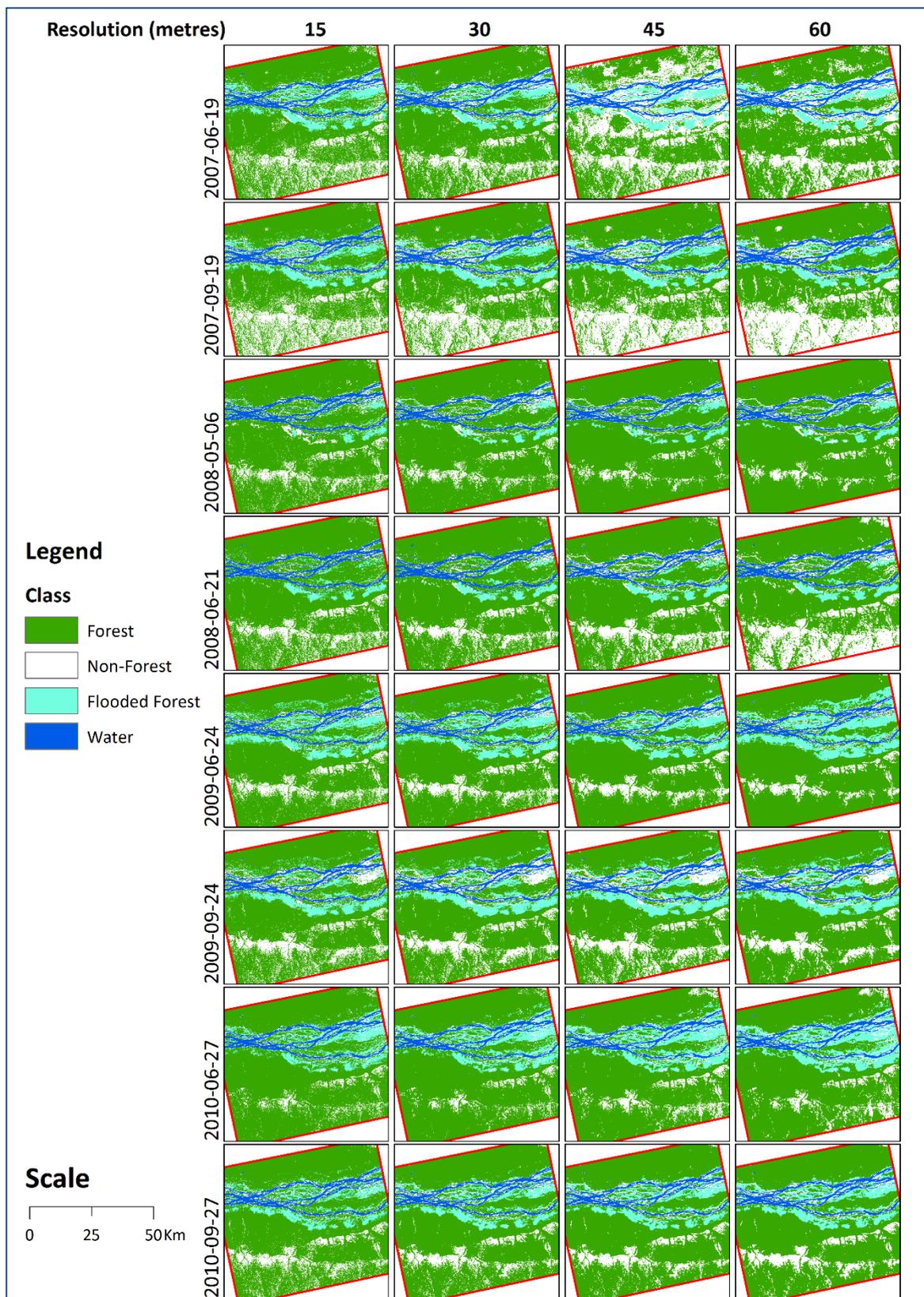


Figure B4.1: Comparison of classifications at 15, 30, 45 and 60m resolutions

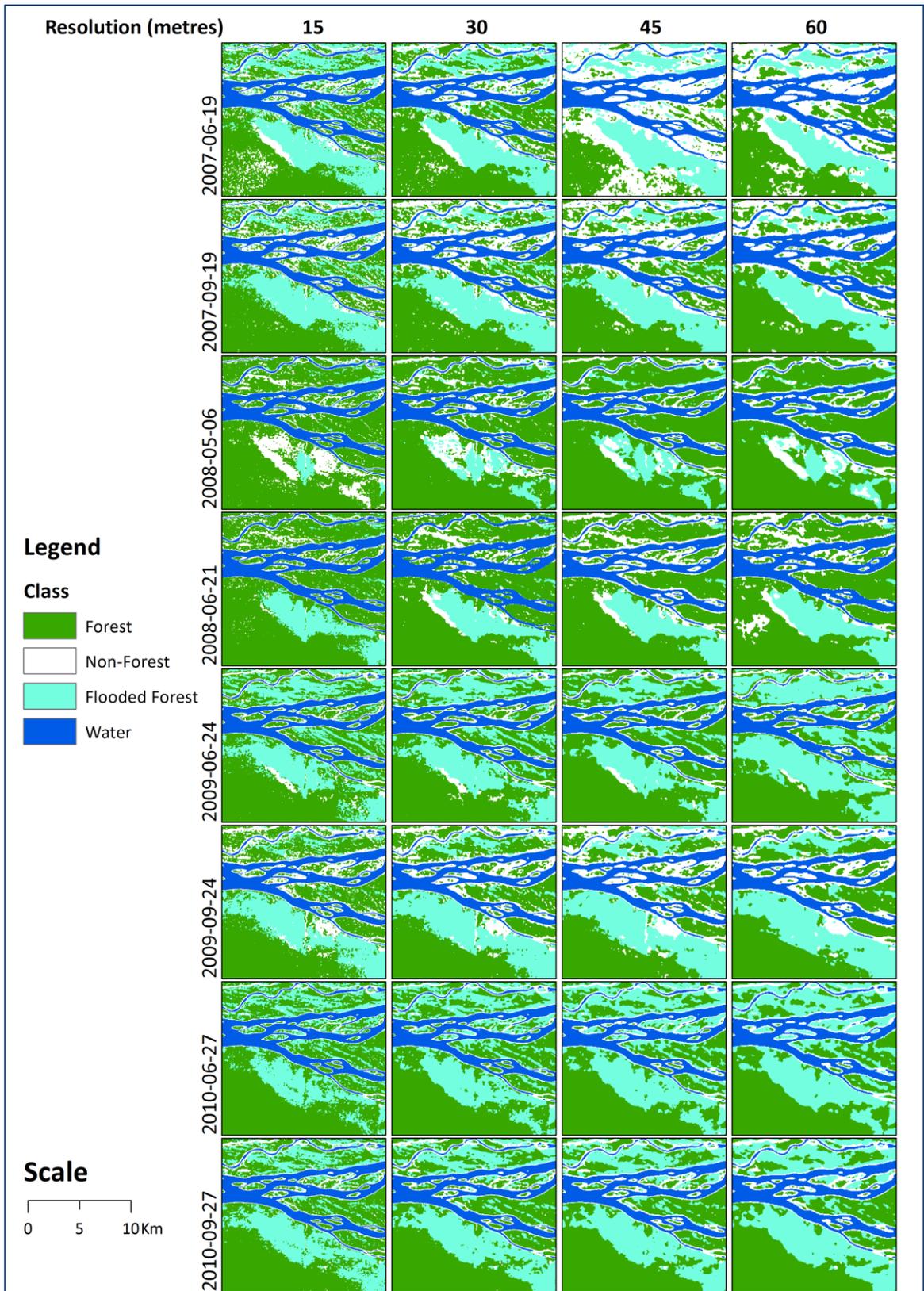


Figure B4.2: Comparison of classifications - flooded forest detail

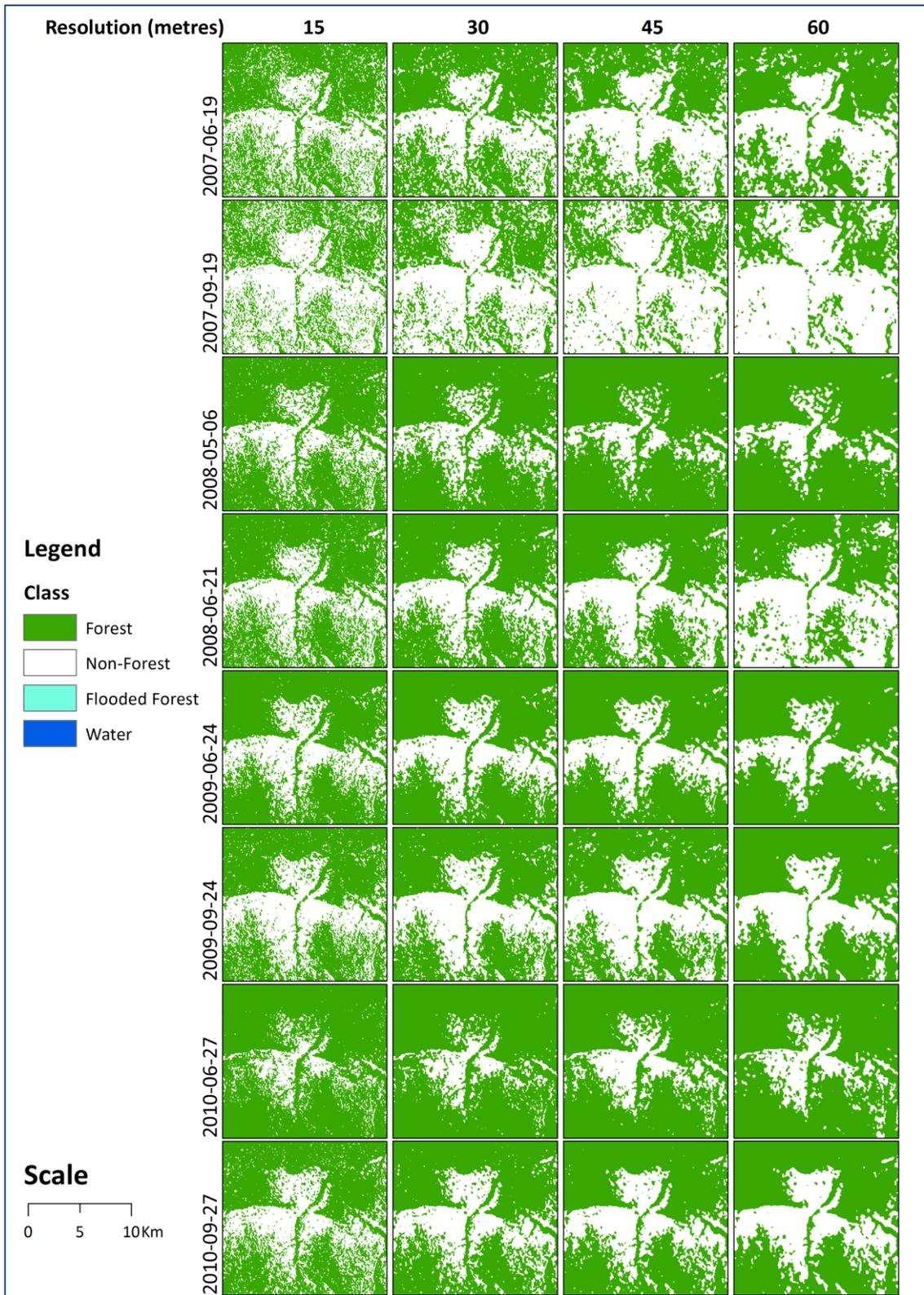


Figure B4.3: Comparison of classifications – non-forest detail

B5 Confusion Matrices from RF classifications

Table B1: Tables of Confusion Matrices from RF classifications at all Resolutions, including Class Mapping Accuracy

Confusion Matrices - 15m Resolution							
Date	2007-06-19						
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC	
Forest	37437	1338	25	0	38800	0.9071238	
Non_Forest	412	10177	0	0	10589	0.8330878	
Flooded_Forest	2058	55	9916	0	12029	0.8226315	
Water	0	234	0	12570	12804	0.9817245	
TOTAL	39907	11804	9941	12570	74222	0.9444639	
Date	2007-09-19						
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC	
Forest	36277	2358	165	0	38800	0.8867297	
Non_Forest	241	10348	0	0	10589	0.7808633	
Flooded_Forest	1870	134	10025	0	12029	0.8221256	
Water	0	171	0	12633	12804	0.9866448	
TOTAL	38388	13011	10190	12633	74222	0.9334564	
Date	2008-05-06						
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC	
Forest	17158	41	0	0	17199	0.8911858	
Non_Forest	75	8570	0	18	8663	0.9278909	
Flooded_Forest	1979	53	7750	0	9782	0.7922715	
Water	0	479	0	9280	9759	0.9491664	
TOTAL	19212	9143	7750	9298	45403	0.9417439	
Date	2008-06-21						
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC	
Forest	17033	166	0	0	17199	0.831527	
Non_Forest	137	8474	0	52	8663	0.9287593	
Flooded_Forest	3148	0	6634	0	9782	0.6781844	
Water	0	295	0	9464	9759	0.9646315	
TOTAL	20318	8935	6634	9516	45403	0.9163491	
Date	2009-06-24						
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC	
Forest	13261	4	0	0	13265	0.8873202	
Non_Forest	104	8714	0	0	8818	0.8659445	
Flooded_Forest	1576	254	10877	0	12707	0.8559849	
Water	0	987	0	10645	11632	0.9151479	
TOTAL	14941	9959	10877	10645	46422	0.9369911	
Date	2009-09-24						
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC	
Forest	13226	39	0	0	13265	0.7749912	
Non_Forest	406	8412	0	0	8818	0.7597543	
Flooded_Forest	3395	1524	7788	0	12707	0.6128905	
Water	0	691	0	10941	11632	0.9405949	
TOTAL	17027	10666	7788	10941	46422	0.8695662	
Date	2010-06-27						
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC	
Forest	17431	138	0	0	17569	0.9451282	
Non_Forest	533	9977	0	0	10510	0.8767135	
Flooded_Forest	341	189	9753	0	10283	0.9484586	
Water	0	543	0	16656	17199	0.9684284	
TOTAL	18305	10847	9753	16656	55561	0.9686111	
Date	2010-09-27						
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC	
Forest	16861	672	36	0	17569	0.9015613	
Non_Forest	110	10400	0	0	10510	0.9023861	
Flooded_Forest	1023	28	9232	0	10283	0.8946603	
Water	0	315	0	16884	17199	0.981685	
TOTAL	17994	11415	9268	16884	55561	0.9606919	

Confusion Matrices - 30m Resolution						
Date	2007-06-19					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	10918	66	0	0	10984	0.9075644
Non_Forest	146	2873	0	0	3019	0.8861814
Flooded_Forest	900	12	2485	0	3397	0.7315278
Water	0	145	0	3477	3622	0.9599669
TOTAL	11964	3096	2485	3477	21022	0.9396347
Date	2007-09-19					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	10184	759	41	0	10984	0.8908328
Non_Forest	91	2928	0	0	3019	0.7351243
Flooded_Forest	357	81	2959	0	3397	0.8606748
Water	0	124	0	3498	3622	0.9657648
TOTAL	10632	3892	3000	3498	21022	0.9308819
Date	2008-05-06					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	4920	0	0	0	4920	0.8860076
Non_Forest	15	2443	0	0	2458	0.7873026
Flooded_Forest	618	14	2139	0	2771	0.7719235
Water	0	631	0	2122	2753	0.7707955
TOTAL	5553	3088	2139	2122	12902	0.9009456
Date	2008-06-21					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	4904	16	0	0	4920	0.8555478
Non_Forest	65	2290	0	103	2458	0.8817867
Flooded_Forest	747	73	1951	0	2771	0.704078
Water	0	50	0	2703	2753	0.9464286
TOTAL	5716	2429	1951	2806	12902	0.9183072
Date	2009-06-24					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	3740	0	0	0	3740	0.9007707
Non_Forest	49	2449	0	0	2498	0.8318614
Flooded_Forest	363	36	3264	0	3663	0.8910729
Water	0	410	0	2909	3319	0.8764688
TOTAL	4152	2895	3264	2909	13220	0.9350983
Date	2009-09-24					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	3735	5	0	0	3740	0.805304
Non_Forest	109	2389	0	0	2498	0.7187124
Flooded_Forest	789	437	2437	0	3663	0.6653017
Water	0	384	0	2935	3319	0.8843025
TOTAL	4633	3215	2437	2935	13220	0.8695915
Date	2010-06-27					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	4905	37	0	0	4942	0.9613877
Non_Forest	150	2827	0	0	2977	0.842623
Flooded_Forest	10	130	2766	0	2906	0.9518238
Water	0	211	0	4702	4913	0.9570527
TOTAL	5065	3205	2766	4702	15738	0.9658152
Date	2010-09-27					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	4889	53	0	0	4942	0.9685024
Non_Forest	33	2944	0	0	2977	0.9078014
Flooded_Forest	73	6	2827	0	2906	0.9728149
Water	0	207	0	4706	4913	0.9578669
TOTAL	4995	3210	2827	4706	15738	0.9763629

Confusion Matrices - 45m Resolution						
Date	2007-06-19					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	1967	2937	0	0	4904	0.3936362
Non_Forest	37	1306	0	0	1343	0.2916481
Flooded_Forest	56	162	1301	0	1519	0.8564845
Water	0	36	0	1577	1613	0.9776813
TOTAL	2060	4441	1301	1577	9379	0.6558268
Date	2007-09-19					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	4229	673	2	0	4904	0.829541
Non_Forest	25	1311	0	7	1343	0.6239886
Flooded_Forest	169	67	1283	0	1519	0.843524
Water	0	18	0	1595	1613	0.9845679
TOTAL	4423	2069	1285	1602	9379	0.8975371
Date	2008-05-06					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	2179	0	0	0	2179	0.8912065
Non_Forest	7	1090	0	0	1097	0.7725018
Flooded_Forest	259	5	973	0	1237	0.7865804
Water	0	309	0	924	1233	0.7493917
TOTAL	2445	1404	973	924	5746	0.8990602
Date	2008-06-21					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	2177	2	0	0	2179	0.876762
Non_Forest	18	1079	0	0	1097	0.8070307
Flooded_Forest	286	43	908	0	1237	0.734034
Water	0	195	0	1038	1233	0.8418491
TOTAL	2481	1319	908	1038	5746	0.9053254
Date	2009-06-24					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	1672	0	0	0	1672	0.8393574
Non_Forest	73	1039	0	0	1112	0.7889142
Flooded_Forest	247	11	1358	0	1616	0.8403465
Water	0	194	0	1273	1467	0.8677573
TOTAL	1992	1244	1358	1273	5867	0.9105164
Date	2009-09-24					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	1672	0	0	0	1672	0.8667703
Non_Forest	42	1070	0	0	1112	0.6876607
Flooded_Forest	215	263	1138	0	1616	0.7042079
Water	0	181	0	1286	1467	0.876619
TOTAL	1929	1514	1138	1286	5867	0.8805182
Date	2010-06-27					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	2166	55	0	0	2221	0.9613848
Non_Forest	32	1297	0	0	1329	0.8600796
Flooded_Forest	0	32	1261	0	1293	0.9752514
Water	0	92	0	2081	2173	0.9576622
TOTAL	2198	1476	1261	2081	7016	0.9699259
Date	2010-09-27					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	2199	22	0	0	2221	0.9691494
Non_Forest	7	1322	0	0	1329	0.9123533
Flooded_Forest	41	4	1248	0	1293	0.9651972
Water	0	94	0	2079	2173	0.9567418
TOTAL	2247	1442	1248	2079	7016	0.9760547

Confusion Matrices - 60m Resolution						
Date	2007-06-19					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	2197	583	0	0	2780	0.7555021
Non_Forest	23	721	0	0	744	0.4914792
Flooded_Forest	105	97	651	0	853	0.7631887
Water	0	43	0	861	904	0.9524336
TOTAL	2325	1444	651	861	5281	0.8388563
Date	2007-09-19					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	2310	470	0	0	2780	0.7736102
Non_Forest	12	710	0	22	744	0.5508146
Flooded_Forest	194	57	602	0	853	0.7057444
Water	0	18	0	886	904	0.9568035
TOTAL	2516	1255	602	908	5281	0.8536262
Date	2008-05-06					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	1246	0	0	0	1246	0.9055233
Non_Forest	9	603	0	0	612	0.7398773
Flooded_Forest	121	16	556	0	693	0.8023088
Water	0	187	0	494	681	0.7254038
TOTAL	1376	806	556	494	3232	0.8969678
Date	2008-06-21					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	1151	95	0	0	1246	0.8180526
Non_Forest	25	582	0	5	612	0.6360656
Flooded_Forest	136	29	528	0	693	0.7619048
Water	0	179	0	502	681	0.7317784
TOTAL	1312	885	528	507	3232	0.8548886
Date	2009-06-24					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	942	0	0	0	942	0.8971429
Non_Forest	38	580	0	4	622	0.7837838
Flooded_Forest	70	5	848	0	923	0.9187432
Water	0	113	0	692	805	0.855377
TOTAL	1050	698	848	696	3292	0.9301337
Date	2009-09-24					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	942	0	0	0	942	0.812069
Non_Forest	34	582	0	6	622	0.6928571
Flooded_Forest	184	111	628	0	923	0.68039
Water	0	107	0	698	805	0.8606658
TOTAL	1160	800	628	704	3292	0.8657351
Date	2010-06-27					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	1207	29	0	0	1236	0.9511426
Non_Forest	33	719	0	0	752	0.8321759
Flooded_Forest	0	29	701	0	730	0.960274
Water	0	54	0	1142	1196	0.9548495
TOTAL	1240	831	701	1142	3914	0.9629535
Date	2010-09-27					
Class	Forest	Non_Forest	Flooded_Forest	Water	TOTAL	MAPPINGACC
Forest	1219	17	0	0	1236	0.9720893
Non_Forest	0	752	0	0	752	0.8973747
Flooded_Forest	18	12	700	0	730	0.9589041
Water	0	57	0	1139	1196	0.9523411
TOTAL	1237	838	700	1139	3914	0.9734287

Bibliography

- Arino, O., Gross, D., Ranera, F., Bourg, L., Leroy, M., Bicheron, P., Latham, J., Di Gregorio, A., Brockman, C., Witt, R., Defourny, P., Vancutsem, C., Herold, M., Sambale, J., Achard, F., Durieux, L., Plummer, S. & Weber, J.-L. (2007) GlobCover: ESA service for Global Land Cover from MERIS. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2007*. pp.2412–2415.
- Baccini, A., Laporte, N.T., Goetz, S.J., Sun, M. & Dong, H. (2008) A first map of tropical Africa's above-ground biomass derived from satellite imagery. *Environmental Research Letters*, 3 (4), p.45011.
- Ban, Y., Gong, P. & Giri, C. (2015) Global land cover mapping using Earth observation satellite data: Recent progresses and challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, pp.1–6.
- Betbeder, J., Gond, V., Frappart, F., Baghdadi, N.N., Briant, G. & Bartholomé, E. (2014) Mapping of Central Africa forested wetlands using Remote Sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7 (2), pp.531–542.
- Betts, R.A. (2000) Offset of the potential carbon sink from boreal forestation by decreases in surface albedo. *Nature*, 408 (6809), pp.187–90.
- Bontemps, S., Defourny, P., van Bogaert, E., Arino, O., Kalogirou, V. & Ramos Perez, J. (2010) Globcover 2009: products description and validation report [Internet]. Available from: <http://due.esrin.esa.int/globcover/LandCover2009/GLOBCOVER2009_Validation_Report_2.2.pdf> [Accessed 6 August 2012].
- Bontemps, S., Herold, M., Kooistra, L., van Groenestijn, A., Hartley, A., Arino, O., Moreau, I. & Defourny, P. (2012) Revisiting land cover observation to address the needs of the climate modeling community. *Biogeosciences*, 9 (6), pp.2145–2157.
- Breiman, L. (2001) Random forests. *Machine Learning*, 45, pp.5–32.

- Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-fernández, J., Llorens, P., Latron, J., Martin, C. & Bittelli, M. (2011) Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. *Remote Sensing of Environment*, 115 (12), pp.3390–3408.
- Bueso-Bello, J.-L., Martone, M., Gonzalez, C., Wecklich, C., Schulze, D., Rizzoli, P., Krieger, G. & Zink, M. (2017) TanDEM-X global forest/non-forest map generation. In: *2017 18th International Radar Symposium (IRS)*. IEEE, pp.1–9.
- Bwangoy, J.R.B., Hansen, M.C., Roy, D.P., De Grandi, G. & Justice, C.O. (2010) Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices. *Remote Sensing of Environment*, 114 (1), pp.73–86.
- Canada Centre for Remote Sensing (2007) Radar Tutorial [Internet]. Available from: <http://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/earth-sciences/files/pdf/resource/tutor/gsarcd/pdf/gsarcd_e.pdf> [Accessed 10 April 2012].
- Canadell, J.G., Le Quéré, C., Raupach, M.R., Field, C.B., Buitenhuis, E.T., Ciais, P., Conway, T.J., Gillett, N.P., Houghton, R.A. & Marland, G. (2007) Contributions to accelerating atmospheric CO₂ growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the National Academy of Sciences of the United States of America*, 104 (47), pp.18866–70.
- Canadell, J.G. & Raupach, M.R. (2008) Managing forests for climate change mitigation. *Science*, 320, pp.1456–7.
- Canadian Space Agency (2017) RADARSAT Constellation [Internet].
- Canadian Space Agency (2015) RADARSAT Systems: Satellite Characteristics [Internet]. Available from: <<http://www.asc-csa.gc.ca/eng/satellites/radarsat/radarsat-tableau.asp>> [Accessed 30 July 2017].
- Cao, M. & Woodward, F.I. (1998) Dynamic responses of terrestrial ecosystem carbon cycling to global climate change. *Nature*, 393, pp.249–252.

- Carreiras, J.M.B., Jones, J., Lucas, R.M. & Gabriel, C. (2014) Land Use and Land Cover Change Dynamics across the Brazilian Amazon: Insights from Extensive Time-Series Analysis of Remote Sensing Data. *PLoS ONE*, 9 (8), p.e104144.
- CBD (2010) Strategic Plan for Biodiversity 2011-2020: Provisional Technical Rationale, possible indicators and suggested milestones for the Aichi Biodiversity Targets. In: *Conference of the Parties to the Convention on Biological Diversity*. Nagoya, Japan, pp.1–20.
- Chambers, J.Q. & Roberts, D.A. (2014) Ecology: Drought in the Congo Basin. *Nature*, 509 (7498), pp.36–37.
- Chang, C.-C. & Lin, C.-J. (2011) LIBSVM : A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1--27, pp.1–39.
- Cloude, S.R. & Papathanassiou, K.P. (1998) Polarimetric SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 36 (5), pp.1551–1565.
- Cloude, S.R., Woodhouse, I.H. & Suárez, J.C. (2011) *Polarimetry and polarimetric interferometry for forestry applications: final report*.
- Cohen Liechti, T., Matos, J.P., Boillat, J. & Schleiss, A.J. (2012) Comparison and evaluation of satellite derived precipitation products for hydrological modeling of the Zambezi River Basin. *Hydrology and Earth System Sciences*, 16, pp.489–500.
- CONAE (2014) SAOCOM - Introducción [Internet]. Available from: <<http://www.conae.gov.ar/index.php/espanol/misiones-satelitales/saocom/objetivos>> [Accessed 6 October 2014].
- Congalton, R.G. & Green, K. (2009) *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. 2nd ed. Boca Raton, FL, CRC Press.
- Consortium Congo 2010 (2014) *ETAT DES LIEUX DE LA BIODIVERSITE DANS LA R.D. CONGO*. Kisangani, DRC.
- Cortes, C. & Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, 20, pp.273–297.

- CPC, N. (2017) NOAA CPC Morphing Technique (“CMORPH”) [Internet]. Available from: <http://www.cpc.ncep.noaa.gov/products/janowiak/cmorph_description.html> [Accessed 31 July 2017].
- Criminisi, A., Shotton, J. & Konukoglu, E. (2011) Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. *Microsoft Research technical report TR-2011-114*.
- Cumming, I.G. & Bennett, J.R. (1979) Digital processing of SEASAT SAR data. In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79*. Washington, D.C., pp.710–718.
- Dargie, G.C., Lewis, S.L., Lawson, I.T., Mitchard, E.T.A., Page, S.E., Bocko, Y.E. & Ifo, S.A. (2017) Age, extent and carbon storage of the central Congo Basin peatland complex. *Nature*, 542 (7639), pp.86–90.
- Dean, T.J., Bell, J.P. & Baty, A.J.. (1987) Soil Moisture Measurement by an Improved Capacitance Technique, Part 1. Sensor Design and Performance. *Journal of Hydrology*, 93, pp.67–78.
- DeFries, R.S., Hansen, M.C., Townshend, J.R.G., Janetos, A.C. & Loveland, T.R. (2000) A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, 6 (2), pp.247–254.
- DeFries, R.S., Houghton, R.A., Hansen, M.C., Field, C.B., Skole, D. & Townshend, J.R.G. (2002) Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (22), pp.14256–14261.
- DiMiceli, C.M., Carroll, M.L., Sohlberg, R.A., Huang, C., Hansen, M.C. & Townshend, J.R.G. (2011) Annual Global Automated MODIS Vegetation Continuous Fields (MOD44B) at 250 m Spatial Resolution for Data Years Beginning Day 65, 2000 - 2010, Collection 5: Percent Tree Cover, University of Maryland, College Park, MD, USA. [Internet]. Available from: <<http://glcf.umd.edu/data/vcf/>> [Accessed 29 August 2013].

- Dinku, T., Ceccato, P., Kopec, E.G., Lemma, M. & Connor, S.J. (2007) Validation of satellite rainfall products over East Africa's complex topography. *International Journal of Remote Sensing*, 28 (7), pp.1503–1526.
- Dong, J., Xiao, X., Sheldon, S., Biradar, C., Zhang, G., Duong, N.D., Hazarika, M., Wikantika, K., Takeuchi, W. & Moore, B. (2014) A 50-m forest cover map in Southeast Asia from ALOS/PALSAR and its application on forest fragmentation assessment. *PloS one*, 9 (1), p.e85801.
- Einzmann, K., Haarpaintner, J. & Larsen, Y. (2012) Forest monitoring in Congo Basin with combined use of SAR C- & L-band. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany*, pp.6573–6576.
- Englhart, S., Keuck, V. & Siegert, F. (2011) Aboveground biomass retrieval in tropical forests — The potential of combined X- and L-band SAR data use. *Remote Sensing of Environment*, 115 (5), pp.1260–1271.
- Englhart, S., Keuck, V. & Siegert, F. (2012) Modeling Aboveground Biomass in Tropical Forests Using Multi-Frequency SAR Data — A Comparison of Methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5 (1), pp.298–306.
- ESA (2017) Introducing Sentinel-2 [Internet]. Available from: <http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2/Introducing_Sentinel-2> [Accessed 14 August 2017].
- ESA (2014) Sentinel Online User Guides: Applications: Land monitoring [Internet]. Available from: <<https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/applications/land-monitoring>> [Accessed 16 August 2014].
- Escorihuela, M.J. & Quintana-seguí, P. (2016) Comparison of remote sensing and simulated soil moisture datasets in Mediterranean landscapes. *Remote Sensing of Environment*, 180, pp.99–114.
- European Space Agency (2012) Sentinel-1 Data Sheet: Facts and Figures [Internet]. Available from: <http://esamultimedia.esa.int/docs/S1-Data_Sheet.pdf>

[Accessed 31 August 2012].

- FAO (2001) *Global Forest Resources Assessment 2000 Main Report*. , pp.321–331.
- FAO (2010) *Global Forest Resources Assessment 2010* [Internet]. Available from: <<http://www.fao.org/docrep/013/i1757e/i1757e.pdf>>.
- FAO (2015a) *Global Forest Resources Assessment 2015. Country Report. Democratic People's Republic of Korea*. Rome, Italy.
- FAO (2015b) *Global Forest Resources Assessment 2015. Country Report. Mexico*. Rome, Italy.
- FAO & JRC (2012) *Global forest land-use change 1990-2005*, by E.J. Lindquist, R. D'Annunzio, A. Gerrand, K. MacDicken, F. Achard, R. Beuchle, A. Brink, H.D. Eva, P. Mayaux, J. San-Miguel-Ayanz & H-J. Stibig. FAO Forestry Paper No. 169. Food and Agriculture Organization of the United Nations and European Commission Joint Research Centre. Rome, Italy.
- Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D. & Amorim Fernández-Delgado, D. (2014) Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research*, 15, pp.3133–3181.
- Ferreira-Ferreira, J., Silva, T.S.F., Streher, A.S., Affonso, A.G., de Almeida Furtado, L.F., Forsberg, B.R., Valsecchi, J., Queiroz, H.L. & de Moraes Novo, E.M.L. (2015) r. *Wetlands Ecology and Management*, 23 (1), pp.41–59.
- Freeman, A. & Durden, S.L. (1998) A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 36 (3), pp.963–973.
- Frost, V.S., Stiles, J.A., Shanmugen, K.S. & Holtzman, J.C. (1982) A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-4 (2), pp.157–166.
- Gaston, G., Brown, S., Lorenzini, M. & Singh, K.D. (1998) State and change in carbon

- pools in the forests of Tropical Africa. *Global Change Biology*, 4, pp.97–114.
- Ghimire, B., Rogan, J., Galiano, V.R., Panday, P. & Neeti, N. (2012) An Evaluation of Bagging, Boosting, and Random Forests for Land-Cover Classification in Cape Cod, Massachusetts, USA. *GIScience & Remote Sensing*, 49 (5), pp.623–643.
- Gibbs, H.K., Brown, S., Niles, J.O. & Foley, J.A. (2007) Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, 2 (4), p.45023.
- Gislason, P.O., Benediktsson, J.A. & Sveinsson, J.R. (2006) Random Forests for land cover classification. *Pattern Recognition Letters*, 27, pp.294–300.
- Goodman, R.C., Phillips, O.L. & Baker, T.R. (2014) The importance of crown dimensions to improve tropical tree biomass estimates. *Ecological Applications*, 24 (4), pp.680–698.
- De Grandi, G., Bouvet, A., Lucas, R.M., Shimada, M., Monaco, S. & Rosenqvist, A. (2011) The K&C PALSAR mosaic of the African continent: processing issues and first thematic results. *IEEE Transactions on Geoscience and Remote Sensing*, 49 (10), pp.3593–3610.
- De Grandi, G. & Leysen, M. (1997) Radar reflectivity estimation using multiple SAR scenes of the same target: technique and applications. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 1997*. Singapore, pp.1047–1050 vol.2.
- De Grandi, G., Malingreau, J.P. & Leysen, M. (1999) The ERS-1 Central Africa mosaic: A new perspective in radar remote sensing for the global monitoring of vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, 37 (3), pp.1730–1746.
- De Grandi, G., Mayaux, P., Malingreau, J.P., Rosenqvist, A., Saatchi, S. & Simard, M. (2000) New perspectives on global ecosystems from wide-area radar mosaics: Flooded forest mapping in the tropics. *International Journal of Remote Sensing*, 21 (6–7), pp.1235–1249.

- De Grandi, G., Mayaux, P., Rauste, Y., Rosenqvist, A., Simard, M. & Saatchi, S.S. (2000) The Global Rain Forest Mapping Project JERS-1 radar mosaic of tropical Africa: development and product characterization aspects. *IEEE Transactions on Geoscience and Remote Sensing*, 38 (5), pp.2218–2233.
- Gross, D., Achard, F., Dubois, G., Brink, A. & Prins, H.H.T. (2017) Uncertainties in tree cover maps of Sub-Saharan Africa and their implications for measuring progress towards CBD Aichi Targets D. Rocchini & M. Disney eds. *Remote Sensing in Ecology and Conservation*.
- Guo, H.-D., Ding, Y., Liu, G., Zhang, D.-W., Fu, W.-X. & Zhang, L. (2014) Conceptual study of lunar-based SAR for global change monitoring. *Science China Earth Sciences*, 57 (8), pp.1771–1779.
- Hall-Beyer, M. (2017) Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *International Journal of Remote Sensing*, 38 (5), pp.1312–1338.
- Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Carroll, M., Dimiceli, C. & Sohlberg, R.A. (2003) Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions*, 7 (10), pp.1–15.
- Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O. & Townshend, J.R.G. (2013) High-resolution global maps of 21st-century forest cover change. *Science*, 342, pp.850–853.
- Hansen, M.C., Roy, D.P., Lindquist, E.J., Adusei, B., Justice, C.O. & Altstatt, A. (2008) A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sensing of Environment*, 112 (5), pp.2495–2513.
- Haralick, R.M., Shanmugam, K. & Dinstein, I. (1973) Texture Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3 (6), pp.610–

621.

Hennig, T.A., Kretsch, J.L., Pessagno, C.J., Salamonowicz, P.H. & Stein, W.L. (2001) The Shuttle Radar Topography Mission. In: C. Y. Westort ed. *Proceedings of the First International Symposium on Digital Earth Moving*. Manno, Switzerland, pp.65–77.

Herold, M. & Johns, T. (2007) Linking requirements with capabilities for deforestation monitoring in the context of the UNFCCC-REDD process. *Environmental Research Letters*, 2 (4), p.45025.

Hsu, C., Chang, C. & Lin, C. (2010) A Practical Guide to Support Vector Classification. Technical Report [Internet]. Available from: <[http://www.datascienceassn.org/sites/default/files/Practical Guide to Support Vector Classification.pdf](http://www.datascienceassn.org/sites/default/files/Practical%20Guide%20to%20Support%20Vector%20Classification.pdf)> [Accessed 15 September 2014].

JAXA (2010) PALSAR 50m Orthorectified Mosaic Product - Central Africa [Internet]. Available from: <http://www.eorc.jaxa.jp/ALOS/en/kc_mosaic/kc_50_c_africa.htm>.

JAXA-EORC (2016) Global 25m Resolution PALSAR-2/PALSAR Mosaic and Forest/Non-Forest Map (FNF) Dataset Description [Internet]. Available from: <http://www.eorc.jaxa.jp/ALOS/en/palsar_fnf/DatasetDescription_PALSAR2_Mosaic_FNF_revD.pdf> [Accessed 30 June 2017].

Jensen, J.R. (1996) *Introductory digital image processing: A remote sensing perspective*. 2nd ed. Englewood Cliffs, New Jersey, Prentice-Hall.

Joyce, R.J., Janowiak, J.E., Arkin, P.A. & Xie, P.-P. (2004) CMORPH : A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *Journal of Hydrometeorology*, 5, pp.487–503.

Justice, C.O., Wilkie, D.S., Zhang, Q., Brunner, J. & Donoghue, C. (2001) Central African forests, carbon and climate change. *Climate Research*, 17, pp.229–246.

Kearsley, E., de Haulleville, T., Hufkens, K., Kidimbu, A., Toirambe, B., Baert, G., Huygens,

- D., Kebede, Y., Defourny, P., Bogaert, J., Beeckman, H., Steppe, K., Boeckx, P. & Verbeeck, H. (2013) Conventional tree height-diameter relationships significantly overestimate aboveground carbon stocks in the Central Congo Basin. *Nature communications*, 4, p.2269.
- Lavalle, M. & Wright, T. (2009) *Absolute Radiometric and Polarimetric Calibration of ALOS PALSAR Products Generated within ADEN*.
- Legendre, P. (1993) Spatial Autocorrelation: Trouble or New Paradigm? *Ecology*, 74 (6), pp.1659–1673.
- Lewis, S.L. (2006) Tropical forests and the changing earth system. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 361 (1465), pp.195–210.
- Lillesand, T., Kiefer, R.W. & Chipman, J. (2008) *Remote Sensing and Image Interpretation*. John Wiley & Sons; 6th Edition, p 44.
- Lindquist, E.J. & D’Annunzio, R. (2016) Assessing global forest land-use change by object-based image analysis. *Remote Sensing*, 8 (8).
- Lopez-Gonzalez, G., Lewis, S.L., Burkitt, M. & Phillips, O.L. (2011) ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data. *Journal of Vegetation Science*, 22 (4), pp.6573–6576.
- Lu, D. & Weng, Q. (2007) A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28 (5), pp.823–870.
- Lucas, R.M., Armston, J. & Fairfax, R. (2010) An Evaluation of the ALOS PALSAR L-Band Backscatter — Above Ground Biomass Relationship Queensland , Australia : Impacts of Surface Moisture Condition and Vegetation Structure. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3 (4), pp.576–593.
- Lund, H.G. (2015) Definitions of Forest, Deforestation, Afforestation, and Reforestation

- [Internet]. Available from:
<https://www.researchgate.net/publication/259821294_Definitions_of_Forest_Deforestation_Afforestation_and_Reforestation> [Accessed 31 July 2017].
- Malhi, Y. & Grace, J. (2000) Tropical forests and atmospheric carbon dioxide. *Trends in ecology and evolution*, 15 (8), pp.332–336.
- Mansourpour, M., Rajabi, M.A. & Blais, J.A.R. (2006) Effects and performance of speckle noise reduction filters on active radar and SAR images. *Proc. ISPRS Symp. on Topographic Mapping from Space (with Special Emphasis on Small Satellites)*, (Ankara, Turkey XXXVI-1/W41.).
- Mayaux, P. & Achard, F. (2010) “REDD-plus” requirements for the Congo Basin countries. In: M. Brady & C. de Wasseige eds. *Monitoring Forest Carbon Stocks and Fluxes in the Congo Basin*. Brazzaville, Republic of Congo, pp.4–9.
- Mayaux, P., Bartholomé, E., Fritz, S. & Belward, A. (2004) A new land-cover map of Africa for the year 2000. *Journal of Biogeography*, 31, pp.861–877.
- Mayaux, P., De Grandi, G., Rauste, Y., Simard, M. & Saatchi, S.S. (2002) Large-scale vegetation maps derived from the combined L-band GRFM and C-band CAMP wide area radar mosaics of Central Africa. *International Journal of Remote Sensing*, 23 (7), pp.1261–1282.
- Mayaux, P., Pekel, J.F., Desclée, B., Donnay, F., Lupi, A., Achard, F., Clerici, M., Bodart, C., Brink, A., Nasi, R. & Belward, A. (2013) State and evolution of the African rainforests between 1990 and 2010. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368 (1625), p.20120300.
- Mayaux, P., Richards, T. & Janodet, E. (1999) A vegetation map of Central Africa derived from satellite imagery. *Journal of Biogeography*, 26 (2), pp.353–366.
- Megevand, C., Mosnier, A., Hourticq, J., Sanders, K., Doetinchem, N. & Streck, C. (2013) *Deforestation Trends in the Congo Basin: Reconciling Economic Growth and Forest Protection*. Washington, D.C., The World Bank.

- Melillo, J.M., McGuire, A.D., Kicklighter, D.W., Moore, B., Vorosmarty, C.J. & Schloss, A.L. (1993) Global climate change and terrestrial net primary production. *Nature*, 363, pp.234–240.
- Mitchard, E.T. a., Saatchi, S.S., Lewis, S.L., Feldpausch, T.R., Woodhouse, I.H., Sonké, B., Rowland, C. & Meir, P. (2011) Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest–savanna boundary region of central Africa using multi-temporal L-band radar backscatter. *Remote Sensing of Environment*, 115 (11), pp.2861–2873.
- Moreira, A., Krieger, G., Younis, M., Hajnsek, I., Papathanassiou, K.P., Eineder, M. & De Zan, F. (2011) Tandem-L: A mission proposal for monitoring dynamic earth processes. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2011*. IEEE, pp.1385–1388.
- Mountrakis, G., Im, J. & Ogole, C. (2011) Support vector machines in remote sensing : A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66 (3), pp.247–259.
- NASA (2017) Landsat 8 Overview [Internet]. Available from: <<https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-overview/>> [Accessed 14 August 2017].
- Njoku, E.G. (1999) AMSR Land Surface Parameters. *Algorithm Theoretical Basis Document Version 3.0, NASA Jet Propulsion Laboratory*.
- Njoku, E.G., Jackson, T.J., Lakshmi, V., Chan, T.K. & Nghiem, S.V. (2003) Soil moisture retrieval from AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing*, 41 (2), pp.215–229.
- NOAA Climate Prediction Center (2006) African Rainfall Estimation Algorithm Version 2.0 [Internet].
- Oliver, C. & Quegan, S. (2004) *Understanding Synthetic Aperture Radar*. Raleigh, N.C., SciTech Publishing, Inc.
- Olofsson, P., Foody, G.M., Stehman, S.V. & Woodcock, C.E. (2013) Making better use of

accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, pp.122–131.

OpenCV (2017) Random Trees [Internet]. Available from: <[http://docs.opencv.org/2.4.13.3/modules/ml/doc/random_trees.html?highlight=random forests](http://docs.opencv.org/2.4.13.3/modules/ml/doc/random_trees.html?highlight=random%20forests)> [Accessed 13 June 2017].

Papathanassiou, K.P., Kugler, F., Lee, S.K., Marotti, L. & Hajnsek, I. (2008) Recent advances in Polarimetric SAR Interferometry for forest parameter estimation. In: *IEEE RADAR Conference*. Rome, Italy, IEEE, pp.1–6.

Qin, Y., Xiao, X., Dong, J., Zhang, G., Roy, P.S., Joshi, P.K., Gilani, H., Murthy, M.S.R., Jin, C., Wang, J., Zhang, Y., Chen, B., Menarguez, M.A., Biradar, C.M., Bajgain, R., Li, X., Dai, S., Hou, Y., Xin, F. & Moore, B. (2016) Mapping forests in monsoon Asia with ALOS PALSAR 50-m mosaic images and MODIS imagery in 2010. *Scientific Reports*, 6 (July 2015), pp.1–10.

Quegan, S., Toan, T.L., Yu, J.J., Ribbes, F. & Floury, N. (2000) Multitemporal ERS SAR analysis applied to forest mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 38 (2), pp.741–753.

Regos, A., Ninyerola, M., Moré, G. & Pons, X. (2015) Linking land cover dynamics with driving forces in mountain landscape of the Northwestern Iberian Peninsula. *International Journal of Applied Earth Observation and Geoinformation*, 38, pp.1–14.

Reiche, J., Lucas, R., Mitchell, A.L., Verbesselt, J., Hoekman, D.H., Haarpaintner, J., Kellndorfer, J.M., Rosenqvist, A., Lehmann, E.A., Woodcock, C.E., Seifert, F.M. & Herold, M. (2016) Combining satellite data for better tropical forest monitoring. *Nature Climate Change*, 6 (2), pp.120–122.

Rexer, M. & Hirt, C. (2014) Comparison of free high resolution digital elevation data sets (ASTER GDEM2, SRTM v2.1/v4.1) and validation against accurate heights from the Australian National Gravity Database. *Australian Journal of Earth Sciences*, 61 (2),

pp.213–226.

Rocchio, L. & Ochs, B. (2007) About the Landsat Data Continuity Mission [Internet]. Available from: <<http://ldcm.gsfc.nasa.gov/about.html>>.

Romijn, E., Herbert, J., Wijaya, A., Herold, M., Angelsen, A., Verchot, L. & Murdiyarso, D. (2013) Exploring different forest definitions and their impact on developing REDD + reference emission levels : A case study for Indonesia. *Environmental Science and Policy*, 33, pp.246–259.

Rosenqvist, A., Finlayson, C.M., Lowry, J. & Taylor, D. (2007) The potential of long-wavelength satellite-borne radar to support implementation of the Ramsar Wetlands Convention. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 17 (3), pp.229–244.

Rosenqvist, A., Rebelo, L.-M. & Costa, M. (2015) The ALOS Kyoto & Carbon Initiative: enabling the mapping, monitoring and assessment of globally important wetlands. *Wetlands Ecology and Management*, 23 (1), p.1.

Rosenqvist, A., Shimada, M., Chapman, B., Freeman, A., De Grandi, G., Saatchi, S.S. & Rauste, Y. (2000) The global rain forest mapping project-a review. *International Journal of Remote Sensing*, 21 (6), pp.1375–1387.

Rosenqvist, A., Shimada, M., Suzuki, S., Ohgushi, F., Tadono, T., Watanabe, M., Tsuzuku, K., Watanabe, T., Kamijo, S. & Aoki, E. (2014) Operational performance of the ALOS global systematic acquisition strategy and observation plans for ALOS-2 PALSAR-2. *Remote Sensing of Environment*, 155, pp.3–12.

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M.A., Mitchard, E.T.A., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., Petrova, S., White, L., Silman, M. & Morel, A.C. (2011) Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (24), pp.9899–904.

Samba, G., Nganga, D. & Mpounza, M. (2008) Rainfall and temperature variations over Congo-Brazzaville between 1950 and 1998. *Theoretical and Applied Climatology*,

91 (1–4), pp.85–97.

- Santoro, M., Fransson, J.E.S., Eriksson, L.E.B., Magnusson, M., Ulander, L.M.H. & Olsson, H. (2009) Signatures of ALOS PALSAR L-Band Backscatter in Swedish Forest. *IEEE Transactions on Geoscience and Remote Sensing*, 47 (12), pp.4001–4019.
- Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C. & Townshend, J.R.G. (2013) Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS Vegetation Continuous Fields with lidar-based estimates of error. *International Journal of Digital Earth*, 1303210312.
- Sgrenzaroli, M., Baraldi, A., De Grandi, G.D., Eva, H. & Achard, F. (2004) A novel approach to the classification of regional-scale Radar mosaics for tropical vegetation mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 42 (11), pp.2654–2669.
- Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R. & Lucas, R. (2014) New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment*, 155, pp.13–31.
- Shimada, M., Kankaku, Y., Watanabe, M. & Motooka, T. (2011) Current Status of the ALOS-2 / PALSAR-2 and the CALVAL Program [Internet]. Available from: <[http://media.asf.alaska.edu/asfmainsite/documents/2011-ceos-workshop/Current Status of ALOS-2, M.Shimada.pdf](http://media.asf.alaska.edu/asfmainsite/documents/2011-ceos-workshop/Current%20Status%20of%20ALOS-2,%20M.Shimada.pdf)> [Accessed 16 June 2015].
- Singh, A. (1989) Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, 10 (6), pp.989–1003.
- Small, D. & Schubert, A. (2008) *Guide to ASAR Geocoding*. Zurich, Switzerland.
- Smith, P.M., Kalluri, S.N.V., Prince, S.D. & Defries, R.S. (1997) The NOAA/NASA PathfinderAVHRR 8-Km Land Data Set. *Photogrammetric Engineering and Remote Sensing*, 63, pp.12–31.
- Swain, P.H. & King, R.C. (1973) Two Effective Feature Selection Criteria for Multispectral

Remote Sensing. *Purdue University LARS Technical Reports*, 39 (42673).

- Takeuchi, S., Suga, Y. & Yoshimura, M. (2001) A Comparative Study of Coherence Information by L-band and C-band SAR for Detecting Deforestation in Tropical Rain Forest. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2001*. Sydney, Australia, IEEE, pp.2259–2261.
- Tansey, K.J., Luckman, A. & Skinner, L. (2004) Classification of forest volume resources using ERS tandem coherence and JERS backscatter data. *International Journal of Remote Sensing*, 25 (4), pp.751–768.
- Tewkesbury, A.P., Comber, A.J., Tate, N.J., Lamb, A. & Fisher, P.F. (2015) A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, 160, pp.1–14.
- Thiel, C.J., Thiel, C. & Schmullius, C.C. (2009) Operational Large-Area Forest Monitoring in Siberia Using ALOS PALSAR Summer Intensities and Winter Coherence. *IEEE Transactions on Geoscience and Remote Sensing*, 47 (12), pp.3993–4000.
- Thomas, N., Lucas, R., Itoh, T., Simard, M., Fatoyinbo, L., Bunting, P. & Rosenqvist, A. (2015) An approach to monitoring mangrove extents through time-series comparison of JERS-1 SAR and ALOS PALSAR data. *Wetlands Ecology and Management*, 23 (1), pp.3–17.
- Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K.P., Plummer, S., Rocca, F., Saatchi, S.S., Shugart, H.H. & Ulander, L.M.H. (2011) The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*, 115 (11), pp.2850–2860.
- Townshend, J.R.G., Hansen, M.C., Carroll, M.L., DiMiceli, C.M., Sohlberg, R.A. & Huang, C. (2011) User Guide for the MODIS Vegetation Continuous Fields product Collection 5.1, version 1 [Internet]. Available from: <http://modis-land.gsfc.nasa.gov/pdf/VCF_C5_UserGuide_Feb2015.docx> [Accessed 29 August 2013].
- Ulaby, F.T., Dubois, P.C. & van Zyl, J. (1996) Radar mapping of surface soil moisture.

- Journal of Hydrology*, 184, pp.57–84.
- UNFCCC (2001) Report of the Conference of the Parties on its seventh session, held at Marrakesh from 29 October to 10 November 2001. Addendum. Part two: Action taken by the Conference of the Parties. Volume I [Internet]. Available from: <<http://unfccc.int/resource/docs/cop7/13a01.pdf>>.
- UNFCCC (2010) Report of the Conference of the Parties on its sixteenth session, held in Cancun from 29 November to 10 December 2010. Addendum. Part Two: Action taken by the Conference of the Parties at its sixteenth session. [Internet]. Available from: <<http://unfccc.int/resource/docs/2010/cop16/eng/07a01.pdf>>.
- UNFCCC (2006) Report of the Conference of the Parties serving as the meeting of the Parties to the Kyoto Protocol on its first session, held at Montreal from 28 November to 10 December 2005 Addendum Part Two: Action taken by the Conference of the Parties serving as the [Internet]. Available from: <<http://unfccc.int/resource/docs/2005/cmp1/eng/08a03.pdf>>.
- USGS (2015a) Basin Excess Rainfall Maps [Internet]. Available from: <<https://earlywarning.usgs.gov/fews/product/54>> [Accessed 12 December 2015].
- USGS (2015b) HYDRO1K Documentation ReadMe [Internet]. Available from: <<https://lta.cr.usgs.gov/HYDRO1KReadMe>> [Accessed 12 December 2015].
- Verdin, K.L. & Verdin, J.P. (1999) A topological system for delineation and codification of the Earth's river basins. *Journal of Hydrology*, 218, pp.1–12.
- de Vries, F.P.P. (1998) *Speckle reduction in SAR Imagery by various multi-look techniques*. The Hague, Netherlands.
- Wagner, W., Luckman, A., Vietmeier, J., Tansey, K.J., Balzter, H., Schmullius, C.C., Davidson, M.W.J., Gaveau, D., Gluck, M., Le Toan, T., Quegan, S., Shvidenko, A., Wiesmann, A. & Yu, J.J. (2003) Large-scale mapping of boreal forest in SIBERIA using ERS tandem coherence and JERS backscatter data. *Remote Sensing of Environment*, 85, pp.125–144.

- Walker, W.S., Stickler, C.M., KelIndorfer, J., Kirsch, K.M. & Nepstad, D.C. (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), pp.594–604.
- Watanabe, M. & Shimada, M. (2006) Relation between coherence , forest biomass , and L-band σ_0 . In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2006*. Denver, CO, pp.1651–1654.
- Weydahl, D.J. & Eldhuset, K. (2012) Combining Polarimetric and Coherence SAR Images for Forest Type Discrimination. In: *9th European Conference on Synthetic Aperture Radar. EUSAR, 2012*. Nuremberg, Germany, pp.235–238.
- Wheeler, J., Rodriquez-Veiga, P., Balzter, H., Tansey, K. & Tate, N. (2017) Forest Mapping of the Congo Basin using Synthetic Aperture Radar (SAR). In: H. Balzter ed. *Earth Observation for Land and Emergency Monitoring; Innovative Concepts for Environmental monitoring from Space*. London, UK, Wiley, p.Ch.4.
- White, L., Brisco, B., Dabboor, M., Schmitt, A. & Pratt, A. (2015) A Collection of SAR Methodologies for Monitoring Wetlands. *Remote Sensing*, 7 (6), pp.7615–7645.
- Woodhouse, I.H. (2006) *Introduction to Microwave Remote Sensing*. Boca Raton, FL, CRC Press.
- Zebker, H.A. & Villasenor, J. (1992) Decorrelation in Interferometric Radar Echoes. *IEEE Transactions on Geoscience and Remote Sensing*, (818), pp.1–19.
- Zhou, L., Tian, Y., Myneni, R.B., Ciais, P., Saatchi, S., Liu, Y.Y., Piao, S., Chen, H., Vermote, E.F., Song, C. & Hwang, T. (2014) Widespread decline of Congo rainforest greenness in the past decade. *Nature*, 509 (7498), pp.86–90.