

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**A Comparison of Classification Techniques for Monitoring and Mapping Land Cover  
and Land Use Changes in the Subtropical Region of Thai Nguyen, Vietnam**

A thesis presented in partial fulfilment of the requirements for the degree of

Master of  
Environmental Management

at Massey University, Palmerston North  
New Zealand.

**Tuyen Ha Van**

2017

## **Abstract**

Deriving land cover/land-use information from earth observation satellite data is one of the most common applications for environmental monitoring, evaluation and management. Many parametric and non-parametric classification algorithms have been developed and applied to such applications. This study looks at the classification accuracies of three algorithms for different spatial and spectral resolution data. The performance of Random Forest (RF) was compared to Maximum Likelihood (MLC) and Artificial Neural Network (ANN) algorithms for the separation of subtropical land cover/land-use categories using Sentinel-2 and Landsat 8 data. The overall, producers' and users' accuracies were derived from the confusion matrix, while local land use statistics were also collected to evaluate the accuracy of classified images. The accuracy assessment showed the RF algorithm regularly outperformed the MLC and ANN in both types of imagery data (>90%). This approach also exhibited potential in dealing with the challenge of separating similar man-made features such as urban/built-up and mining extraction classes. The ANN algorithm had the lowest accuracy among the three classification algorithms, while Landsat 8 imagery was most suitable for the classification of subtropical mixed and complex landscapes.

As the RF algorithm demonstrated a robustness and potential for mapping subtropical land cover/land-use, this study chose it to monitor and map temporal land cover/land-use changes in Thai Nguyen, Vietnam between 2000 and 2016. The results of this temporal monitoring revealed that there were substantial changes in land cover/land use over the course of 16 years. Agricultural and forest land decreased, while urban and mining extraction land expanded significantly, and water increased slightly. Changes in land cover/land-use are strongly associated with geographic locations. The conversion of agriculture and forest into urban/built-up and mining extraction land was detected largely in the Thai Nguyen central city and southern

regions. In addition, further GIS analysis revealed that approximately 69.6% (100.2km<sup>2</sup>) of new built-up areas had occurred within 2km of primary roads, and nearly 96% (137.6km<sup>2</sup>) of new built-up expansion was detected within a 5-km buffer of the main roads. This study also demonstrates the potential of multi-temporal Landsat data and the combination of remote sensing, GIS and R programming to provide a timely, accurate and economical means to map and analyse temporal changes for long-term local land use development planning.

**Keywords:** Random forest; Land cover mapping; Remote Sensing; Vietnam

## Acknowledgements

I would first like to express my gratitude and sincere thanks to my supervisors, **Mike Tuohy** and **Matthew Irwin** at Institute of Agriculture and Environment, Massey University for their guidance, valuable advice, encouragement, and endless patience through the entire process of this thesis research. They always allowed me to work on my own idea, but steered me on the right direction whenever I needed assistance.

My special thanks also go to all lecturers for their enthusiastic lectures and encouragement during my first year at Massey University, and to my classmates for cheering me up and giving constructive advice. Special thanks also devote to all Vietnamese friends and scholars at Massey University for their continuous encouragement and support.

I would also like to thank my friends back home in Vietnam who involved in fieldtrip survey for this thesis research, particularly Mr. **Tuan**, Mr. **Hung** and Mr. **Hoang**. This research would not be complete without their enthusiasm and assistance.

I would also love to take this opportunity to thank New Zealand government for sponsoring my home data collection, and for the scholarship to pursue the Master's degree in Environmental Management at Massey University. Special thanks also dedicate to **Jamie**, **Dave**, **Logan** and all staff at International Student Support Office, Massey University for making sure that all scholars like me could obtain the best academic achievements and enjoyable life experience while in New Zealand.

Finally, I must express my very profound gratitude to my family for providing me with continuous support and love throughout my years of study.

## Table of Contents

Abstract.....	ii
Acknowledgements.....	iv
Table of Contents.....	v
List of Tables.....	ix
List of Figures.....	x
List of Abbreviations.....	xii
1 Introduction.....	1
1.1 Background.....	1
1.2 Land cover/ land-use in Thai Nguyen, Vietnam.....	3
1.3 Research problem.....	7
1.3.1 Research goals.....	8
1.3.2 Research concerns.....	8
1.4 The organization and structure of the research thesis.....	9
2 LITERATURE REVIEW.....	10
2.1 Remote sensing and GIS.....	10
2.1.1 Remote sensing.....	10
2.1.2 Geographic Information System.....	13
2.2 Image Processing.....	14
2.2.1 Pre-processing.....	14
2.2.2 Classification algorithms.....	16
2.3 Temporal land cover and land-use monitoring and mapping.....	23

2.4	Change detection and Accuracy assessment .....	25
2.4.1	Change detection.....	25
2.4.2	Accuracy assessment .....	26
3	STUDY AREA and MATERIALS.....	28
3.1	Study Area.....	28
3.2	Biophysical and Scio-economic Characteristics .....	30
3.2.1	Topology and Hydrology .....	30
3.2.2	Geology, Soil and Mineralogy.....	31
3.2.3	Flora and Fauna.....	31
3.2.4	Population and Economic Development.....	32
3.3	Data Collection.....	32
3.3.1	Satellite Data and Thematic Maps .....	33
3.3.2	Training and Reference Data .....	35
3.3.3	Software .....	37
4	METHODS .....	37
4.1	Image Pre-processing .....	37
4.1.1	Radiometric Correction.....	38
4.1.2	Enhancement and Transformations.....	38
4.1.3	Mosaic from Multiple Images.....	40
4.2	Classification Scheme .....	41
4.3	Comparison of Land Cover Classifiers .....	48

4.3.1	Classification Algorithms .....	48
4.3.2	Sentinel-2 and Landsat 8 Accuracy Assessment .....	52
4.4	Land Cover Monitoring and Mapping for Thai Nguyen.....	54
4.4.1	Landsat 7 and 8 Classification Accuracy Assessment.....	55
4.4.2	Change Detection.....	56
5	RESULTS AND DISCUSSION .....	56
5.1	Comparison of Classifiers .....	57
5.1.1	Land Cover Maps Derived from Sentinel-2.....	57
5.1.2	Land Cover Maps Derived from Landsat 8 .....	63
5.2	Temporal Land Cover Change Monitoring.....	69
5.2.1	Land cover/land use change.....	69
5.2.2	Classification accuracy assessment.....	71
5.2.3	Land cover change patterns .....	73
6	CONCLUSIONS AND RECOMMENDATIONS .....	76
6.1	Conclusions .....	76
6.1.1	Comparison of land cover classification techniques.....	76
6.1.2	Temporal land cover monitoring and mapping.....	77
6.2	Recommendations .....	78
6.2.1	Comparison of land cover classification techniques.....	78
6.2.2	Land cover monitoring and mapping.....	78
	References.....	79



Appendices..... 85

Appendix 1: TNMT land use statistics ..... 85

Appendix 2: Accuracy statistics for 20-m Sentinel-2 data using the three classification algorithms ..... 86

Appendix 3: Accuracy statistics for 10-m Sentinel-2 data using the three classification algorithms ..... 87

Appendix 4: Accuracy statistics for Landsat 8 data using the three classification algorithms 88

Appendix 5: Accuracy statistics for Landsat 7 data using the RF classification algorithms... 89

## List of Tables

Table 3.1 Data layers used in this study .....	33
Table 3.2 Training data statistics for Sentinel-2 and Landsat 8 data (pixels).....	36
Table 4.1 Description of various land cover classes in Thai Nguyen, Vietnam.....	43
Table 5.1 Summary of accuracy statistics for three classifiers using Sentinel-2 data (%) .....	59
Table 5.2 Summary of area statistics for the three classifiers using Sentinel-2 data.....	61
Table 5.3 Summary of accuracy statistics for the three classifiers using Landsat 8 data (%) .	65
Table 5.4 Summary of area statistics for the MLC, ANN and RF algorithms using Landsat 8 data.....	66
Table 5.5 Summary of classification statistics for 2000 and 2016 .....	70
Table 5.6 Summary of Landsat ETM+ and Landsat OLI classification accuracies (%) .....	72
Table 5.7 Matrix of land cover/land-use and changes (area in km <sup>2</sup> ) between 2000 and 2016	75

## List of Figures

Figure 1.1 Spatial distribution of industrial zone development by 2020, Thai Nguyen.....	6
Figure 2.1 Sentinel-2 spectral and spatial resolution.....	12
Figure 2.2 Spectral reflectance of green vegetation, Landsat 8 data .....	15
Figure 2.3 An example of land cover classification using spectral class.....	16
Figure 2.4 Right-skewed distribution of reflectance for mining class, Sentinel-2 band 3.....	17
Figure 2.5 Land cover studies publications using the ANN approach between 1990 and 2005 .....	22
Figure 3.1 A map of the study area.....	29
Figure 3.2: Elevation and water network for Thai Nguyen derived from DEM.....	30
Figure 3.3 Clouds in the Landsat 8 image .....	34
Figure 4.1 A false colour composite of the study area (Sentinel-2 bands 3, 4 and 8) .....	40
Figure 4.2 Natural colour Sentinel-2 imagery without colour balancing (A), and colour balancing (B).....	41
Figure 4.3 Normalized differenced vegetation index of the study area (Sentinel-2 bands 8 and 4) .....	42
Figure 4.4 Rural houses (left) and urban built-up infrastructure (right) .....	44
Figure 4.5 Examples of agriculture; harvested rice field (left) and corn fields (right).....	45
Figure 4.6 Mixed forests (left) and forest plantation (right).....	46
Figure 4.7 Tea plantations (left) and tea intercropping (right), Dai Tu district, Thai Nguyen	46
Figure 4.8 A part of Cong River (left) and Coc lake (right) .....	47
Figure 4.9 Coal mining activities in Dai Tu district (left) and iron exploitation (right) in Dong Hy district, Thai Nguyen province.....	48
Figure 4.10 An example of decision tree .....	50
Figure 4.11: The operational procedure of feed-forward neural network .....	52

Figure 4.12 Reference data for assessing the Sentinel-2 and Landsat 8 classification.....	53
Figure 4.13 Reference data for assess the accuracy of Landsat 7 (collected based on Google Earth and ESRI Imagery 2000).....	55
Figure 5.1 Land cover/land-use maps derived from Sentinel-2 data using MLC, ANN and RF algorithms .....	58
Figure 5.2 The similarity in mining extraction and urban feature surface of the study area ...	60
Figure 5.3 The comparison of area statistics reported between classification algorithms for Sentinel-2 data for the study area.....	62
Figure 5.4 The effect of tree sizes on the RF model accuracy using Sentinel-2 data.....	63
Figure 5.5 Land cover/land-use maps derived from Landsat 8 data using MLC, ANN and RF algorithms .....	64
Figure 5.6 The comparison of area statistics between algorithm-based and TNTM statistics	66
Figure 5.7 The comparison of area statistics reported between classification algorithms using Landsat 8.....	67
Figure 5.8 The effect of tree sizes on the model accuracy using Landsat 8 .....	68
Figure 5.9 The variable importance identified by the RF model for Landsat 8 bands .....	69
Figure 5.10 Land cover classification products derived from Landsat data for 2000 and 2016 .....	70
Figure 5.11 Area statistics by land cover for 2000 and 2016 and the areas of change.....	71
Figure 5.12 Variations in the percentage of each land cover between local statistics and the RF classification derived from the Landsat 8 .....	73
Figure 5.13 Major changes in two land cover/ land-use categories using Landsat data between 2000 and 2016.....	74

## **List of Abbreviations**

ROI: Region of Interest

DN: Digital Number

ETM+: Enhanced Thematic Mapper Plus

OLI/ TIR: Operational Land Imager/ Thermal Infrared

GIS: Geographic Information System

RS: Remote Sensing

MLC: Maximum Likelihood Classifier

ANN: Artificial Neural Network

RF: Random Forest

CART: Classification and Regression Tree

PCA: Principal Component Analysis

TOA: Top of Atmosphere

ENVI: Environment for Visualizing Images

NDVI: Normalized Differenced Vegetation Index

UTM: Universal Transverse Mercator

WGS84: World Geodetic System 84

NASA: National Aeronautics and Space Administration

USWGS: United States Geological Survey

VNIR: Visible/Near-infrared

SWIR: Shortwave Infrared

RGB: Red, Green and Blue

PCC: Post Classification Comparison

TNMT: Thai Nguyen Department of Natural Resource and Environment