Study of Dynamic Process of Land Use Land Cover Changes in Vietnam

July 2022

Phan Cao Duong

Study of Dynamic Process of Land Use Land Cover Changes in Vietnam

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
the University of Tsukuba
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Environmental Studies
(Doctoral Program in Sustainable Environmental Studies)

Phan Cao Duong

Abstract

Land use/land cover (LULC) information is essential for environmental studies toward sustainable development. However, a profound understanding of change processes, patterns, and rates is challenging due to the lack of comprehensive and consistent LULC data/maps. To improve the quality of LULC maps, remote sensing and artificial intelligence techniques have recently been utilized increasingly, but there are still limitations such as low accuracy or resolutions, and the shortage of reliable reference datasets in a long-term period. Therefore, a comprehensive framework of LULC monitoring is fundamental for future land assessment. This study aims to develop such a framework for long-term time-series monitoring of land use/land cover change (LULCC) on the Vietnam-wide scale. Four major tasks were done as follows.

First, this dissertation estimated the uncertainties of currently existing LULC products in Vietnam. To this end, high-resolution LULC maps were first generated for LULCC hot spots at a sub-national scale. Reliable data sets of reference samples were also established through extensive field surveys across the country (in 2015, 2016, 2018, 2019, and 2020), statistical inventory data, and visual interpretation from high-spatial-resolution images in Google Earth. Then, a synthesis was conducted for LULC maps in Vietnam. Results showed that few products of LULC have been produced at the Vietnam-wide scale while most of them focus on specific LULC types such as forests and croplands. Global LULC databases are available on a national scale. They, however, contain uncertainties and inconsistencies such as low accuracy (< 78%), various spatial resolution (10 – 1,000 m), and differences between the definitions of the LULC categories. These products seldom meet the prime requirement of projects' objectives. Hence, the improvement of LULC mapping is in urgent need, providing reliable data for further environmental research.

Given the above-mentioned importance, the second section aimed at proposing a novel approach (Ensemble Learning Updating Classifier/ELUC), which could be applied with various classification algorithms and datasets to simplistically generate new classifications or renew existing classifications with a remarkable accuracy improvement. Applying miscellaneous features of Landsat-8 images, the ELUC of a random-forest-based algorithm produced sequences of single-time classifications with a mean overall accuracy of 84%. Through the study period, these sequences were subsequently joined to achieve a final classification that reached an overall accuracy of 94%. Also, the ELUC of the random-forest-based algorithm outperformed that of Kernel-Density-Estimation with a 5% overall accuracy higher. These outcomes confirm

the effectiveness of the ELUC for a remarkably consistent land use/cover estimation in a data-rich environment.

The third chapter was to develop a comprehensive framework for the automatic production of the first Vietnam-wide annual land use/land cover datasets (VLUCDs) from 1990 to 2020. Specifically, after comprehensively assembling and preprocessing on the GEE platform, various data sources were integrated, including the informative Landsat TM, ETM+ and OLI, Sentinel SAR GRD and MSI images, and ground-truth data. Together with the ELUC, an automatic training migration model (ATMM) was adopted to monitor Vietnam LULC. The obtained VLUCDs had overall accuracy ranging from 85.7 $\pm$ 1.3% to 92.0 $\pm$ 1.2% with the ten primary dominant LULC and 77.6 $\pm$ 1.2% to 84.7 $\pm$ 1.1% with the eighteen secondary dominant LULC over the study period. This confirms the potential of the proposed framework for the systematically long-term monitoring of LULC in Vietnam. Also, the dynamic process of major LULC changes was estimated across mainland Vietnam. The most dominant LULC type was forests, accounting for approximately half of the entire country, followed by croplands (16.3%), rice fields (14.2%), and open water (including parts of saltwater, 8.1%). Grassland and scrubland occupied a relatively similar proportion (2.8%) while the smallest LULC was residential areas (1.3%). Over the study period, major LULC changes frequently occurred in forests, agriculture, aquaculture, wetlands, and residential land. Specifically, despite slight recoveries in 2000 and 2010, the net loss of forests (19,940 km^2) mainly transformed into croplands over 30 years. Some productive croplands were converted to urban areas, which increased approximately ten times. A threefold increase in aquaculture caused a major loss of wetlands (1,914 km^2). Changes varied according to region, but the most dynamic regions were the western north, the southern center, and the south. These findings can provide evidence-based information on formulating and implementing clear land management policies. The explicitly spatio-temporal VLUCDs can be benchmarks for regional and global LULC validation and utilized for a variety of applications in the research of environmental changes toward the Sustainable Development Goals.

The remaining part of the dissertation was to quantify the socioeconomic and biophysical drivers of LULCC with a focus on forests. Utilizing the VLUCDs, the dynamic changes in forest land were quantified from 1990 to 2020. To decide the major drivers of changes, a synthesis of previous articles focusing on forest transition studies in Vietnam at various spatial-scale levels was conducted. Subsequently, a machine learning technique was utilized to measure the drivers of the forest changes. Our results indicated that although the forest area increased from 2005 to 2010, it underwent a decrease over the whole study period. There was a dramatic conversion between forest and agricultural land, especially in the northwest and central highland areas. This conversion was mainly driven by agricultural and plantation expansion/shifting, accessibility/infrastructure, population growth/migration, and distance to systems such as irrigation, drainage, and mining/industry. The identification of the drivers is likely to help to enhance the accuracy of the land use/land cover change prediction. Also, these findings provide scientific evidence about the dynamics and drivers of forest

changes at the nationwide and decadal scales and thus can support informing the core policies of forest management in Vietnam.

Acknowledgement

First, I gratefully acknowledge the financial support of the Project for Human Resource Development Scholarship by Japanese Grant Aid (JDS). I acknowledge the support of the Hydraulic Construction Institute - Vietnam Academy for Water Resources for allowing me to be absent from work to do my Ph.D. in Japan.

I would like to thank my chief supervisor, Dr. NASAHARA Nishida Kenlo, for his guidance and mentorship. He has helped me to become an independent researcher and to overcome unfamiliar circumstances. He is not only my supervisor but also my friend who helps me enjoy a happy life in Japan. I very much appreciate my co-supervisors, Dr. TSUJIMURA Maki and Dr. ASANUMA Jun, and the rest of the dissertation committees, Dr. MIZUNOYA Takeshi, Dr. MATSUSHITA Bunkei, Dr. AKITSU Kawaguchi Tomoko, Dr. UCHIDA Taro, Dr. YABAR Mostacero Helmut Friedrich, and Dr. YAMAMOTO Sachiko who have given thoughtful questions, critiques, profound comments, and constructive feedback to enhance this study.

I would like to thank the organizations and their professional colleagues who have assisted this work. Specifically, I appreciate support from the Japan Aerospace Exploration Agency and the Remote Sensing Technology Center of Japan, especially Dr. TADONO Takeo, Dr. HAYASHI Masato, and Ms. OHGUSHI Fumi for data provision and sharing the JAXA High Resolution Land Use Land Cover Map Products of Vietnam. I express my thanks to Dr. HOANG Thanh Tung, Mr. TA Hoang Trung, Mr. TRUONG Van Thinh, and Mr. NGUYEN Ho for helping to collect reference data through Vietnam. I sincerely appreciate support from my current organization (Hydraulic Institute - Vietnam Academy for Water Resources) in Vietnam, especially Dr. TRAN Dinh Hoa, Dr. NGUYEN Quoc Dung, Dr. NGUYEN Thanh Cong, Dr. TRAN Van Thai, and Mr. THAI Quoc Hien for their tremendous support and positive encouragement.

I acknowledge all of the faculty, staff, and friends from the Graduate School of Life and Environmental Sciences and the Watershed Management Laboratory who have helped me during my PhD life. Also, I would like to thank the people outside of the University of Tsukuba who have worked as my academic and professional mentors, including Dr. DO Hoai Nam, Dr. VU Ba Thao, Dr. BUI Tien Dieu, Dr. TRAN Dang An, Dr. PHAM Tien Dat, and Dr. PHAM Duc Chinh, all helped to provide constructive feedback and profound comments on my study.

I very much appreciate my families for their lifelong help, especially my parents for doing hard to secure my approach to a high-quality education. Finally, I would like to express my special thanks to my wife, Giang, who has been my biggest source of love

and encouragement throughout this entire process.

Contents

List of Figures

List of Tables

List of Publications

Peer-reviewed publications

1. **Duong Cao Phan**, Ba Thao Vu, Dang An Tran, Vuong Trong Kha, and Kenlo Nishida Nasahara. "Dynamics and determinants of forest changes across mainland Vietnam in the recent three decades." **(Accepted in Recent Advanced Research in Water Resources and Environmental Systems - A SpingerNature book)**; (Part of Chapter 7 of this dissertation).

2. Ho Nguyen, Ta Hoang Trung, **Duong Cao Phan**, Thong Anh Tran, Thi Hai Ly Nguyen, Kenlo Nishida Nasahara, Alexsander Prishchepov, and Norbert Holzel. **2022**. "Transformation of Rural Landscapes in the Vietnamese Mekong Delta from 1990 to 2019: A Spatio-Temporal Analysis." *Geocarto International*, June, 1–23; https://doi.org/10.1080/10106049.2022.2086623

3. **Duong Cao Phan**, Ta Hoang Trung, Van Thinh Truong, Taiga Sasagawa, Thuy Phuong Thi Vu, Dieu Tien Bui, Masato Hayashi, Takeo Tadono, and Kenlo Nishida Nasahara. **2021**. "First Comprehensive Quantification of Annual Land Use/Cover from 1990 to 2020 across Mainland Vietnam." *Scientific Reports* 11 (1); (Chapter 6 of this dissertation). https://doi.org/10.1038/s41598-021-89034-5.

4. **Duong Cao Phan**, Ta Hoang Trung, Van Thinh Truong, and Kenlo Nishida Nasahara. **2021**. "Ensemble Learning Updating Classifier for Accurate Land Cover Assessment in Tropical Cloudy Areas." *Geocarto International*, February, 1–18; (Chapter 5 of this dissertation). https://doi.org/10.1080/10106049.2021.1878292.

5. **Duong Cao Phan**, Ta Hoang Trung, Kenlo Nishida Nasahara, and Takeo Tadono. **2018**. "JAXA High-Resolution Land Use/Land Cover Map for Central Vietnam in 2007 and 2017." *Remote Sensing*; (Chapter 4 of this dissertation). https://doi.org/10.3390/rs10091406.

6. Phuong Thao Thi Ngo, Tien Dat Pham, Viet Ha Nhu, Thu Trang Le, Dang An Tran, **Duong Cao Phan**, Pham Viet Hoa, José Lázaro Amaro-Mellado, and Dieu Tien Bui. **2021**. "A Novel Hybrid Quantum-PSO and Credal Decision Tree Ensemble for Tropical Cyclone Induced Flash Flood Susceptibility Mapping with Geospatial Data." *Journal of Hydrology* 596. https://doi.org/10.1016/j.jhydrol.2020.125682.

7. Viet Ha Nhu, Phuong Thao Thi Ngo, Tien Dat Pham, Jie Dou, Xuan Song, Nhat Duc Hoang, Dang An Tran, **Duong Cao Phan**, et al. **2020**. "A New Hybrid Firefly-Pso Optimized Random Subspace Tree Intelligence for Torrential Rainfall-Induced Flash Flood Susceptible Mapping." *Remote Sensing* 12 (17). https://doi.org/10.3390/RS12172688.

8. Van Thinh Truong, Thanh Tung Hoang, **Duong Cao Phan**, Masato Hayashi, Takeo Tadono, and Kenlo Nishida Nasahara. **2019**. "JAXA Annual Forest Cover Maps for Vietnam during 2015-2018 Using ALOS-2/PALSAR-2 and Auxiliary Data." *Remote Sensing* 11 (20). https://doi.org/10.3390/rs11202412.

9. Van Thinh Truong, **Duong Cao Phan**, Kenlo Nishida Nasahara, and Takeo Tadono. **2019**. "How Does Land Use/Land Cover Map's Accuracy Depend on Number of Classification Classes?" *Scientific Online Letters on the Atmosphere* 15: 28–31. https://doi.org/10.2151/SOLA.2019-006.

10. Do Hoai Nam, Tran Dinh Hoa, **Duong Cao Phan**, Duong Hai Thuan, and Dang Thanh Mai. **2019**. "Assessment of Flood Extremes Using Downscaled CMIP5 High-Resolution Ensemble Projections of near-Term Climate for the Upper Thu Bon Catchment in Vietnam." *Water* (Switzerland) 11 (4). https://doi.org/10.3390/w11040634.

11. Do Hoai Nam, **Duong Cao Phan**, Duong Hai Thuan, Dang Thanh Mai, and Nguyen Quoc Dung. **2018**. "Assessment of Near-Term Runoffresponse at a River Basin Scale in Central Vietnam Using Direct CMIP5 High-Resolution Model Outputs." *Water* (Switzerland) 10 (4). https://doi.org/10.3390/w10040477.

12. **Duong Cao Phan**, Alexandra Nauditt, Hoai Nam, and Nguyen Tung Phong. **2016**. "Assessment of Climate Change Impact on River Flow Regimes in The Red River Delta, Vietnam – A Case Study of the Nhue-Day River Basin." *Journal of Natural Resources and Development* 06: 81–91. https://doi.org/10.5027/jnrd.v6i0.09.

Under review

1. Tien Dat Pham, Nam Thang Ha, **Duong Cao Phan**, Neil Saintilan, Andrew Skidmore, Nga Nhu Le, Hung Luu Viet, Wataru Takeuchi, and Danial A. Friess. "Advances in Earth Observation and Machine Learning for Quantifying Blue Carbon." **(Under review in Earth-Science Reviews)**.

2. Thanh Tung Hoang, **Duong Cao Phan**, Takeo Tadono, and Kenlo Nishida Nasahara. "A Spatiotemporal Analysis of Deforestation in Vietnam over the Last Two Decades." **(Under review in Land Use Policy)**.

In preparation

1. Ho Nguyen, **Duong Cao Phan**, Duong Vu, Luan Pham, Alexander Prischepov, and Norbert Hoelzel. "Monitoring Land-use/Land-cover Transition in the Vietnamese Mekong Delta: An overview of existing map products." *(Target journal: GIScience & Remote Sensing)*

2. **Duong Cao Phan**, Thao Ba Vu, Taro Uchida, Kenlo Nishida Nasahara, Dieu Tien Bui, Do Hoai Nam, and Tran Dinh Hoa. "Recent Advances in Machine Learning and Remote Sensing for Modelling Flash-flood Susceptibility Areas: Current status, Challenges, and Prospects." *(Target journal: GIScience & Remote Sensing)*

Data publications

1. Annual Land Use and Land Cover Maps across Mainland Vietnam from 1990 to 2020 (Released in September 2021 / Version 21.09).
https://www.eorc.jaxa.jp/ALOS/en/dataset/lulc/lulc_vnm_v2109_e.htm

2. High-Resolution Land Use and Land Cover Map of the Central Region of Vietnam (Released in July 2018 / Version 18.07).
https://www.eorc.jaxa.jp/ALOS/en/dataset/lulc/lulc_vnm_v1807_e.htm

3. Photos from our field surveys (Here); A list of location notices of different land use/land cover (see Appendix A.3 for more details).

List of Acronyms and Abbreviations

Abbreviations	The description/full form
AGRHYMET	Agro-meteorology and operational hydrology
ANN	Artificial neural networks
AOI	Areas of interest
ArcGIS	Aeronautical Reconnaissance Coverage Geographic Information System
ASA	Arid and semi-arid lands
ATMM	automatic training migration model
AVHRR	Advanced very high resolution radiometer
CCI	Climate Change Initiative
DEM	Digital elevation model
DSM	Digital surface model
DT	Decision tree
ELUC	Ensemble Learning Updating Classifier
EMS	Electromagnetic spectrum
ENSO	El Niño southern oscillation
ENVI	Environment for Visualizing Images
Envisat	Environmental Satellite
EO	Earth Observation
ERDAS	Earth Resources Data Analysis System
ERS	European Remote Sensing Satellite
ESA	European Space Agency
ESRI	Environmental Systems Research Institute
ETM	Enhanced thematic mapper
ETM+	Enhanced Thematic Mapper plus
EVI	Enhanced vegetation index
FAO	Food and Agricultural Organization of the United Nations.
FLAASH	Fast line-of-sight atmospheric analysis of spectral hypercubes
FRA	Global Forest Resources Assessment
FROM-GLC	Finer Resolution Observation and Monitoring of Global Land Cover
GBFM	Global Boreal Forest Mapping
GE	Google Earth
GEE	Google earth engine

Continued on next page

Continued from previous page

Acronyms	The description/full form
GLCNMO	Global Land Cover by National Mapping Organizations
GIS	Geographic information systems
GMES	Global monitoring for environment and security
GPS	Global positioning system
GRD	Ground range detected
ha	Hectare
HH	Horizontal transmit and horizontal receive
HIS	Hue
HRV	High Resolution Visible
HRLULC	High-Resolution Land Use and Land Cover Map
HV	Horizontal transmit and vertical receive
IR	Infrared
ISODATA	Iterative self-organising data analysis
JAXA	Japan Aerospace Exploration Agency
JDS	Japanese Grant Aid
K-NN	K-nearest neighbour
km^2	Square Kilometre
LCC	Land Cover Change
LEDAPS	Landsat ecosystem disturbance adaptive processing system
LiDAR	Light detection and ranging
LULC	Land Use Land Cover
LULCC	Land Use Land Cover Change
m^3	Cubic metres
MERIS	Medium resolution imaging spectrometer
MSS	Multispectral scanner
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NDWI	Normalised difference wetness index
NIR	Near Infrared
NOAA	National Oceanic and Atmospheric Administration
OBIA	Object-based Image Analysis
OLI	Operational Land Imager
PALSAR	Phase Array type L-band SAR
PCA	Principal Component Analysis
QGIS	Quantum Geographic Information System
RADAR	Radio detection and ranging
RESTEC	Remote Sensing Technology Center of Japan
RF	Random Forest
RGB	Red Green Blue
ROI	Region of Interest

Continued on next page

Continued from previous page

Acronyms	The description/full form
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SAVI	Soil and Vegetation Index
SD	Standard Deviation
SDGs	Sustainable development goals
SLC	Scan Line Corrector
SPOT	Satellite Pour L'Observation de la Terre
SVM	Support Vector Machine
SWIR	Short-wave infrared
TCWI	Tasselled cap wetness index
TM	Thematic Mapper
TOA	Top of Atmosphere
TRMM	Tropical Rainfall Measuring Mission
UAV	Unmanned aircraft vehicle
USAID	United States agency for international development
USGS	United Sates Geological Service
UTM	Universal transverse Mercator
VH	Vertical transmission and horizontal receiving
VHR	Very High Resolution
VIIRS	Visible Infrared Imaging Radiometer Suite
VLUCDs	Vietnam-wide annual land use/land cover data sets
VV	Vertical transmission and vertical receiving
WCMC	World Conservation Monitoring Centre
WRI	Water ratio index

Chapter 1

Introduction

1.1 Background of the research problem

Land use/land cover (LULC) information has a pivotal role in the sustainable development of society. Land cover is what people observe on the Earth's surface such as vegetation, bare soil, building, etc. Land use is the use of the land which is managed for specific purposes (e.g., cultivation and recreation). A plot of land might simultaneously be utilized for different purposes. LULC may change differently according to time and space. A change in land cover means the alteration of land cover types (e.g., forest to cropland) or characteristics (e.g., degradation and structure). A change in land use is characterized by changes in management practices, intensification, etc. A LULC change (LULCC), which is caused by both anthropogenic and natural factors, can lead to alterations in, for example, global energy and biogeochemical cycles [2], services of ecosystems [3], climate [4], biodiversity (e.g., pollinator) [5], carbon sequestration [6], and sea-level rise [7]. Therefore, LULCC data provide land policymakers with scientific evidence-based information to understand the fundamental alterations of land dynamic conditions. In other words, the data support to manage sustainably the environment.

Even the recognition of its importance, a proper understanding of LULCCs at a large scale is challenging due to the shortage of accurate and consistent LULC maps or data [8]. Most of the existing LULC data lack fine temporal or spatial information (available a very few times or at a coarse spatial resolution) and differences in LULC classification schemes [9]. Particularly, remote-sensing-based data may have a fine spatial resolution but they are available for a single or few times. Existing long-term LULC data are usually observed for a single land type such as cropland [10] or a predefined LULC class system (the term "*class*" is used to describe a LULC type/category in this study), which may not meet the requirement of most projects [11]. Many of these time-series data have not been completely checked the accuracy in terms of time and space, owing to the lack of coherent reference data [12]. On the contrary, statistical and inventory data have been conducted for a long period, but they are mainly constrained by small (at village, district, and province scales) or developed regions, fragmentation, or focusing only on land use. These inventory data may not be publicly opened to all users, especially in

the developing world. Meanwhile, recent evidence suggests that the LULCC has been more dynamic in the developing regions such as Southeast Asia but likely decreased in the other regions and on the global scale [13]. These high dynamic changes of LULC have caused difficulties in effectively monitoring LULC in such a developing region. It is also noted that most developing countries are located in subtropical or tropical climate regions, which are frequently covered by dense clouds, challenging the accurate assessment of LULC from space. Other difficulties in consistently estimating LULC in the developing world are the lack of coherent reference data and the high degree of uncertainties in the current LULC reconstruction data [14]. These challenges are certainly true in the case of Vietnam.

As a developing country, Vietnam has increasingly experienced a considerable change in LULC. However, available LULC products mainly focus on small administrative areas or have few predefined times [15]. Recently, there are several inter-provincial LULC data sets such as the central and southern LULC maps of the years 2007 and 2017, and the northern maps of the years 2007 and 2015 [16]. Truong et al. has mapped annual land cover between 2015 and 2019 [17] with a focus on forest types. However, the accuracy of non-forest categories is low and might not be utilized for other primary purposes. As such, although the report of the Ministry of Agricultural and Rural Development (MARD, 2016) has shown a continuous net forest gain, Hensen et al., with a systematically comprehensive review, has reported forest loss in Vietnam [18]. It means that the rates, patterns, and processes of the change in forests on the national scale have not been fully understood. Case studies over the past two decades have provided important information on LULCC in Vietnam. Typical examples are the expansion of urban areas [19] and a boom in the local aqua-cultural industry [20]. Hence, timely, accurate, and comprehensive LULC products at the national scale can provide a better profound understanding of the rate, pattern, and change processes of LULC in Vietnam. This information can support policymakers in forming crucial decisions on sustainable development and resource management. The LULC products and generated ground-truth reference data can be benchmarks for validating regional and global land cover databases.

Given the above-mentioned motivation, an important objective of this study is to annually produce a comprehensive LULC database across mainland Vietnam from 1990 to 2020. To this end, a general methods was established to overcome current paramount concerns, including the lack of reference data, cloud cover problems, and the limitations of existing classification methods. First, a full standard set of extensive reference data was established for training and validating a proposed classification method. Specifically, nationwide on-site detailed surveys have been conducted annually from 2015 to 2020. This work was supported by the locals and remote sensing experts. Along with these data, further reference information from statistics, inventory, existing fragmented maps, and high-spatial-resolution satellite images in Google Earth were extracted carefully. To minimize labor, cost, and time for obtaining further reference data, a potential training migration method/model (in this study, "*model*" is a simple description of a system or process that can be used in calculations or predictions of what

might happen) was applied to acquire training data from 1990 to 2014. Secondly, in order to overcome data shortage and cloud cover problems, an innovative classification method was proposed to take advantage of all the available remote sensing data, including Landsat, Sentinel, and ALOS sensors. Specifically, a systematic review was conducted to find out the most suitable classification algorithm ("*algorithm*" is a set of mathematical instructions or rules that, especially if given to a computer, will help to calculate an answer to a problem). Previous research has reported that although a large number of advanced classification methods have been developed, there is not the best one for all. Each method has its own advantages and disadvantages, and thus may only perform effectively with the case studies [21]. Then, experiments were conducted to test the effectiveness of the best-recommended methods and found that a random forest is the most effective one. This also accords with earlier observations, which showed that the random forest shows the out-performance of its rivals such as fuzzy adaptive resonance theory-supervised predictive mapping (Fuzzy ARTMAP), support vector machine (SVM), artificial neural network (ANN), Mahalanobis distance (MD), and spectral angle mapper (SAM) [22]. Subsequently, a new random-forest-based method was developed in this study. Unlike the common use of single-time classification, time-series images were first classified independently. The second step in this process was to develop an approach for optimizing post-classification results. The proposed method achieved an outstanding performance according to a benchmark established by Truong et al. [23]. It enables the generation of an explicit annual LULC database at a Vietnam-wide scale. This can be utilized for a tremendous variety of applications in the research of environmental changes in Vietnam. The method is expected to deploy elsewhere at a broader scale, even the global LULC quantification towards the Sustainable Development Goals.

1.2 Research aims and objectives

The essential aim of this dissertation is to create a new comprehensive approach to estimate the long-time dynamics of land use/land cover changes in mainland Vietnam. The approach is effective and efficient by utilizing the public availability of multiple remotely sensed sources and ground-based data. Connecting the analyzed changes to socioeconomic and biophysical variables, the determinants of the changes are discovered to provide policymakers with evidence-based information for land and resource management. To this end, specifically principal objectives are as follows.

- To build a common dataset of ground-truth data for LULC classification and valuation across mainland Vietnam.
- To reveal the degree of uncertainties and inconsistencies of the currently existing LULC databases in mainland Vietnam.
- To develop an effective method for mapping accurately LULC in tropical cloudy regions, including mainland Vietnam.

- To establish a comprehensive framework to quantify annual LULC from 1990 to 2020 across mainland Vietnam.
- To connect the LULCC to socioeconomic and biophysical variables for determining the drivers of the changes.

1.3 Dissertation structure

My thesis is composed of eight themed chapters. The first chapter of this thesis provides a brief introduction of the background of the research problem, research objectives, and thesis organization. The second chapter conducts a comprehensive literature review. This includes the importance of land use/land cover (LULC) assessment, the LULC databases at global, national, and Vietnam-wide scales, remote-sensing-based approaches for LULC analyses, and the challenges and opportunities for improving LULC monitoring at national or above scales. The third chapter is concerned with the study area and materials deployed in this study. The main issues addressed in this thesis are chapters 4, 5, 6, and 7. Chapter four analyses the uncertainties and consistencies of some global LULC databases at the Vietnam-wide scale. In chapter 5, a practical approach has been developed for accurate LULC assessment in tropical cloudy regions, including Vietnam. The sixth chapter presents a coherent framework for comprehensive quantification of annual LULC from 1990 to 2020 across mainland Vietnam. Chapter seven provides an estimation of LULC changes and drivers over the recent three decades in Vietnam. The remaining part of the thesis proceeds as follows: a conclusion of the study findings, the limitations, and recommendations for future work.

Chapter 2

Literature review

2.1 The importance of land use/land cover information updating

The literature on environmental studies has highlighted the importance of LULC information. For example, the significant loss of forests impacts carbon-cycle balances, which can cause changes in climate system [24]. Such changes are expected to alter natural habitats, which may be a major driver of biodiversity losses [25]. LULC information is the main indicator in mapping natural hazards such as landslides, debris flow, and flash floods [26, 27]. Comprehensive assessment of LULCC is an important debate topic in international associations such as the Paris Agreement [28]. Meanwhile, LULC changes constantly and global LULCC are much more considerable than previous assessment [13]. Most publicly available LULC databases have limitations, for example, a limited number of predefined LULC classes or a coarse spatial-temporal resolution. Hence, more regularly updating and detailed LULC information is in great need. Such information can provide a profound understanding of spatial-temporal dynamic changes in LULC and evidence-based information for various environmental studies, including planning and policy makers towards socio-economic sustainable development.

2.2 Land use/land cover studies

2.2.1 A global scale

Given the above-mentioned essence, numerous projects have been recently developed into creating global LULC databases. Reliably applicable databases include the USGS EROS Archive - Land Cover Products - Global Land Cover Characterization [29], Global Land Cover Facility [30], Global Land Cover 2000 [31], GlobCover 2009 [32], CCI Global Land Cover [33], MCD12Q21 MODIS [34], GLCNMO [35], FROM-GLC [36], GLC-GCS30 [37], CGLS-LC100 [38], and Esri 10-Meter Land Cover [39]. Among these databases, the most state-of-the-art products are described in detail in Appendix A.1.

As shown in Appendix A.1, there is considerable improvement of the global LULC databases. The spatial resolution of the data have currently reached 10 m in comparison to 1,000 m in the 1990s. There is also a remarkable increase in the overall accuracy from 67% to 86%. Several data have a high temporal resolution, for example, the MODIS LC and CCI, which have observed yearly since the 2000s. However, the global LULC data have limitations. These data used different land cover classification systems and the number of classes varies according to the individual databases, ranging from 10 to 32 classes. The reference data (see Section 3.3.1 for the detailed definition of "*reference data*" in this study) of several databases are different. For example, the GLCNMO has 904 training sites while the ESRI-LC10 has 23,693,961 training sites. In addition, these LULC databaes utilize different input data and classification methods. This might cause inconsistencies among these data and hard to find which data is the most reliable and applicable.

2.2.2 A regional and national scale

There is a growing body of literature that recognises a need of continental/national-scale LULC to obtain more detail information at a smaller scale. One of the first continental scale databases is the land-cover database created for sub-saharan Africa (1982 - 1991). They were created using remote-sensing-based spectral bands and indices, for example, vegetation and surface temperature. The database, however, has a coarse spatial resolution (1,000 m) and shortage of representative reference data [40]. To increase the detail of mapping, L. Durieux use the fusion of the Global Boreal Forest Mapping radar and the Medium resolution imaging spectrometer images to map the vegetation of Siberia region at a spatial resolution of 300 m [41]. Several attempts have established more reliable LULC databases such a cropland mapping at the African continental scale [42] and a LULC change over the continental US [43]. Recent work has tried to monitor LULC at finer spatial resolution (30 m or 10 m) and higher classification accuracy. An example is the 30-m spatial resolution wall-to-wall LULC of Australia (1985 - 2015) [44], and the Land Cover Mapping at 10 m resolution over Europe [45].

The pioneers in LULC surveys at a national scale were in developed nations such as the USA, Japan, and Australia. Then, developing countries such as India, Nepal, Saudi Arabia, and Iran had programmes on LULC surveys [46]. Particularly, in the 2000s, several attempts created nationwide maps, such as LULC monitoring in Spain [47], in Netherlands [48], in the USA [49], and in Portugal [50]. Recently, several countries have been updated their national LULC databases, for examples, a land cover database to circa 2001 and 2011, a update of nationwide LULC database of China in 2010. Numerous countries have nationwide LULC product but very few countries have multi-date LULC products. For example, the US has released the National Land Cover Database 2016 with 7 products dates including 2001, 2004, 2006, 2008, 2011, 2016, and 2016 [51].

2.2.3 Land use/land cover in Vietnam

To understand LULC in Vietnam, a synthesis is conducted, focusing the studies from 1990 to 2020. There have been approximately 300 publications on LULC studies. In particular, from all databases in Web of Science, an advance search with the query expression "TI = (Determinants OR drivers OR causes OR dynamics AND land*) AND TS = (Vietnam OR "Viet Nam" AND land*) AND TS = (agricul* OR crop* OR *fores* OR defor* OR refor* OR degrad* OR aquacul* OR cropland*)" is carefully conducted. Also, an advanced search in Google Scholar is employed, though there are duplicates. After the duplicates removed, 545 publications are selected in total, of which 297 studies focused on LULC in Vietnam. Going through the titles and abstracts, 48 publications work qualitatively or quantitatively on the determinants of driving changes in LULC, especially forests, agriculture, or/and aquaculture. These are the most dynamic LULC types in Vietnam. After the careful reading of the full articles, the study topics (e.g., forest loss and forest gain), key determinants, study area locations, study periods, and methods used by the selected articles were recorded correctly. Major determinants were decided based on the selected publications if they noted such a major determinant. Similar determinants were merged to generalize the determinants. They are described in more detail in Appendix A.6.

In Vietnam, satellite-based methods have been employed to generated numerous LULC products, but most focus on small-scale regions or have few predefined times [15]. Several attempts have been made to create inter-provincial LULC databases such as the seven-category LULC maps for the central and southern Vietnam in 2007 and 2017, and the northern Vietnam in 2007 and 2015 [16]. Recently, Vietnam-wide maps were produced to map annual forest cover from 2015 to 2019 with the primary focus on forest monitoring [52]. As such, the accuracy of non-forest LULC classes might be insufficient for other objectives. Meanwhile, there has been a highly dynamic LUCC which varies among different regions in Vietnam. Despite the report of continuous net forest gain by the Ministry of Agricultural and Rural Development (MARD, 2016) a systematically comprehensive review has reported forest loss in Vietname [18]. The rates and patterns of changes at the nation scale may not be fully understood. Hence, timely, accurate, and comprehensive LULC products can provide a profound understanding of LUCC patterns and processes. This information can supports policymakers in forming crucial decisions on sustainable development and resource management. The maps may be benchmarks for quantifying regional and global land cover products.

2.3 Remote-sensing-based analyses of land use/land cover

2.3.1 Remotely sensed data

Remote sensing data are fundamental for the LULC monitoring and assessment. They provide valuable information at multiple scales in time and space. There are also multiple spectra (e.g., Red, Blue, and Green), indices (e.g., Normalized Difference

Vegetation Index), hyper-spectral data (e.g., Earth Observing One – Hyperion), multiple angular data (e.g., Multiangle Imaging Spectroradiometer), night-time light data (e.g., Visible Infrared Imaging Radiometer Suite), microwave data (e.g., ALOS data), and lidar data (e.g., Digital Surface Model). The data have been advanced in terms of resolutions and access. For example, Landsat had a 16-day repeat orbit in the past but it has an 8-day repeat orbit currently. Recently, a 5-day repeat coverage of Sentinel-1 and -2 images with a 10-m spatial resolution has been available in public. Medium spatial resolution satellite data are easily obtained from different archives, for example, the USGS EarthExplorer, Landviewer, and Copernicus Open Access Hub. Fine resolution data are also available with orders. In summary, various remotely sensed data have been effectively utilized for the analysis of LULC changes.

2.3.2 The usage of geospatial software for image processing

Different geospatial software tools were used for remotely sensed image processing. Basic tools are ArcGIS, ENVI, Quantum GIS, ERDAS Imagine, and IDRISI. Recently, open-source software has been used more frequently such as Google Earth Engine, Python, Matlab, and Rstudio. The open-source software has advantages for geospatial analysts due to its flexible, available, and simple modification with regularly updated effective plug-ins. Popular software is described in Appendix A.2

2.3.3 Image pre-processing techniques

Remote sensing data may contain noise and errors, which can be removed and/or corrected by using effective pre-processing techniques. They include radiometric corrections (atmospheric and topographic corrections), geometric corrections (orthorectification and registration), and image enhancement [53]. In addition, for different data sources, coordinate re-projection and resampling steps are critical needs. Data with a UTM projection were warped to the WGS84 latitude-longitude projection. Although various resampling methods have been developed, widely used methods are nearest neighbourhood, bilinear interpolation, cubic convolution, Lanczos, and mode methods.

2.3.4 Land cover classification system

Defining a standard land cover classification system (LCCS) is a crucial step in the practical land cover assessment. It should be delineated precisely depending on the objectives of users and the availability of mapping resources. Most LULC maps employ the theory and framework of the International Geosphere-Biosphere Programme (IGBP) [54], the Land Cover Classification System (LCCS; https://www.fao.org/3/x0596e/x0596e00.htm), and the Coastal Change Analysis Program (C-CAP) Land Cover Classifications [55]. However, some categories in these LCCSs do not fully reflect at local or national scales. Taking Vietnam as an typical example, where snow and ice do not exist, while one cropland category does not represent the diverse croplands in Vietnam. Although

detailed classifications of high and low developed built-up areas play a fundamental role in urban planning and management for the rapid urbanization of Vietnam, they are not include in the previous LULC products. Therefore, a new LULC classification system was developed by remaining the appropriate categories of the above-mentioned LCCSs and adding new proper categories based on the local biophysical environment and end-users' recommendations.

2.3.5 LULC classification methods

Various approaches have been applied in the classification of remote sensing data for LULC analyses. Early approaches were visual interpretation and pattern recognition, which were mostly developed in the 1970s and 1980s [56]. The following decades witnessed the effective utilization of computer-based and knowledge-based approaches, including pixel-based, object-based, and hybrid methods [57]. Specifically, widely used algorithms are Decision Tree (DT), Random Forest (RF), Multilayer Perceptron (MP), Artificial Neural Network (ANN), K-nearest Neighbours Algorithm (k-NN), and Support Vector Machine (SVM). Among these algorithms, extensive research has shown that the RF properly outperforms the others [57, 22]. Object-based methods likely outperformed the others in the classification of specific land types such as forest, agriculture, and wetlands; they, however, required a careful selection of optimal segmentation scales, which challenges most analysts.

Thanks to the advancement of remote sensing technologies and computational sciences, deep learning (DL) has been successfully employed in LULC analyses. DL models were developed from neural networks. In the beginning, there were supervised models such as convolutional neural networks (CNN) and recurrent neural networks (RNN). The supervised models required a large number of training data, which are costly and time-consuming for comprehensively collecting. To overcome such a training collection issue, analysts developed unsupervised models, for example, deep belief networks (DBN), generative adversarial networks (GAN), and autoencoders (AE). Several attempts have been made to review the application of deep learning in remote sensing, especially LULC classification [58, 59].

2.3.6 Accuracy assessment

Assessing the accuracy of LULC maps and area of changes is important to quantify the quality of provided information. It also helps to understand the uncertainty of the data. Although several approaches have been well-developed, there is not the best one [60, 61, 62]. Recently, a "good practice" has been recommended by Pontus Olofsson et al. The detail of this approach is rigorously described in [63]. In general, there are three main steps, namely sampling design, feedback design, and analysis. It is noted that the accuracy validation and area assessment are done after checking and removing clear errors in the classified maps. A sampling method is essential to minimize the cost

of accuracy assessment. It is prohibitive to collect a large number of reference data. Another consideration is that data utilized for accuracy validation must be independent of the samples utilized for creating the classified maps.

2.3.7 Change analysis

The analysis of changes in land cover starts from comparison between old and new LULC maps. We are particularly interested in the characteristics of the changes, namely, location, extent, timing and speed. Extensive research has shown that the principal methods of change analyses can be pre-classification and post-classification comparison; the post-classification technique is more effectively utilized in LULC change analyses [64]. The post-classification comparison can minimize negative effects due to the differences between sensors, climate, and environmental conditions; it provides with more detailed information about LULC and thus is widely used to quantify the magnitude and speed of dynamic changes in LULC. In practice, change indicators including the area of net annual change (a (km^2.year^{-1}); Eq. (2.1)), the percentage of net annual change (p (year^{-1}); Eq. (2.2)), and the speed of net annual change (r (year^{-1}); Eq. (2.3)) are commonly used [65].

$$a = \frac{A_{t2} - A_{t1}}{t_2 - t_1} \tag{2.1}$$

$$p = \left(\frac{A_{t2} - A_{t1}}{A_{t1}(t_2 - t_1)} \right) * 100 \tag{2.2}$$

$$r = \left(\frac{1}{t_2 - t_1} \right) * \ln \left(\frac{A_{t2}}{A_{t1}} \right) \tag{2.3}$$

where A_{t1} and A_{t2} (km^2) are the area of the land cover type in the observation years t_1 and t_2 respectively ($t_1 < t_2$).

2.4 Challenges and opportunities for improved land use/cover monitoring at a national or broader scale

Although LULC databases have been increasingly well-established, there are limitations and challenges. First, the detailed information of the existing large-scale LULC databases is sometimes unclear. Most of the existing LULC databases have been conducted at a global scale, which frequently has a coarse spatio-temporal resolution and low accuracy. Specifically, the accuracy of the databases varies according to local and national locations, where have different levels of landscape or LULC heterogeneity. Such heterogeneity challenges the accurate assessment of LULC. To overcome this challenge, several attempts have been made to create LULC databases at a national scale, mainly in developed countries. Developing countries, however, experience a shortage of informative LULC data. These data may be established by governmental projects,

but they are rarely shared publicly. Second, statistical and ground-based reference data are rarely collected or not openly available at large scales. Unfortunately, the LULC community has focused on creating classification algorithms or models rather than creating standard reference data. Numerous models have developed, but there is not the best one. Selecting the best model depends on specific study cases. Last but not least, the existing large-scale LULC databases have high uncertainties and inconsistencies regarding the definition of map categories and the accuracy validation [66]. Recent evidence has shown that there is a profound difference in the area of a single LULC class derived from disparate remotely sensed data [13]. Although multiple LULC databases have been developed recently, it is still not known which database is the most comprehensive one. Therefore, to build a comprehensive one is a critical need for understanding and quantifying large-scale LULC dynamics.

Recently, the advances in remotely sensed technologies and computational capacities have provided potential chances for LULC assessment. New models have been increasingly developed with more stability, generalizability, and reliability, but low computational cost. These models are deployed with powerful computing systems or public cyberinfrastructures such as Google Earth Engine (GGE) and Amazon Web Service (AWS). In these systems, various software and numerous satellite data are compatible, integrated, or built-in, allowing users to conveniently achieve their goals. To utilize these potential opportunities, the LULC community is expected to challenge the following tasks. First, a comprehensive assessment of publicly existing LULC databases is a crucial task, which provides the uncertainties and inconsistencies of the data in detail. Another challenging task is to standardize, harmonize, and combine multiple data sources to establish accurate and consistent LULC databases. These databases then should be shared publicly with systematic platforms and protocols. To this end, one of the important objectives is to create a great volume of ground-truth data for reference data, which are effectively utilized to train and validate models.

Chapter 3

Study area and materials

3.1 Study area

The study area is mainland Vietnam (Fig. 3.1). The country has approximately 97 million population (2018; Fig. 3.2) and covers a land area of over 300,000 km^2. It includes the Red River Delta and the Mekong River Delta which is the third-largest delta in the world. The country has diverse topography with over 75% of the total area being hills and mountains, covering mainly by tropical rain-forests. Climate is changeable according to different regions (six climate regions) but dominated by a tropical monsoon type with the mean annual humidity of 84%; the mean annual rainfall between 1,200 and 3,000 mm and the mean annual temperature from 21 – 27 ^oC (Fig. 3.3) [67]. Its diverse climate and topography form the rich biodiversity and landscape heterogeneity of Vietnam's LULC. However, there are the identifying characteristics of LULC in different climatic regions. While the southern part is mainly covered by rice, aquaculture, and orchards, the northern part is primarily occupied by forests and plantations, except for the Red River Delta. In the northern centre, the dominant lands are evergreen broadleaf forests and annual croplands whereas woody crops, deciduous broadleaf forests, and evergreen needle-leaf forests are dominant lands in the southern centre.

3.2 Data acquisition and pre-processing

3.2.1 Remote sensing data

Multiple sources of remote sensing data are utilized in this work. First, Landsat TM, ETM+, and OLI Surface Reflectance Tier 1 with a 30-m spatial resolution, Sentinel MSI Level-2A and SAR GRD with a 10-m spatial resolution were collected properly. The Landsat provided by the USGS while the Sentinel provided by the ESA are publicly available in Google Earth Engine with several preprocessing steps being done. For example, the Landsat and the Sentinel MSI have been atmospherically corrected while the Sentinel SAR GRD has gone through thermal noise removal, radiometric calibration,

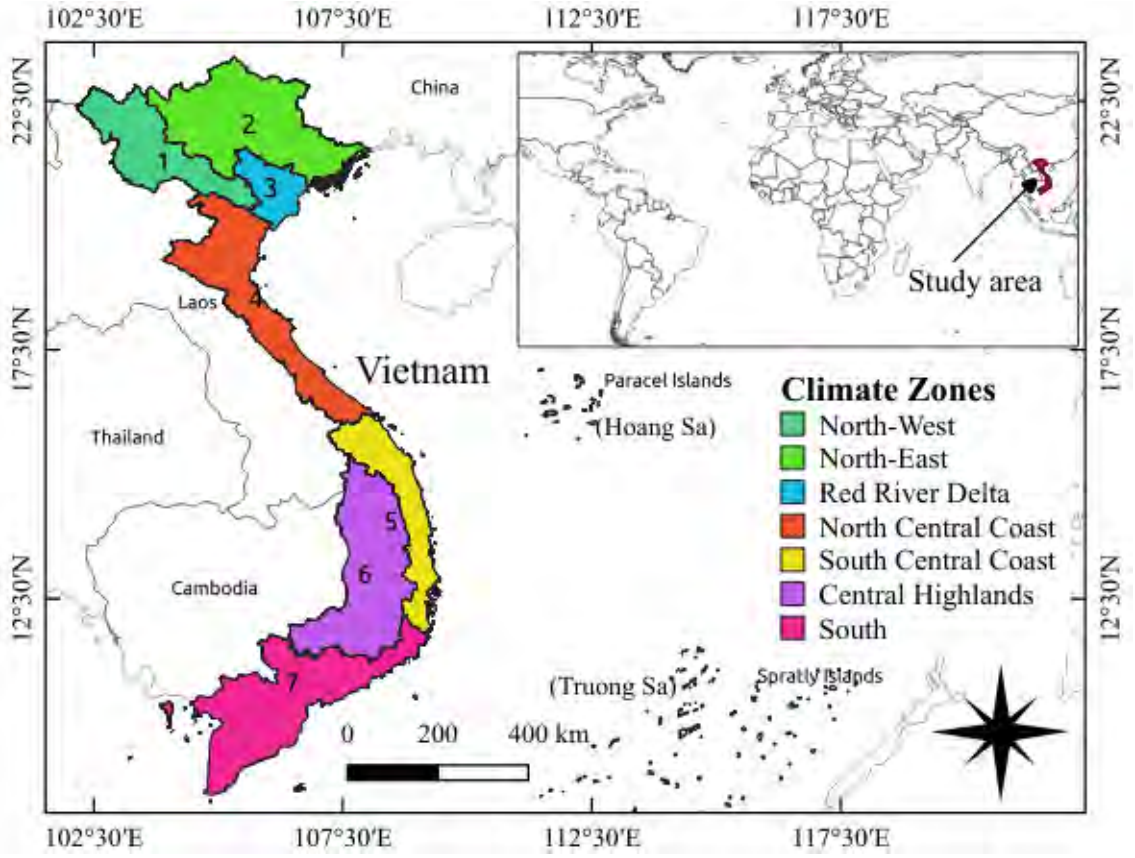


Figure 3.1: Study area; the location of Vietnam in the world and seven main zones of climate.

terrain correction using Sentinel-1 Toolbox and Shuttle Radar Topography Mission (SRTM) data. Regarding geometric correction, over 99% of the datasets from the GEE archive have high geometric accuracy with the error being less than half a pixel. Images with the geometric error greater than half a pixel are removed from this study. Further processing is done to support time-series analyses. To reduce illumination impacts from elevation, aspect and slope, the topographic correction is performed using the Modified Sun-Canopy-Sensor Topographic Correction algorithm for the Landsat and Sentinel MSI. The Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) is applied to perform atmospheric correction for Landsat TM and ETM+, and the Land Surface Reflectance Code (LaSRC) is adopted for Landsat OLI. All Landsat images are masked and removed clouds, cloud shadows and saturation pixels utilizing the Function of Mask (CFMASK). Sen2Cor is adopted to correct atmospheric issues and to mask clouds for Sentinel MSI. Finally, because of the different solar and view angles of the Landsat OLI and the Sentinel MSI, normalizing the bidirectional reflectance distribution function (BRDF) is applied for the data. For the Sentinel SAR GRD data, a further process was speckle filtering. The filtering was done using Lee filter, which is superior due to its capacity of maintaining point targets, edge, linear spaces and texture information. The mosaic images of ALOS AVNIR-2, ALOS PALSAR, and ALOS-2 PALSAR-2 mosaic were also utilized effectively. These data are derived from the Japan

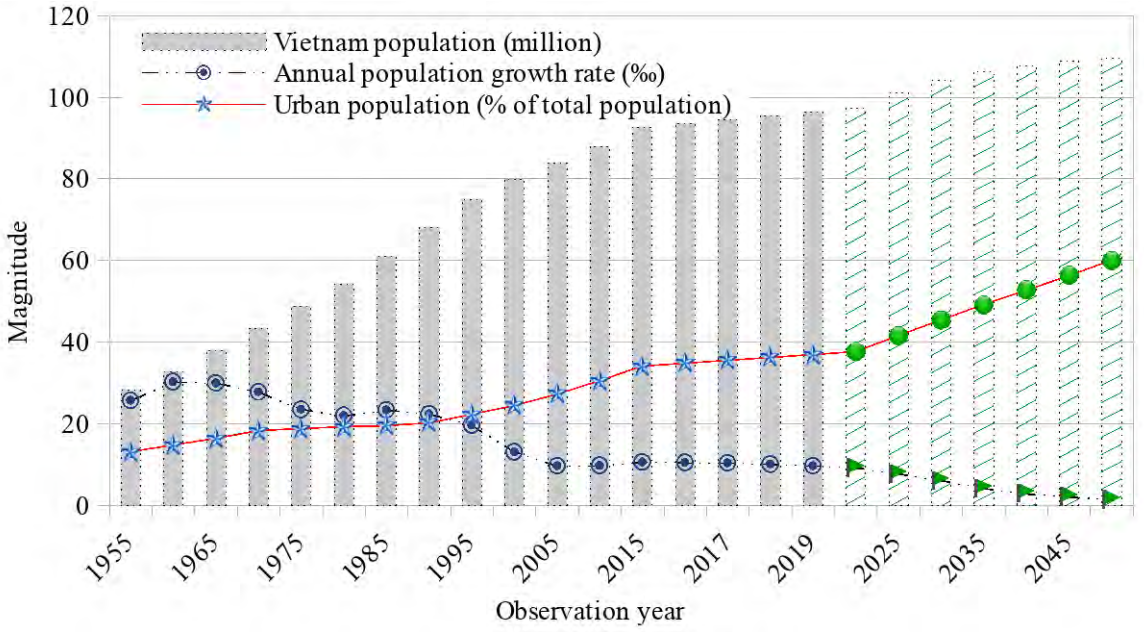


Figure 3.2: Trends and projections of Vietnam population; Annual growth rate (‰) and urban population (% of total population) from 1955 to 2020 with 5 year interval projections until 2050. The data are derived from the Population and Houses census, the General Statistics Office of Vietnam.

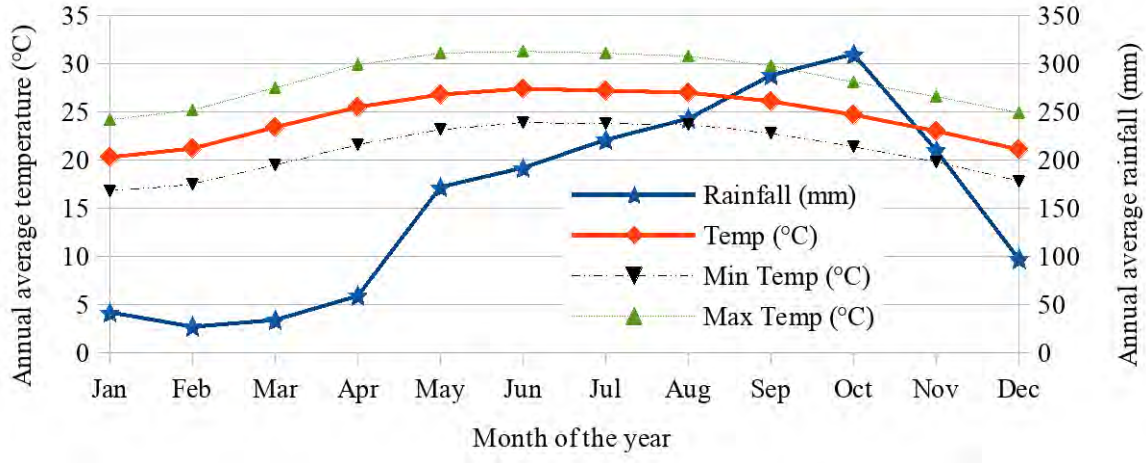


Figure 3.3: Annual average temperature and rainfall from 1900 to 2020 in Vietnam. The data are derived and computed from the Administrative unit, Lands and Climate, the General Statistics Office of Vietnam.

Aerospace Exploration Agency (JAXA). They are preprocessed with basic atmospheric corrections, cloud masking, geometric corrections, and speckle filtering. All the data are re-projected to Universal Transverse Mercator (UTM) projection (Zone 47-49 N and WGS-84 datum) and resambled into a designated spatial resolution using a bicubic interpolation method. All the processing steps are automatically completed using the Geospatial Data Abstraction Library (GDAL), the Geographic Resources Analysis Support System, and Python programming in Ubuntu 18.04.5 LTS.

3.2.2 Proxy data of the biophysical and socio-economic determinants

Major biophysical and socio-economic determinants were selected based on the evidence of the case-study synthesis of LULCC drivers. They are provided in Appendix A.7 with reference sources. Most data were annually collected at a commune-level scale from the General Statistics Office of Vietnam (mainly from 1995 to 2019; household surveys). Also several data were created such as distance from roads, rivers and irrigation systems. All of these data were originally arranged in different formats, including vector, raster and tabular data. The data had different spatial resolution but finer than the commune level. Henceforth, an enormous effort was performed to organize, clean, and check the data quality before they was utilized for further analysis. While keeping the original spatial resolution of the commune-level scale data (e.g., paddy yield and agriculture expansion), mean values were calculated at the commune-level scale for the finer spatial resolution data (e.g., distance to drainage and soil types). At the commune-level scale, mean, rate of change and standard deviation values were computed over the study period for time-series data (e.g., annual precipitation from 1990 to 2020) while mean values were calculated for the constant data such as elevation and slope, which do not significantly alter over time or are not available for multiple periods.

3.2.3 Ancillary data

Recent evidence suggests that ancillary information can help increase the accuracy of LULC mapping [52, 16, 17]. Herein, various ancillary data including elevation, slope, and aspect are used. These data are derived from ALOS Global Digital Surface Model or “ALOS World 3D-30m (AW2D30)" [68]. Transport systems, buildings, distance to rivers, coastlines, and soil classes are also added as covariates. Whereas the transport systems and buildings are derived from the OpenStreetMap, river networks and soil classes are derived from the OpenDevelopmentMekong.

3.3 Field surveys and reference data

3.3.1 Reference data extracted from field surveys

Reference data is an essential key in mapping LULC. The term "*Reference data*" refers to data that are used for training and validating a LULC classification model. In other words, ambiguous reference data cause confusion to the classification model. Then, a misclassification map is produced. To overcome such a misclassification, references must be ground-truth data, which are extracted from field surveys. However, LULC field surveys are costly and time-consuming, especially at a nationwide scale and in the past. Hence, apart from field survey data, it is necessary to make use of available reliable data, including the statistical inventory data and high-resolution images in Google Earth. Before creating reference data, a eighteen-class LULC system (Appendix A.4) was first designed based on the standard Land Cover Classification System, the

local biophysical environment, and end-users recommendations in Vietnam. Visual interpretations with high-resolution images in Google Earth were utilized to properly identify LULC types at the nationwide scale. At a local scale, unsure areas were noted for double check with detailed field surveys. Surveys in 2015, 2016, 2018, 2019, and 2020 were properly conducted across the country. During these surveys, GPS photos were taken using CASIO H20G, GoPro HERO 6, and GoPro HERO 7. In-depth interviews are also conducted with the locals and local experts. Based on the local knowledge, reference photos, and high-resolution images from Google Earth, then approximately 130,000 reference sites were designed carefully. A reference site was an area of a homogeneous land type, which covered an area of a circle with a radius of 75 m. A size of the circle was selected because it must cover at least nine pixel of the largest-spatial-resolution satellite image used. This confirmed that the reference was defined explicitly and properly in the site. Subsequently, the reference sites were divided in two individual parts for training and validating (80% against 20%) the classification models. The visualization of the reference data is briefly shown in Fig. 3.4.

3.3.2 Details of the field surveys

Introduction

The accuracy of a LULC map is not properly evaluated without ground-truth data. It demonstrates the importance of the ground-truth data which are only acquired from field surveys. However, a field survey for large-scale LULC studies are expensive and time-consuming. An excellent design for such a survey is essential to collect comprehensively and effectively ground-truth data. Herein, the currently available informative LULC data such as previous LULC maps, inventory data, and high-resolution images from Google Earth were used in this study. Based on these data, a map of high and low confidence areas of LULC types was created. A high confidence area meant all the data used had the same LULC type in the area. A low confidence area meant the disagreement between the different data used in terms of LULC types. Using this map, the sites of the surveys were carefully selected; of which, 70% sites were the high confident areas and the others were the low confidence areas. Then, the surveys were planned within five years. Each year, some members were assigned to conduct the surveys. In 2015 and 2016, Hoang Thanh Tung and Nasahara Nishida Kenlo conducted a survey for the northern part of Vietnam (see [69] for more details). Ta Hoang Trung and I carefully designed for a survey through the northern to southern parts of Vietnam in 2018. However, I could not go to the fields since I got a problem with my health. Truong Van Thinh, Ta Hoang Trung, and Nasahara Nishida Kenlo conducted the survey. Truong Van Thinh, Hoang Thanh Tung, and Nasahara Nishida Kenlo made another survey in 2019. In 2020, Hoang Thanh Tung and I did the final survey from Hanoi to the most southern part of Vietnam. The detailed routes of each year and the number of survey sites are presented in figure 3.4.

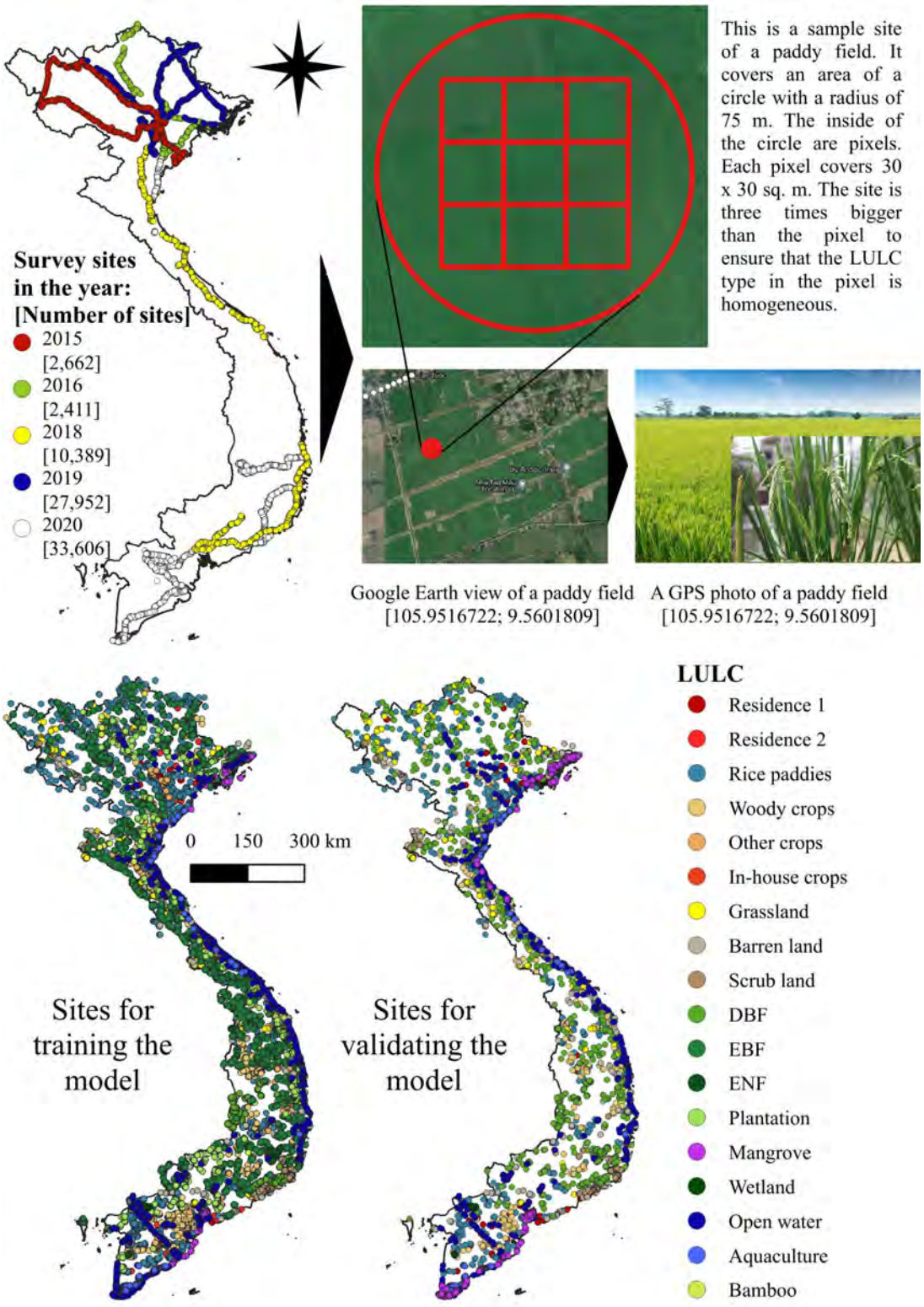


Figure 3.4: Distribution of the field survey routes and reference data of LULC in Vietnam. The field surveys were conducted in the year 2015, 2016, 2018, 2019, and 2020. Herein, the reference data are displayed for the year 2020, including training and validating sets, which are independent to each other.

Description of activities

Main activities during the surveys:

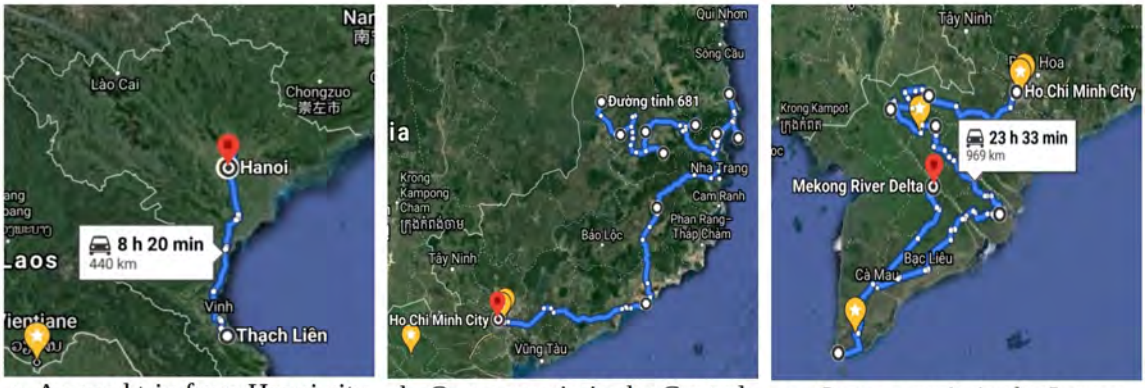
1. Taking GPS photos of the target sites. The location of the photos were published on Google Map (see Appendix A.3 for more details).
2. Conducting in-depth interviews with the locals and local experts in LULC. I recorded the interviews in the Vietnamese language. The data were stored in our server and shared upon requests.
3. Locating the target sites in Google Earth (see Appendix A.3 for more details).
4. Measuring and estimating the tree height, perimeter, and diameter.
5. Recognizing the different LULC types and their conditions.

Field surveys in central and southern Vietnam in 2020

Overview of the trip

1. Main participants: Hoang Thanh Tung and Phan Cao Duong.
2. Date: from Feb 17 to Mar 4, 2020.
3. Equipment: a memory card reader, a CASIO H20G, GoPro HERO 6, and their accessories (Fig. 3.5).
4. Purpose: The purpose of this survey is to collect empirical reference data and conduct actively individual interviews. The reference data are "ground truth" data which is vital for testing and validating LULC predicted map products what I am doing.
5. Summary: I (Phan Cao Duong) conducted an intensive survey through Vietnam namely (1) a round trip between Hanoi city and Ha Tinh province; (2) one-way trip in the Central Highlands from Dak Lak to Phu Yen, Khanh Hoa, Lam Dong and Binh Thuan (5 provinces); and (3) one-way trip in the Lower Mekong River Delta (LMRD) from Ho Chi Minh City to Dong Thap, Long An, Tien Giang, Vinh Long, Tra Vinh, Soc Trang, Bac Lieu, Hau Giang provinces and Can Tho city (Fig. 3.6). We, however, could not visit from the corner to corner of these places due to the limitation of resources. In fact, I visited planned places where I had not been sure about the LULC facts. During the trip, I used rental cars or/and public transportation services and took approximately 30,000 GPS photos. I also conducted 20 in-depth personal interviews with records in the Vietnamese language. These data contain tremendously useful information that saved on our laboratory server for further careful analyses. Although large numbers of the task have been accomplished, I would like to brief key activities and significantly surprising findings as follows.

Figure 3.5: Main equipment: a memory card reader, a CASIO H20G, GoPro HERO 6, and their accessories.



a. A round trip from Hanoi city to Ha Tinh province

b. One-way trip in the Central Highlands

c. One-way trip in the Lower Mekong Delta

Figure 3.6: The main survey routes: the red symbol is a final destination of a trip and the other end of the blue line is a starting place.

A round trip between Hanoi city and Ha Tinh province

I took a round-way bus between Hanoi city and Ha Tinh provinces. During the trip, I observed and took pictures of LULC at the target areas. I conducted interviews with locals whenever I met them. In Ha Tinh province, detailed survey and in-depth interviews were conducted in Xuan Loc, My Loc, Dong Loc, Trung Loc and Khanh Loc communes with local guides. Specifically, the population density was low in these areas and their main income earned from agricultural products. However, farming jobs have been recently less attractive to the young generation. They preferred educational advancement. As a result, most of them have been moving to crowded cities for

earning a better life. The most dominant LULC types of the visited areas are orchards, rice paddies, crops, and forests. Rice paddies are cultivated two times a year, i.e., in winter-spring season starting in December and harvesting in April, and in summer season starting from May to September (Fig. 3.7c). Three remaining months are not cultivated but for soil improvement. The area of rice paddies has been almost unchanged for 40 years. In contract, crops have been converted between corn, cassava, bean, peanut and sesame with two seasons per year (Fig. 3.7a and b). The conversion depended on the recommendation of authorities or/and policies. Although orchards are less dominant than rice and crops, they are mainly planted surrounding foothills or hills (Fig. 3.8b). Regarding plantation forests, I visited Khe Tho area (Fig. 3.8a). There is a big reservoir with acacia plantation surrounding and forest hills nearby and next to mixture natural forests (Fig. 3.8c). The acacia forests were planted on rental land from the government. These forests are harvested after 5 to 10 years after plantation, depending on the regions. The forests then are reproduced for a new season.

a. (GPS Position: 18 deg 23' 11.92" N, 105 deg 42' 55.82" E; 2020, Feb 20th)

b. (GPS position: 18 deg 23' 3.41" N, 105 deg 42' 44.87" E; 2020, Feb 20th)

c. (GPS position: 18 deg 23' 25.51" N, 105 deg 42' 57.82" E; 2020, Feb 20th)

Figure 3.7: Croplands in Dong Loc commune (Can Loc district, Ha Tinh province): a cassava field (a), a corn field (b), and a rice field (c).

a. (GPS Position: 18 deg 22' 26.60" N, 105 deg 42' 27.45" E; 2020 Feb 20th)

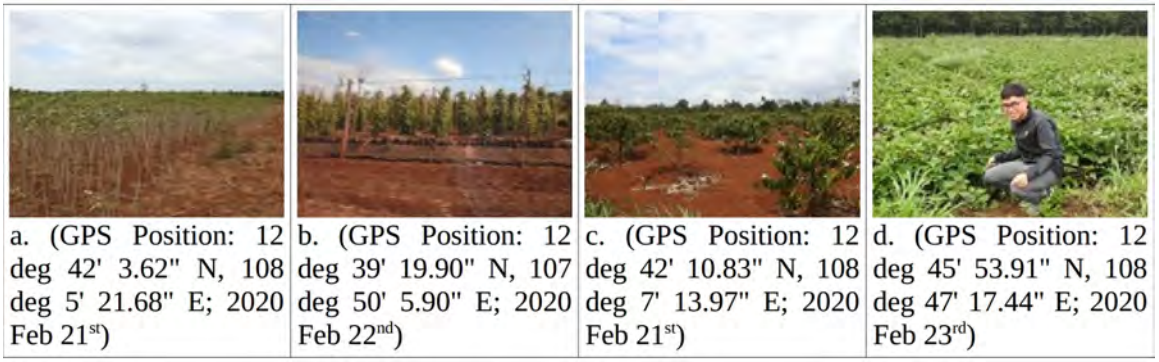
b. (GPS Position: 18 deg 22' 37.81" N, 105 deg 42' 38.59" E; 2020, Feb 20th)

c. (GPS position: 18 deg 23' 21.49" N, 105 deg 43' 48.74" E; 2020, Feb 20th)

Figure 3.8: Other land types in Dong Loc commune (Can Loc district, Ha Tinh province): Khe Tho reservoir (a), grapefruit (b), and a plantation acacia forest (c).

Surveys in the Central Highlands

In the Central Highlands, Tung and I visited five provinces, i.e., Dak Lak, Phu Yen, Khanh Hoa, Lam Dong, and Binh Thuan. Dak Lak and Lam Dong provinces were highlands with the dominant LULC types being agricultural crops or perennial crops, orchards, and forests. To be specific, Dak Lak was famous for perennial crops namely cassava, pepper, coffee (mainly Robusta, Arabica, and Liberica) and sweet potato (Fig. 3.9). Regarding the plantation process, coffee, pepper, and rubber took from three to five years to have the first full harvest while cassava and sweet potato were usually planted two seasons per year. Also, there were vast fields of forests both plantation and nature (Fig. 3.10). The main plantation forests were acacia and rubber. Rubber could reach mature trees after three years and be harvested latex lasting for 40 years whereas acacia was normally harvested and reproduced within five or ten years. Remaining natural forests were national parks such as Yok Don and Yang Sin managed by the Vietnamese government. These forests were mixed forests of deciduous and evergreen broad-leaf forests. There were similar characteristics of LULC in Lam Dong province. However, it had vast areas of natural evergreen needle leaf forest and agricultural lands such as flower and vegetation as the observation and interview with local people (Fig. 3.11).



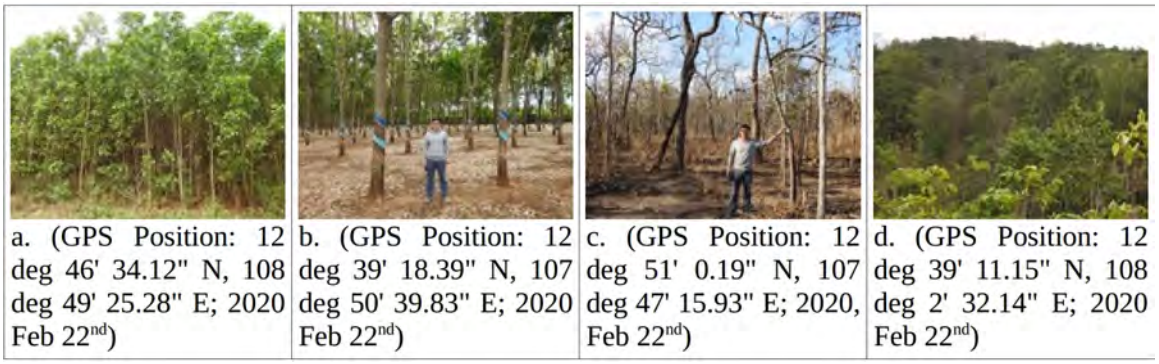
a. (GPS Position: 12 deg 42' 3.62" N, 108 deg 5' 21.68" E; 2020 Feb 21st)

b. (GPS Position: 12 deg 39' 19.90" N, 107 deg 50' 5.90" E; 2020 Feb 22nd)

c. (GPS Position: 12 deg 42' 10.83" N, 108 deg 7' 13.97" E; 2020 Feb 21st)

d. (GPS Position: 12 deg 45' 53.91" N, 108 deg 47' 17.44" E; 2020 Feb 23rd)

Figure 3.9: Croplands in Dak Lak province: a cassava field (a), mixture of pepper and coffee (b), a pepper field (c) and a rubber field (d).



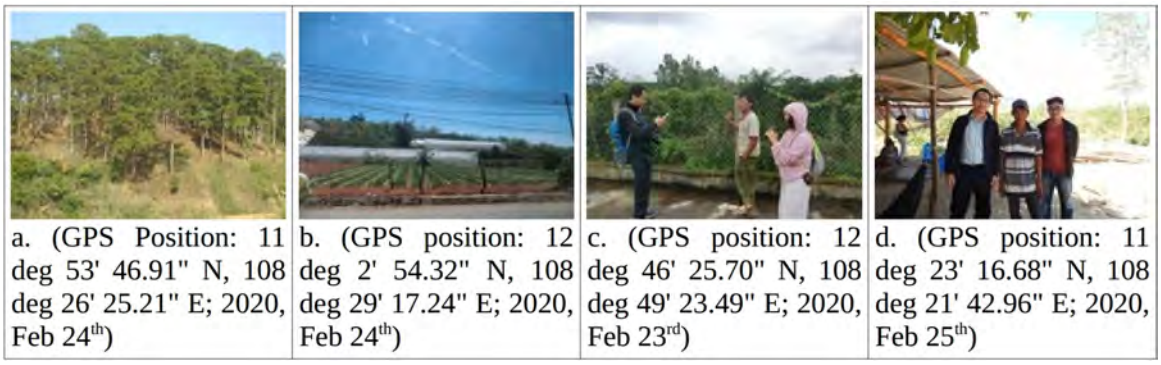
a. (GPS Position: 12 deg 46' 34.12" N, 108 deg 49' 25.28" E; 2020 Feb 22nd)

b. (GPS Position: 12 deg 39' 18.39" N, 107 deg 50' 39.83" E; 2020 Feb 22nd)

c. (GPS Position: 12 deg 51' 0.19" N, 107 deg 47' 15.93" E; 2020, Feb 22nd)

d. (GPS Position: 12 deg 39' 11.15" N, 108 deg 2' 32.14" E; 2020 Feb 22nd)

Figure 3.10: Plantation in Dak Lak province: a acacia plantation forest (a), a rubber plantation forest (b), a deciduous broad leaf forest (c) and evergreen broad leaf forest (d).



a. (GPS Position: 11 deg 53' 46.91" N, 108 deg 26' 25.21" E; 2020, Feb 24th)

b. (GPS position: 12 deg 2' 54.32" N, 108 deg 29' 17.24" E; 2020, Feb 24th)

c. (GPS position: 12 deg 46' 25.70" N, 108 deg 49' 23.49" E; 2020, Feb 23rd)

d. (GPS position: 11 deg 23' 16.68" N, 108 deg 21' 42.96" E; 2020, Feb 25th)

Figure 3.11: Surveys in Lam Dong province: a pine forest area (a), a sweet potato crop field (b), interviews with the local people (c) and (d).

In the visited areas, most local people came from different places in Vietnam. They have come and lived here since the 1990s. Their annual income was mainly earned from crops. However, some of the crop products were very cheap in the season of price drops which could sometimes take years. They sold only 5,000 VND (0.2 USD) per kilogram of fresh coffee seed but they had spent half of the money for harvest and the other half for fertilizer and initial investment. It meant they did not have profit from their crops. Due to the price fluctuation of the market economy in agricultural products, several local people abandoned or sold their croplands and went to big cities for earning a better living, especially the young generation. Others have been willing to change to other crops with a risky and significantly initial investment. More importantly, local people rarely recognized climate change and its impacts on their crops. I also visited Khanh Hoa and Binh Thuan provinces which are coastal areas. Over the visited areas, crops were mainly dragon fruit which were cultivated inland plain regions. There were subtropical shrub-land areas for coastline protection. There were still many bare lands and sand dune along the coastline (Fig. 3.12). Weather in these areas was quite hot and dry.

a. (GPS position: 11 deg 5' 50.5" N, 108 deg 22' 11.5" E; 2020, Feb 25th)

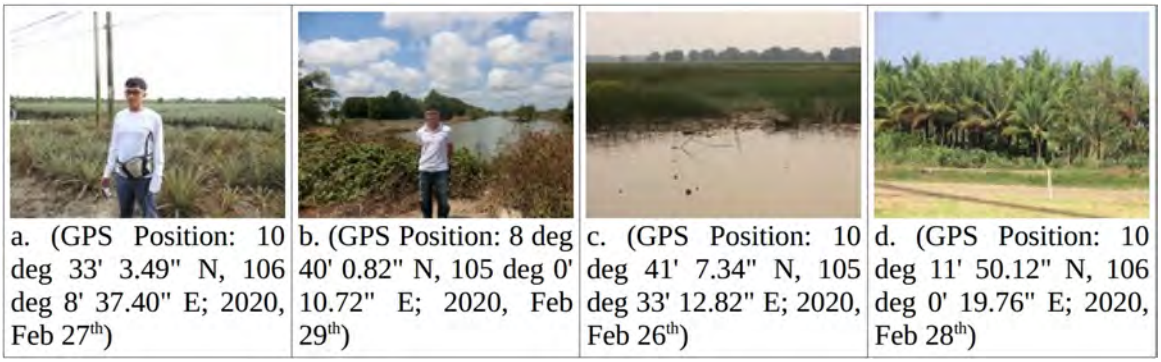
b. (GPS position: 11 deg 7' 5.90" N, 108 deg 29' 30.35" E; 2020, Feb 25th)

c. (GPS position: 11 deg 7' 4.52" N, 108 deg 29' 30.71" E; 2020, Feb 25th)

Figure 3.12: Surveys in Khanh Hoa and Binh Thuan provinces: dragon fruit (a), subtropical shrubland (b) and sand dune over visited survey (c).

The last region of the surveys was the Vietnamese Mekong River Delta (VMD) in southern Vietnam. The delta is a vast land of rivers, swamps and islands. It is the third largest delta in the world and recognized as one of the most important zone for agricultural production and biodiversity. Currently, its dominant LULC is mainly agricultural land (e.g., rice paddies, potato, pineapple, chilly and cassava), aquaculture land (e.g., shrimp, crab and fish), wetland (marsh, melaleuca and mangrove) and orchard land (Fig. 3.13). To understand the dynamics of LUCC in the region, I conducted a completed survey through 10 provinces namely Dong Thap, Long An, Tien Giang, Vinh Long, Tra Vinh, Soc Trang, Bac Lieu, Ca Mau, Hau Giang and Can Tho. Each province had its special characteristics of LULC. For example, Dong Thap was well known for lotus, rice paddies and mango while Long An was home to pineapple and potato. During the trips, I interviewed local people, in particular the elderly (maximum 85 years old). Some of them were first comers and contributed significantly to the development of the region in 1980s. I then visited their farm land (Fig. 3.14). As a result of the interviews, I understood that the Vietnamese Mekong River Delta had been a huge zone of wetland (mainly marsh) and seasonal rice paddies inland, and mangrove forests distributing along coastline before the starting of World War II. During the World War II, many of mangrove forests were destroyed by dioxin scattered by U.S. army to exterminate Vietnamese army sheltering in these areas, particularly in Ca Mau. Approximately twenty percent of the mangrove forest were destroyed. Gained independence in the year 1975s, Vietnamese government had harvested the remaining mangrove forest for sake of profit before distributing widely contaminated land, marsh land, and harvested land to local citizens. The land then had encouragingly planted mangrove forest with the free supply of seedling from the government since 1980s. Subsequently, seasonal rice had been expanded beginning in riparian zones. Seasonal rice was a natural breed of rice with their height of from 100 to 300 cm, depending on the fluctuation of the tide in the cultivation zones. Interestingly, based on these seasonal rice, the special characteristics of a region could be estimated such as the maximum inundation of the areas.

The data from surveys and interviews have shown that the LULC in the VMD has changed rapidly since the year 1990s. In 1986, a new reform known as "doi moi" (renovation) was implemented causing the initial expansion and intensification of agricultural land, especially rice paddies and aquaculture, followed by the rapid transformation of land use between mangrove forest, agriculture and aquaculture. Recently, the process of land conversion has been occurring significantly. The main reasons were owing to the fluctuation of domestic market economy and farming conditions such as water supply and saltwater intrusion. This was in agreement with the current analysis ([1, 20]). The current landscapes in the VMD are coastal mangrove forests, inland wetland of marsh-mangrove mixture, aquaculture, crops, and orchards. The Mekong Delta is currently a hot spot and has faced serious problems such as drought and salt intrusion since 2016. Other issues such as loss of mangrove forests and coastline erosion



a. (GPS Position: 10 deg 33' 3.49" N, 106 deg 8' 37.40" E; 2020, Feb 27th)

b. (GPS Position: 8 deg 40' 0.82" N, 105 deg 0' 10.72" E; 2020, Feb 29th)

c. (GPS Position: 10 deg 41' 7.34" N, 105 deg 33' 12.82" E; 2020, Feb 26th)

d. (GPS Position: 10 deg 11' 50.12" N, 106 deg 0' 19.76" E; 2020, Feb 28th)

Figure 3.13: Surveys in the Vietnamese Mekong River Delta: a pineapple field (a), shrimp and crab aquaculture farming (b), marsh and melaleuca wetland (c), and coconut orchard (d).

a. (GPS Position: 10 deg 40' 32.68" N, 105 deg 32' 26.12" E; 2020, Feb 27th)

b. (GPS Position: 10 deg 40' 29.75" N, 105 deg 31' 56.29" E; 2020, Feb 27th)

c. (GPS Position: 10 deg 40' 30.46" N, 105 deg 31' 56.44" E; 2020, Feb 27th)

Figure 3.14: Surveys and interviews in Dong Thap province: interview with the elderly and experienced locals (a), a rice paddy (b) and wetland (c).

have mentioned in various recent studies without the comprehensive analysis of causes and driving forces. For example, a severe drought and increasing saltwater intrusion significantly impacted on rice paddies and domestic water supplies in a few places such as Ben Tre and Tra Vinh. As my observations, rice fields were planned to the region of potential saltwater intrusion while saltwater aquaculture was farmed in the area of freshwater zones. The main reason of the severe drought could be a poor strategic planning. The poor strategic planning means the construction of large numbers of dikes and sluice gates along rivers' embankment. Even though these dikes and sluice gates might protect flooding during rainy season, be actively regulated water supplies and improving soil contaminated by metal, some of them could result in major issues such as the prevention of deposition process of suspended sediment into farm land, causing the lowering of the ground. Also, although water flow into the VMD decreased due to the increasing number of dam constructions in the upstream, water could be easy taken from rivers. Water from rivers, however, was restrictively used due to pollution. The pollution was normally due to different sources, but one of mainly observed sources was untreated wastewater from industrial fishery farms. I found that small producers could not use river water for their aquaculture and farming lands in Ca Mau province.

One of the most interesting activities was serious and detailed discussions with JDS alumni, friends and colleagues, particularly with leading experts from universities of Can Tho and Dong Thap. These experts were prof. Duong Van Ni, prof. Nguyen Thi Hong Diep, prof. Le Anh Tuan and Ms. Nguyen Ho. They are studying on different topics and using LULC products. They therefore have special interest in my research topics. Also, they gave us invaluable feedback and comments as well as showed us environmentally growing concerns in the VMD.

Remark and conclusion of the surveys

During the trip, I acquired a great number of lessons. I were safely back and the survey went smoothly with surprising results and findings. My survey was cooperated with nice weather, friends, colleagues, and local experts.

Regarding negatives, before the survey, I had asked help from my colleague who worked on LULC in the VMD. When I were in the Central Highlands, he said that he had got a serious sick and he could not gone with us. Surprisingly, he felt better several days later and could guide us somewhere nearby. He then asked his friends guided us the other places. These friends were essential to my survey. They introduced local knowledge and helped us approach local citizens much easier. For this lesson, I would recommend that I need a fallback plan when I conduct a field survey. In other words, it should have been better to contact as many as my friends as possible.

Secondly, as can be seen from my schedule, I did a 14-day intensive survey without a day off. I often woke up at 5 AM and got home at 11 PM. Unfortunately, on the way to Can Tho airport back to Hanoi city, I got the shortness of breath. I worried about the prehistoric pathology of my lung or maybe COVID-19. My friends took me to a clinic but they did not have a specialized doctor rather than a nurse on duty because it was Sunday. After a quick check, the nurse recommended us going to a local hospital nearby. I then went and took several measurements. A doctor requested standard measurements with X-ray, ECG, and others. Fortunately, everything was not a big problem. The doctor said that I might get tired due to be hard-working. He recommended me to take a rest. I therefore missed my flight and had to stay an extended night in Can Tho. This lesson taught me that I should not have set a long-term tough schedule.

Finally, I regretted not having visited several hotspots. Thanks to interviews with local people and experts, I found out some places where has been facing issues such as the shortage of clean freshwater, subsidence and landslides (Figure 13).

In summary, although this was a very tough trip, I greatly enjoyed it. I not only obtained a large number of reference data which were vital to my current research, but also discovered significant and surprising findings. I built a vast network with leading experts who helped us to understand how important land use/land cover information was. I must have had more tasks like this and will discover more about local knowledge of the regions. Importantly, this awesome task could not be successful without the considerable support from JICE/JICA, help from my supervisor, colleagues, friends

and local people. I valued and appreciated their help.

Chapter 4

Inadequate reflection of regional and global land cover datasets in mainland Vietnam

Abstract: Robust remote monitoring of land cover changes is essential for a range of studies such as climate modeling, ecosystems, and environmental protection. However, since each satellite data has its own effective features, it is difficult to obtain high accuracy land cover products derived from a single satellite's data, perhaps because of cloud cover, suboptimal acquisition schedules, and the restriction of data accessibility. In this study, I integrated Landsat 5, 7, and 8, Sentinel-2, Advanced Land Observing Satellite Advanced Visual, and Near Infrared Radiometer type 2 (ALOS/AVNIR-2), ALOS Phased Array L-band Synthetic Aperture Radar (PALSAR) Mosaic, ALOS-2/PALSAR-2 Mosaic, Shuttle Radar Topography Mission (SRTM), and ancillary data, using kernel density estimation to map and analyze land use/cover change (LUCC) over Central Vietnam from 2007 to 2017. The region was classified into nine categories, i.e., water, urban, rice paddy, upland crops, grassland, orchard, forest, mangrove, and bare land by an automatic model which was trained and tested by 98,000 reference data collected from field surveys and visual interpretations. Results were the 2007 and 2017 classified maps with the same spatial resolutions of 10 m and the overall accuracies of 90.5% and 90.6%, respectively. They indicated that Central Vietnam experienced an extensive change in land cover (33 ± 18% of the total area) during the study period. Gross gains in forests (2,680 km^2) and water bodies (570 km^2) were primarily from conversion of orchards, paddy fields, and crops. Total losses in bare land (495 km^2) and paddy (485 km^2) were largely to due transformation to croplands and urban & other infrastructure lands. In addition, the results demonstrated that using global land cover products for specific applications is impaired because of uncertainties and inconsistencies. These findings are essential for the development of resource management strategy and environmental studies.

4.1 Introduction

Land use/cover change (LUCC) is increasingly impacting on the Earth's surface biophysics, biogeochemistry, and biogeography at any rate or scale such as ecosystem services [70, 71, 72], water balance [73], climate [74], biodiversity conservation [75], and agriculture [76]. It means land use/cover information is important for natural resources planning and management [77]. In Central Vietnam, the land cover has substantially altered as a result of rapid socio-economic development activities over recent years [78, 79]. Future changes are also anticipated to occur [80], since the region has an economic growth rate of approximately 10% a year, which is higher than the average of Vietnam. The fast-growing economy has rapidly converted forest and agricultural lands into industrial or service zones [81]. The development also increases the region's energy requirement, followed by the mass development of hydro-power plants because of Central Vietnam's suitable geography, topography, and hydrological regime for the hydropower plants. These plants are changing the land cover around them [82]. On the other hand, natural disasters such as drought, floods, and typhoons are also causing land cover changes [83].

This change has negatively affected a variety of resources such as biodiversity, carbon sequestration [78, 81], and food security [67] over the region. Specifically, a decline in rice yields (by 30%), carbon storage (by 15%), and sequestration (by 12%) due to the expansion of infrastructure lands are predicted until 2100 [84]. The reduction of rice yields results in concern for food security; Vietnam is the second-largest exporter of rice [67]. The expansion of built-up land is projected to have influenced urban heat islands in several cities (e.g., Hanoi). Not only does the LUCC insignificantly boost the peak mean air temperature, but the number of hot spots is also growing, particularly in the new infrastructure zones [85]. In addition, Vietnam has seen a gain in forest cover that is estimated at 1696 million hectares [86], while Central Vietnam has seen a dramatic decrease in the forest due to conversion into agricultural land, resulting in the increasing emission of carbon dioxide [87]. The conversions of the forest into agriculture also have led to an increase of about 30% in surface runoff and approximately 55% in sediment yield from 2000 to 2008 in Dong Nai province [88, 89]. Hence, monitoring of the LUCC is necessary for the sustainable management of natural resources and the environment in the region and the achievement of Sustainable Development Goals (SDGs), especially the goal of Life On Land (Goal 15): "Protect, restore, and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss."

However, it is not easy to detect the land cover of a region with existing maps, i.e., global land cover maps. They seem to suffer from low accuracy and coarse spatial resolution. Specifically, Global Land Cover 2000 (GLC2000) [90], GlobCover 2009 [91], the International Geosphere-Biosphere Program Data and Information System's (IGBP) DISCover land cover [92], and the Moderate Resolution Imaging Spectroradiometer land cover (MCD12Q1) [93] have an overall accuracy of 68.6% (22 categories), 67.5% (22 categories), 66.9% (17 categories), and 78.3% (17 categories) respectively. These maps

have a spatial resolution of 300 m to 1,000 m. Even with the Finer Resolution Observation and Monitoring—Global Land Cover (FROM-GLC) [94] has a spatial resolution of 30 m — it suffers from a lower overall accuracy (67.1%). This means that global land cover maps designed for particular global purposes are likely to be uncertain and inconsistent for specific local, subnational, or national applications while most counties lack national scale land cover mappings. Mapping land cover for a specific nation with the assistance of users in that nation is designed to meet particular user requirements. This is also the ongoing project of JAXA EORC Ecosystem Research Group, which is producing country-by-country global land cover maps. The major problem of remotely sensed mappings can be because of insufficient satellite imagery available for land cover mappings, particularly due to suboptimal receiving agendas, limited data accessibility, and cloud cover. Cloud covers 65% of the global surface and 70% of the tropical surface in a year [95]. In addition, Landsat 5, 7, and 8 were designed to acquire data with a 16-day cycle but most areas have not been constantly imaged every 16 days due to the effects of seasonality, solar zenith angle, cloud cover, and because priority is given to the continental US [96]. The temporal resolution of Landsat 7, apart from the broken scan-line corrector (SLC-off; from 31/05/2003), is frequently longer than 16 days [97]. Before 2010, data accessibility is also a serious problem, even though Landsat data are freely available via the Earth Resources Observation and Science (EROS) Center, users cannot acquire this data without a brief project description that must be approved [98]. To close these gaps, a fusion of multiple remote sensing data with ancillary data is one of the best solutions.

The fusion of optical and radar satellite images (e.g., Landsat and L-band SAR) has recently been proven to be an advancement for monitoring land cover [99] and forests [100] in tropical areas. With the development of the recent European Space Agency Sentinel-1, -2, and -3 with high spatial resolution, the fusion of multiple sensors is more prevalent and effective. These new data have been effectively combined with Landsat for urban mapping [101] with the mosaics of Advanced Land Observing Satellite-2 Phased Arrayed L-band Synthetic Aperture Radar-2 (ALOS-2 PALSAR-2) for mangrove and forest monitoring [102]. Integration of multiple optical and radar sensors with different electromagnetic spectra can recognize various land cover features better than a single sensor [103]. However, most studies have integrated multiple sensors to map land cover in a certain time instead of temporal land cover changes; or to map a specific land cover type [104]. Fewer studies have estimated the advancement of data fusion among various land cover types, or used high-resolution data fusion, e.g., Advanced Land Observing Satellite Advanced Visible and Near Infrared Radiometer type 2 (ALOS AVNIR-2), for mapping temporal changes in land cover.

This study aimed to generate land cover maps over Central Vietnam from 2007 to 2017, using a kernel density estimation and remotely sensed data from multiple sensors. This research shows the potential of combining multiple remotely sensed data and ancillary data for mapping large land cover dynamics. Results improve the understanding of land cover dynamics over Central Vietnam and can contribute to resource management and policy-maker decisions. The results also demonstrate the

uncertainties of global land cover products.

4.2 Materials and methods

4.2.1 Study area

The research site is over Central Vietnam ($13^o00'$ – $20^o00'$ N, $105^o50'$ – $109^o12'$ E; Fig. 4.1a) surrounded by the ocean to the east, Laos, and Cambodia to the west, Thanh Hoa province in the north, and Phu Yen and Dak Lak provinces in the south. Its total area is approximately 95,000 km^2 with three main areas: North Central Coast, South Central Coast, and Central Highlands (with the highest elevation at 3142 m above sea level). They have a variety of landscapes from deltas, hill lands, mountainous regions, or highlands, to coastal zones with a diverse climate from humid subtropical, monsoon to tropical savanna climates. The region experiences four seasons: spring (February to April), summer (May to July), fall (August to October), and winter (November to January). The mean annual rainfall is 700–5000 mm [105] and the mean annual temperature is 23.9–25.9 ^{o}C, which significantly controls the crop seasons over the region [67]. The diverse climate, complex topography, and various ethnicities lead to complex geography and landscape with the dominant land cover types of rice paddy, crops, grassland, wetland, urban, forest, bare land, and mangrove.

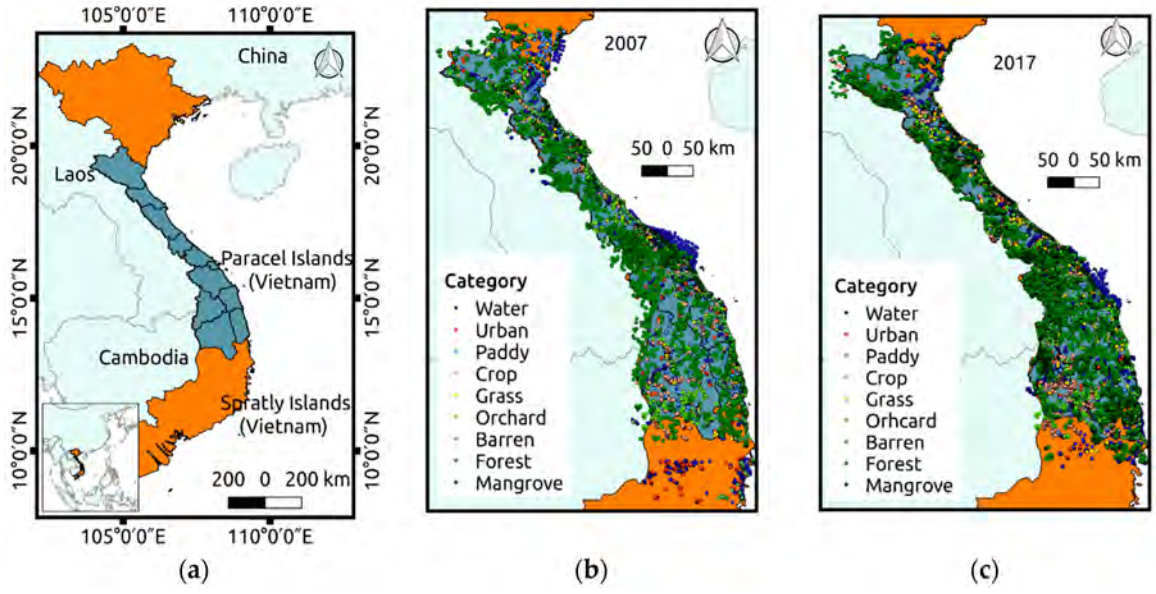


Figure 4.1: (**a**) Study area in Central Vietnam; (**b** and **c**) distribution of reference data for the year 2007 and 2017 respectively.

4.2.2 Classification scheme and reference data design

A land cover and land use category system was established to identify the dominant land cover types for certain purposes in the study region [106, 107]. Based on the local

knowledge and the Land Cover Classification System, this research used nine land cover types including water or open water, urban & built-up or residential land, rice paddy, orchards or woody crops, other crops, grassland, barren, forest, and mangrove A.4; I also referred to a past paper [108] to produce a consistent land cover map for larger scales.

I found that imbalanced stratified random sampling obtains higher accuracy than balanced sampling, which has in agreement with a past study [109]. This study, therefore, employed an imbalanced stratified random sampling to design reference data for training and testing the classifier. Approximately 3,000 reference data were collected from the field survey and about 95,000 were extracted from Google Street View, Degree Confluence Project, Mapillary, and Google Earth by using a visual interpretation (Figs. 4.1b and 4.1c). About 65% of the data was used to train the classification model, while the others were used for accuracy assessment. To achieve strict training sample standards, every extracted reference data must cover a homogeneous land cover type region with a diameter greater than 20 m. The data contains geo-location, land cover category, observation time, and the GPS photo if available.

4.2.3 Data and image preprocessing

This study used a variety of multi-temporal satellite imagery from multiple sensors (Table 4.1). For the 2007 reference year, I used the mosaic images of ALOS AVNIR-2 and ALOS PALSAR of the Japan Aerospace Exploration Agency (JAXA); the calibrated top-of-atmosphere (TOA) reflectance products of Landsat 5 Thematic Mapper (TM) and Landsat 7 Enhanced Thematic Mapper Plus (ETM+) of the United States Geological Survey (USGS). For the year 2017, I used the Sentinel-2 Multi-spectral Instrument (MSI), Landsat 8 Operational Land Imager (OLI), and ALOS-2 PALSAR-2 mosaic. All the data were divided into 24 small square tiles with each of them having a magnitude of one degree by one degree in longitude and latitude degrees or less. Data with a UTM projection were warped to the WGS84 latitude-longitude projection. A bilinear interpolation resampling to World Geodetic System (WGS84) latitude-longitude coordinates with 18.94 s × 18.94 s resolution (about $10 \times 10\,m^2$) was performed for individual data. For optical data, I preprocessed basic atmospheric corrections, cloud masking, and geometric corrections if the data had not been corrected by producers or errors were found, while the radar data were masked slope effects and filtered speckle; both image types were used for classification.

For geometric correction, while most of the Landsat L1T and Sentinel-2 products had geometric accuracy of within $\pm$ 1 pixel and $\pm$ 0.3 pixels, respectively, a few images were less accurate. The less accurate images were checked and removed based on control points derived from the Global Land Survey (GLS) 2000 data [110]. The ALOS AVNIR-2 level 1B2 product has been geometrically corrected by JAXA.

For clouds and cloud shadow masking, I used the function of a mask (Fmask) for Landsat images. Fmask employed temperature bands to identify clouds at various altitudes [111, 112]. The elevation of the cloud was later applied in shadow detection.

Table 4.1: Dataset organization, layer composition for each sensor type in each dataset, and the total number of images for each position.

Sensor Type	Year of Acquisition	Images per site	Spatial Resolutions (m)	Temporal Resolution
Sentinel-2	2017	10	10 and 60	10 days
Landsat 8 OLI	2017	8	30	16 days
ALOS AVNIR-2	2007	5	10	46 days
Landsat 7 ETM+	2007	5	30	16 days
Landsat 5 TM	2007	5	30	16 days
ALOS PALSAR Mosaic	2007	1	25	1 year
ALOS-2 PALSAR-2 Mosaic	2017	1	-	-
SRTM 1 Arc-Second Global 2000	1	30	-	-
Open street map	-	1	-	-

The shadow was identified thanks to the cloud projection and additional determination of shadow pixels. The Sentinel-2 and ALOS AVNIR-2 images, however, did not have such thermal bands, which were used to mask clouds and cloud shadows. They faced difficulty with clouds and cloud shadows masking [112]. Hence, RGB and NIR (and SWIR for Sentinel-2) bands were used to detect clouds; the ratio of blue and green reflectance was employed to identify shadow [113]. Although the method did not use thermal bands, its performance reached similar accuracies to the VIIRS Cloud Mask [114, 115] and VIIRS I-band Cloud Mask [116] methods which used thermal bands.

To enhance the accuracy of the optical images, I calculated optical indices that were primarily served to identify different land types in this research. While the Enhanced Built-Up and Bareness Index (EBBI [117]), the Normalized Difference Built-Up Index (NDBI [118]), the Urban Index (UI [119]), and the Normalized Different Bareness Index (NDBaI [120] were employed for distinguishing built-up and bare land. The Normalized Difference Vegetation Index (NDVI [121]), Enhanced Vegetation Index (EVI [122]), Soil Adjusted Vegetation Index (SAVI [123]), and Normalized Difference Water Index (NDWI [124]) can be used for the sake of promptly monitoring vegetated regions in complex heterogeneous landscapes to discriminate water, croplands, plantations, and forests [125, 126].

The ALOS PALSAR and ALOS-2 PALSAR-2 Mosaic included dual-polarized (HH and HV) channels. To increase the effectiveness of land cover classification [127], each channel was filtered speckle using a Refined Lee filter with the European Space Agency Sentinel Application Platform Toolbox v.6.0.0 (ESA SNAP; available online: http://step.esa.int/main/toolboxes/snap/). The digital number (DN) of HH and HV channels were converted into sigma naught values in decibel units (dB) using the

Table 4.2: The reference of spectral band information Landsat 5, 7 and 8, Sentinel-2, and ALOS AVNIR-2 used for the calculation of optical indices.

Data	Channel	Spectral range (μm)	Electromagnetic Region
Landsat 8	Band 1	0.435–0.451	Coastal Aerosol
	Band 2	0.452–0.512	Visual Blue (VBlue)
	Band 3	0.533–0.590	Visible Green (VGreen)
	Band 4	0.636–0.673	Visible Red (VRed)
	Band 5	0.851–0.879	Near Infrared (NIR)
	Band 6	1.566–1.651	Short Wave Infrared (SWIR1)
	Band 7	2.107–2.294	Short Wave Infrared (SWIR2)
	Band 10	10.60–11.19	Thermal Infrared (TIR)
Landsat 5 and 7	Band 1	0.45–0.52	Visual Blue (VBlue)
	Band 2	0.52–0.60	Visible Green (VGreen)
	Band 3	0.63–0.69	Visible Red (VRed)
	Band 4	0.77–0.90	Near Infrared (NIR)
	Band 5	1.55–1.75	Short Wave Infrared (SWIR1)
	Band 6	10.40–12.50	Thermal Infrared (TIR)
	Band 7	2.09–2.35	Short Wave Infrared (SWIR2)
Sentinel-2	Band 1	0.433–0.453	Coastal Aerosol
	Band 2	0.458–0.522	Visual Blue (VBlue)
	Band 3	0.543–0.578	Visible Green (VGreen)
	Band 4	0.650–0.680	Visible Red (VRed)
	Band 8	0.785–0.899	Near Infrared (NIR)
ALOS AVNIR-2	Band 1	0.42–0.50	Visual Blue (VBlue)
	Band 2	0.52–0.60	Visible Green (VGreen)
	Band 3	0.61–0.69	Visible Red (VRed)
	Band 4	0.76–0.89	Near Infrared (NIR)

following Eq. (4.1).

$$\sigma^0 = 10.\log_{10}(DN^2) + CF \tag{4.1}$$

where σ^0 is the radar backscatter per unit area, and CF is the calibration factor (CF = 83.0 dB) [128].

For anthropogenic activity detection, I took advantage of SRTM 1 Arc-Second Global and OpenStreetMap (OSM) data collected from the National Aeronautics and Space Administration (NASA) and OpenStreetMap Foundation (OSMF), respectively. The SRTM 1 Arc-Second Global data were used to extract topography information, such as slope, whereas the level of human activity was identified by extracting a rasterized distance map to the road network using the OSM.

All the preprocessing steps were automatically completed by employing C++ and python with the support of the Geospatial Data Abstraction Library (GDAL), Ge-

ographic Resources Analysis Support System, Geographic Information System v.7.2 (GRASS—GIS), and Quantum Geographic Information System v.2.18 (QGIS).

4.2.4 Classification method

The overall flowchart included three main stages: image preprocessing, image classification and accuracy assessment, and change analysis (Fig. 4.2). Following a previously described method [129], this study employed Bayesian logic together with a kernel density estimation (KDE) [130]. More specifically, I estimated the probability density function of features (such as vegetation indices, band reflectance, etc.) for each category by generating small Gaussian functions around training data (which is KDE) in the feature space, and combined them to generate posterior probability using Bayes' theorem. After the computation of the posterior probability of every land cover category for each image, I generated the joint posterior probability from all the overlapped images acquired from different periods and sensors. The final choice of the category was based on the highest joint probability among all land cover categories. This method is suitable for detecting the seasonal change of land surface (phenology) as a classification key in a large area due to its fully-automatic robustness. More importantly, this KDE approach is more accurate than support vector machines and the maximum likelihood classification. I offer a detailed description of the processes as follows.

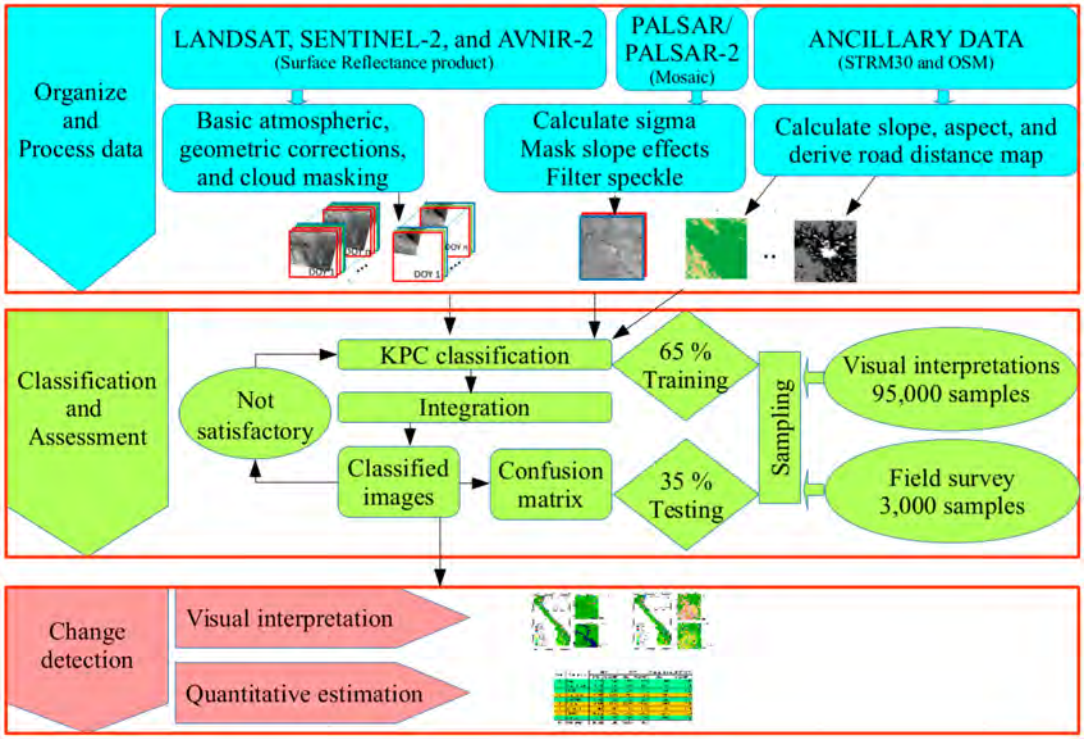


Figure 4.2: Overall flowchart of land cover/use change monitoring and analysis in this study.

The posterior probability of a category was computed based on a D-dimensional vector of input data x, i.e., spectral bands of reference and two-dimensional values

representing observation date $[t_1, t_2]$ (Eq. (4.2)).

$$[t_1, t_2] = \left[\cos\left(2\pi\frac{DOY}{DOY_{\max}}\right), \sin\left(2\pi\frac{DOY}{DOY_{\max}}\right)\right] \tag{4.2}$$

where DOY is the date of the observation year (Julian day), and DOY_{max} (=365.25) is the average maximum day of the years.

For every image, the posterior probability of a category C_k ($k = 1, 2, \ldots, M$; M is the number of land cover categories; $M = 9$) was determined by using the Bayesian rule based on the input data x (Eq. (4.3)).

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)} = \frac{p(C_k)p(x|C_k)}{\sum_{k=1}^{M} p(C_k)p(x|C_k)} \tag{4.3}$$

where $p(C_k)$ is the prior probability of C_k (which is assumed to be a uniform distribution), and $p(x|C_k)$ is a category-conditional probability of x; $p(x|C_k)$ was estimated based on the training data using kernel density estimation (KDE). KDE is used to compute the probability distribution of data as the sum of kernel functions that are of the same form and centered on each training data, by using the Gaussian kernel (Eq. (4.4)) and Scott's rule of thumb (Eq. (4.5)) as follows.

$$p(x|C_k) = \frac{1}{N_k}\sum_{n=1}^{N_k}\left\{\prod_{d=1}^{D}\frac{1}{h_d}K(\frac{x_d - x_{n,d}}{h_d})\right\} \tag{4.4}$$

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-u^2}{2}\right) \tag{4.5}$$

$$h_d = N^{\frac{-1}{D+4}}.\sigma_d \tag{4.6}$$

where, N_k is the number of training data of a category C_k, h_d is a bandwidth parameter estimated by Eq. (4.6), N is the total number of training data ($N = N_1 + N_2 + ... + N_M$), and σ_d denotes the standard deviation of d-th dimension of training data $\{x_{n,d}|1 \leq n \leq N\}$.

In the next step, at each pixel for each category, I integrated the posterior probability of all overlapped images by multiplying the posterior probability of all images. In general, the higher the joint posterior probability of a category, C_k, the more possible the land cover category C_k is. However, in reality, even if the true land cover is category C_k, the $p(C_k|x)$ of one image sometimes might be close to or equal to zero due to noise, cloud, or insufficient training data. If it occurs, it would make the joint of the posterior probability of category C_k also close to or equal to zero, because multiplying by zero always gives zero. This means that even if the $p(C_k|x)$ of the most overlapped images are as high as 1, the final prediction of the land cover cannot be in the category C_k. To overcome this issue, the posterior probability of each image should not be too close to zero. To realize it, I used Eq. (4.7) ([131]. The final probability of a category C_k, $p'(C_k)$ was estimated by Eq. (4.8).

$$p'(C_k|x) = ap(C_k|x) + \frac{1-a}{M} \tag{4.7}$$

$$p^{'}(C_k) = \prod_{i=1}^{S} p_i^{'}(C_k|x_i) \tag{4.8}$$

where a is a constant value ($a = 0.7$), and S is the number of images.

The final choice of a category is the category with the highest joint probability among all land cover categories. Supposedly, at a pixel r of a classified land cover map with two categories, i.e., water and urban, has the joint probability of water: $p^{'}(C_{water})$ = 0.6 and the joint probability of urban: $p^{'}(C_{urban})$ = 0.4. The highest joint probability of pixel r is 0.6 and the land cover of pixel r is water. To shorten data processing and complex landscapes, I extracted imagery data into 24 small square tiles, each of them having a magnitude of 1^o # 1^o longitude and latitude degrees or lesser (Fig. 4.2). The classification process was performed independently for each tile using Saclass Software version 1.7 developed by the University of Tsukuba and JAXA [132, 129].

4.2.5 Accuracy assessment

Assessment of classification accuracy of 2007 and 2017 maps was the most important part to determine the quality of these maps. While there was a large number of accuracy measures used to estimate algorithm performance, it is crucial to carry out accuracy estimation for any category if the classification results are valuable for changing detection [133]. We, therefore, chose a confusion matrix [134] because it is one of the most general methods that is easy to understanding and has useful values. An imbalanced stratified random approach was used to extract approximately 27,000 reference data from ground truth data and visual interpretations. These reference data represented all the designed land cover categories of the region. In addition, a Kappa coefficient of agreement was used to measure the range of single classification accuracy [135]. It is presented by the following formula (Eq. (4.9)).

$$K = \frac{p_o - p_e}{1 - p_e} \tag{4.9}$$

where p_o is the observed proportional agreement between actual and predicted categories defined as $p_o = \frac{1}{n}\sum_{i=1}^{M} f_{ii}$ and p_e is the expected agreement by chance defined as $p_o = \frac{1}{n^2}\sum_{i=1}^{M} f_{i+}f_{+i}$, where M is the number of land cover categories, f_{i+} is the total for the ith row and f_{+i} is the total for the ith column in the confusion matrix.

4.3 Results

The classified maps and areas of land cover change within the 10 years are shown in Fig. 4.3 and their accuracy assessments (confusion matrix) are shown in Tables 4 and 5 of this paper [16]. The overall accuracies of the maps for 2007 and 2017 are 90.5% (kappa

coefficient of 90%) and 90.6% (kappa coefficient of 90%), respectively. Most categories have accuracy for users and producers greater than or close to 90%, except for grass and orchards. Water, bare land, paddy, and forest have the highest accuracies that are over or close to 95%, followed by urban and crops that account for approximately 91% and 90%, respectively. Orchards and grassland have the lowest accuracies (< 85%) in the two maps. The reason for misinterpreted classification of orchard and grassland may be the correspondent spectral characteristics between orchards, grass, and the other categories.

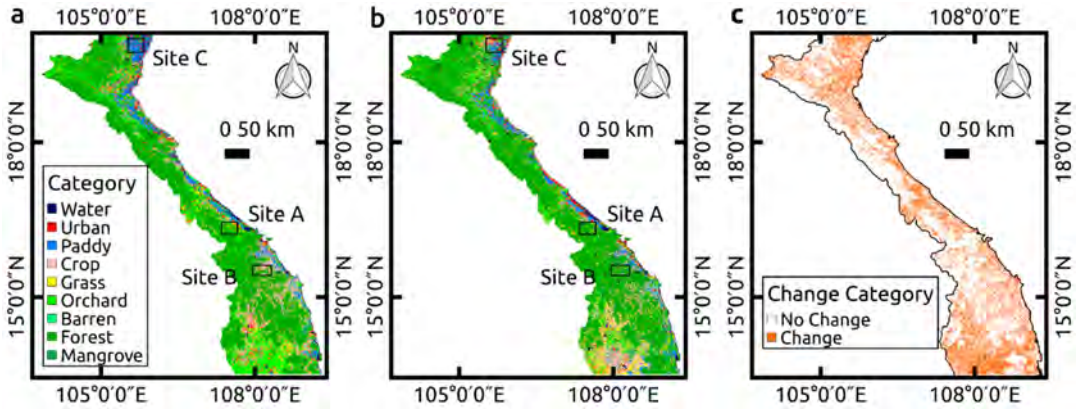


Figure 4.3: Land cover maps in (**a**) 2007, (**b**) 2017, (**c**) areas of land cover change within the 10-year period over Central Vietnam, and A, B, and C are the selected sites for change analysis in Thua Thien Hue, Quang Nam, and Thanh Hoa provinces, respectively.

Central Vietnam experienced an extensive change in land cover from 2007 to 2017 (Figure 4.3c). A total of 31,380 $\pm$ 16,920 km^2 (33 $\pm$ 18% of total area) changed with the major changes occurring in orchards, forests, and croplands in the coastal areas and central highlands. In 2007, central Vietnam covered an estimated area of about 94,000 km^2 with 93% of the total area being vegetated areas and the other being water bodies, bare land, and urban & built-up. The vegetated area included dense tree cover, i.e., forest (about 51,000 km^2) and mangrove (1,000 km^2), and dynamic land cover, i.e., paddy fields (6,500 km^2), grassland (2,800 km^2), cropland (9,900 km^2), and orchards (15,500 km^2). Over the recent decade, water, urban & built-up, cropland, grassland, forest, and mangrove areas increased by approximately 560, 40, 1,100, 2,680, and 930 km^2, respectively, while bare land and paddy fields decreased the similar amount of about 500 km^2. Surprisingly, the period witnessed a sharp decline in orchards (4,600 km^2).

4.4 Discussion

It is not easy to create an accurate map for a cloudy area such as Central Vietnam, a tropical region, where clouds cover 70% of the area over the year [95]. The methods were customized to Vietnam's unique conditions by integrating a variety of satellite images from multiple sensors with ancillary data. This is an effective approach for

land cover mapping for cloudy and large regions as the result of the availability of free satellite images and the development of image processing and computational power. For instance, by using an automated image preprocessing approach, with an automated classification method, and the support of servers, I could produce a 10 m spatial resolution land cover map over Central Vietnam (94,000 km^2) within approximately 3 days. Much more effort was frequently required for the careful work of data organization, field survey, and land cover change estimation. Further effectiveness and efficiency of this approach can be explained as follows.

The research showed the potential of combining multi-sensor remote sensing data for land cover classification and change detection in tropical areas. The outperforming of data fusion (e.g., Landsat and L-band SAR) over individual sensors for improving overall accuracy has been used for land cover monitoring in tropical regions such as Indonesia [136, 137] and West Africa [138], which obtained similar overall classification accuracies. Nevertheless, these researches mapped land cover for smaller areas (< 900 km^2) compared with the current study (94,000 km^2). Note that this study corroborated the research by Hoang et al. [108], which used the combination of multisensor data (i.e., Landsat 5 and 8 and ASTER-VA version) for analyzing land cover in Northern Vietnam that demonstrated the potential of the fusion of multiple sensors for mapping heterogeneous landscapes. However, my maps are better than their maps in terms of the overall accuracy and spatial resolution (90.5% and 10 m vs. 81.0% and 15 m). These two maps have the same number of LULC categories (nine categories). A possible explanation for this might be that I used finer spatial resolution images (10 m against 15 m), more reference data (98,000 against 65,000), and the improvement of posterior probability integration (Eq. (4.7)). Another explanation is that, instead of using 4 bands (Blue, Green, Red, and Infrared), I used the best combination of bands and a set of spectral indices.

Surprisingly, while most categories have users' accuracy (UA) and producers' accuracy (PA) higher than 90%, a few limitations can be found in the accuracy of grass, crops, and orchard categories (Table 4 and 5 [16]). Specifically, grass and orchard have the lowest UA while misclassification between grass, crop, and orchard has occurred in both years. A possible explanation for this might be that orchard and crop categories include a large variety of orchard and crop types leading to a significant variance in spectral reflectance patterns. Another probable reason may be that several grasslands for raising cattle are cultivated as croplands (temporary crops followed by harvest and a bare soil period), confusing whether they are grassland or cropland. The misclassification between crop, orchard, and urban may be due to the specialty of the traditional Vietnamese farm, a form of domestic agriculture in which food gardening, fish rearing, and animal husbandry are wholly combined [139, 140]. However, these systems are frequently used in small complex areas of mixed land cover/land use, causing difficulty in distinguishing land cover. To overcome the misclassification, it may be worth considering the use of very high-resolution remotely sensed data (e.g., Ikonos, QuickBird, and Kompsat-2) [141, 142] or classifying these categories into different subcategories before integrating them into a single one. The issue of the misclassification between the

mixed and complex land covers is an intriguing one that could be explored in further research.

4.4.1 Uncertainties of global land cover maps over central Vietnam

With a spatial resolution of 10 m and accuracy of 90.5% (Kappa coefficient: 0.9), my maps are better than the existing maps which are considered to be of coarse spatial resolution (30 – 1,000 m) and low accuracy (< 80%; Fig. 4.4). Fig. 4.5 compares my maps with (**a**) Climate change initiative (CCI) 300-m land cover V2 for the year 2015 released by the European Space Agency (ESA), (**b**) the GlobeLand30 map for the year 2015 published by the National Geomatics Center of China, (**c**) the MCD1Q1 0.5 km MODIS-based global land cover climatology for the year 2001–2010 published by the USGS, (**d**) the Global PALSAR-2 25-m Forest/Non-forest map for the year 2017, and (**e**) the Global PALSAR 25-m Forest/Non-forest map for the year 2007 from JAXA based on visual interpretation.

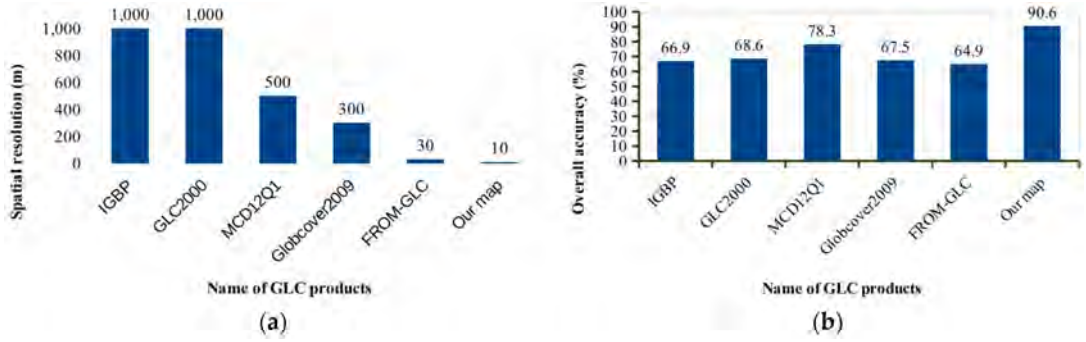


Figure 4.4: A comparison of my maps and previous global land cover maps in the spatial resolution (**a**) and in the overall accuracy (**b**) over Central Vietnam.

Results show that global land cover maps seem to be uncertain and inconsistent. To be specific, the CCI map tended to misclassify most inland water, whereas underestimated urban & built-up areas and overestimated croplands (Fig. 4.5a). Although the GlobeLand30 could detect well inland water, it was likely to underestimate urban & other infrastructure lands and overestimate grassland (Fig. 4.5b). These issues may be the result of using coarse spatial resolution satellite images, cloud cover, or reference data shortcomings. The MCD1Q1 0.5 km MODIS-based global land cover climatology tended to misinterpret most inland water as wetland regions and overestimate cropland (Fig. 4.5c), perhaps because of the difference in land cover type definition or the use of very coarse spatial satellite imagery (500 m). For forest estimation, based on Google Earth view and the 2017 forest map in this research, I found that FNF maps probably misclassified some forest areas and could not accurately detect inland water, since a large number of reservoirs disappeared on the map (Figs. 4.5d and 4.5e). These problems may be the result of using only SAR images (ALOS PALSAR or ALOS-2 PALSAR-2). Although the SAR images are not blocked by clouds or cloud shadows, it often suffers from speckle which can be reduced by using noise reduction filters,

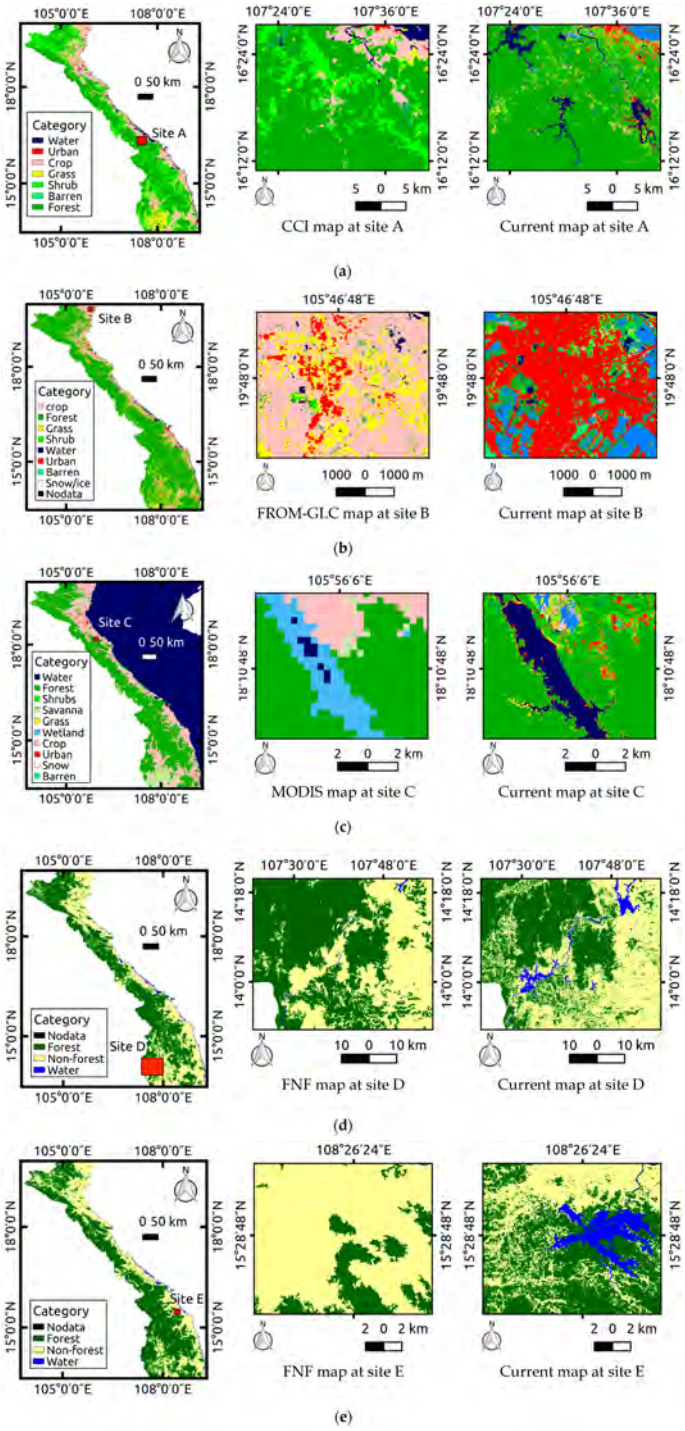


Figure 4.5: A comparison of current maps and the existing global land cover maps over Central Vietnam, using visual interpretations: (**a**) Climate change initiative (CCI) 300-m land cover V2 for the year 2015 released by ESA; (**b**) GlobeLand30 map for the year 2015 published by the National Geomatics Center of China; (**c**) MCD1Q1 0.5 km MODIS-based global land cover climatology for the years 2001–2010 published by the USGS; (**d**) Global PALSAR-2 25-m Forest/Non-forest map for the year 2007, and (**e**) Global PALSAR 25-m Forest/Non-forest map for the year 2007 from JAXA.

however, still constraining classification accuracy [143]. In summary, global land cover maps contain large uncertainties for environmental studies.

4.4.2 Ten-year land cover change over Central Vietnam

Central Vietnam has a heterogeneous landscape that experienced rapid and extensive changes from 2007 to 2017. To observe the mass conversion of land use in Central Vietnam due to recent socioeconomic transformation [144, 145], three testing locations were chosen to detect qualitative changes between 2007 and 2017. The testing locations are sites A, B, and C in Thua Thien Hue, Quang Nam, and Thanh Hoa provinces, respectively (Fig. 4.6). In site A, many reservoirs have been constructed, which is a common phenomenon over Central Vietnam. These reservoirs converted orchard to water surface while many neighboring forests changed to crops. This finding is in agreement with other satellite analysis [146, 147] and can explain why orchards decreased while croplands increased over the recent decade. Site B illustrated a shift from croplands to forests, probably the result of recent government policy to reforest some parts of Vietnam by providing financial and technical resources [148, 149]. This forest gain also agrees with other satellite analysis [150] and demographic statistics [151] showing the forest area increased by 1.696 million hectares on the national scale from 2005 to 2015. Site C presents the change of paddy to crops or urban & built-up areas. This could explain the decrease of paddy fields and the increase of croplands in the region. Another reason for the decrease of paddy fields can be from conversion to aquaculture because of decreasing rice productivity as the result of the intrusion of saltwater [82].

The region has been experiencing extensive changes, particularly in the decrease of paddy fields and the increase of inland water surface that can be detected easily based on this study. These changes may have generated unprecedented new ecosystems that impact environmental sustainability and food security. Results show that more than 21 huge dams have been constructed at upstream rivers (e.g., Huong, Vu Gia—Thu Bon, Dong Nai, and Sre Pok) and many more are now planned. These dams can block suspended sediment from upstream areas, which may cause large-scale shoreline erosion and land loss. Also, the construction of upstream dams restricts downstream river flow leading to a decrease in water level at estuaries, while the sea water level is expected to rise [151, 152]. This can also result in severe erosion and intense saltwater intrusion in lowland areas, followed by the expansion of salinity effects on plant growth and yield such as rice [153, 154], and the conversion of rice to aquaculture or other lands leading to the decrease of rice productivity. Because Vietnam is the second largest exporter of rice, domestic food production and international rice trade are likely to be at risk unless a sustainable development strategy is considered shortly.

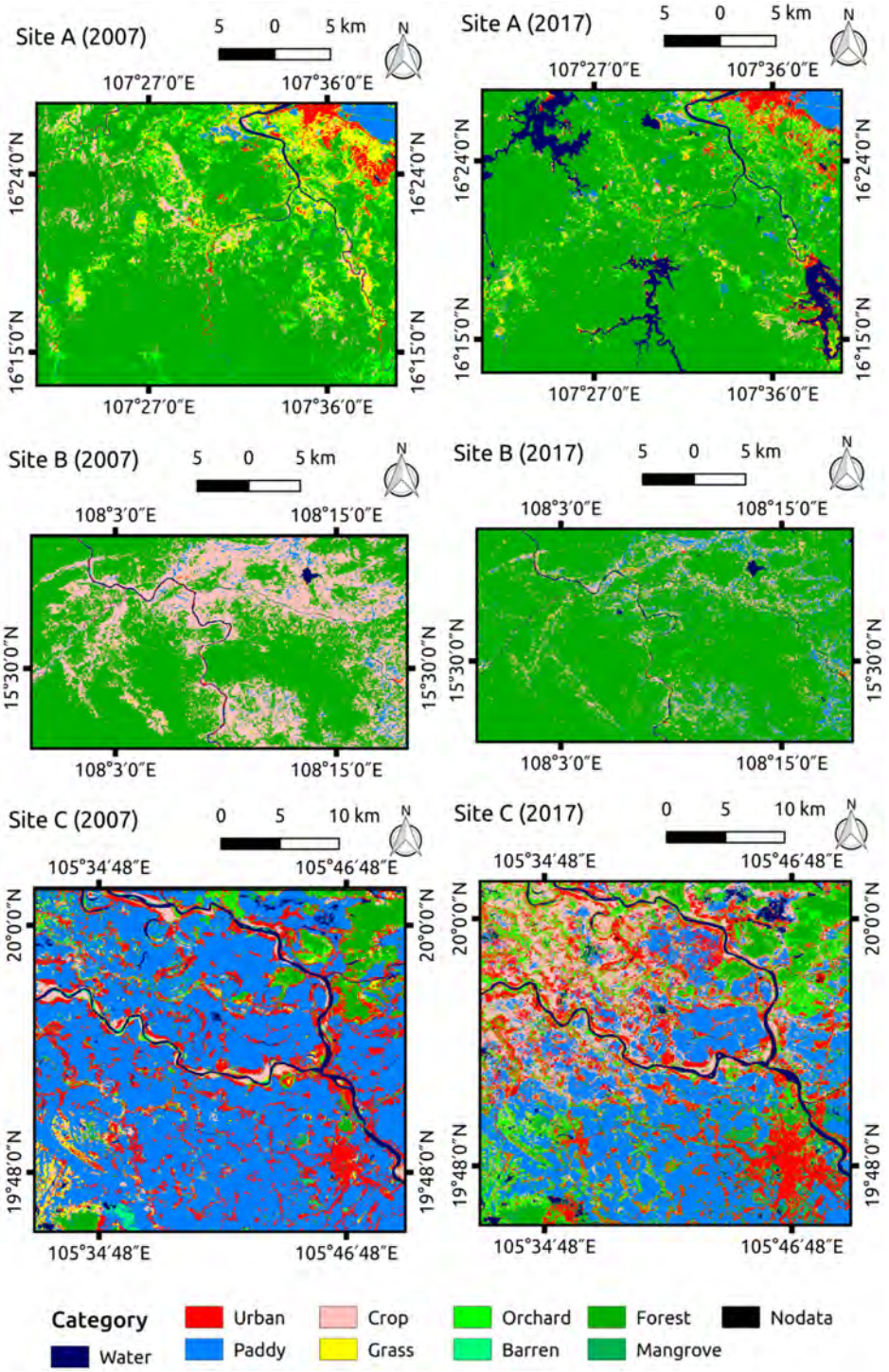


Figure 4.6: The selected sites for land cover change detection for the period 2007 to 2017 over Central Vietnam; Site A, B, and C are in Thua Thien Hue, Quang Nam, and Thanh Hoa provinces, respectively.

4.4.3 Potential application and future work

This study significantly emphasizes the importance of data fusion in remote sensing field that has been applied in recent LULC mapping studies [136, 138]. As in the above

discussions, Central Vietnam, a cloudy area, has been a challenging area for mapping with LULC, especially with high spatial resolution, due to the scarcity of data availability, which can be customized by my approach. Such high spatial resolution maps of the heterogeneous and large-scale areas have rarely been carried out, particularly in the year before 2015 (when Sentinel-2 was launched). This study made use of ALOS AVNIR-2 to generate a past-time 10-m LULC map, which may serve as a baseline map to compare observed changes or critical data for local or national long-term land use planning.

The findings and maps presented in this research can be used for multipurpose applications. First, the construction of hydroelectric reservoirs can result in problems such as shoreline erosion, salinity intrusion, extreme water level variations, and sediment delivery issues, which have been occurring in the region. Also, the region is sensitive and vulnerable to the influences of climate change and consequent sea level rise. These problems link to land cover/land use changes. While several mapping projects have been using coarse resolution satellite imagery for land cover change detection, they may not be effective for a complex and fragmented landscapes such as Central Vietnam [155], and my maps seem to be more suitable for land cover change analysis over Central Vietnam, and can provide the policy-makers and scientific associations with input data for the further discussion of environmental management, in particular water balance, sediment estimation, and food security. In addition, the maps can be a critical data for managing ecosystems and biodiversity such as for Global Forest Resources Assessment (FRA) of FAO. They may also serve as baseline maps for other land cover/land use projects such as the "Land Use Status, Change, and Impacts in Vietnam, Cambodia, and Laos" of NASA. Although my maps achieved a certain level of overall accuracy, they may be insufficient for a quantitative analysis of changes with a certain level of statistical significance. To improve the accuracy, new satellite data (e.g., Sentinel-1), or full polarimetric SAR data should be considered for next steps. On the other hand, the combination of this approach with others may be necessary. Finally, due to the uncertainties of global land cover products, it may be better to create individual national-scale maps and combine them into a global map instead of generating a whole global map.

4.5 Conclusions

Based on the kernel density estimation, I produced land cover maps of the over Central Vietnam between 2007 and 2017 using high-resolution remotely sensed data from multiple sensors. These maps (nine categories) have a spatial resolution of 10 m and an overall accuracy of 90.6% (kappa coefficient 0.9). This accuracy and spatial resolution are higher than that of existing land cover maps which tend to have a coarse resolution (30 m to 1,000 m) and low accuracy (< 80%), causing uncertainties for users. This study indicates the potential of multisensor fusion for monitoring land cover dynamics in a cloud and large area.

Results of this study show that although global land cover products are fundamen-

tal variable for global specific applications, there remains a considerable amount of uncertainties and inconsistencies for particular applications at local and national scales which could be solved by using products in this study.

Anthropogenic pressures on the land cover system over Central Vietnam are growing because of the rapid socioeconomic development process. Over the recent decade, forest areas have significantly expanded due to government efforts to reforest by changing policy and providing technical resources. However, the quality of forest in Central Vietnam remains a mystery, and a major concern in ensuring forest management in Vietnam or the Sustainable Development Goal (SDG) 15.1.1. Urban & other infrastructure areas have expanded around crowded cities such as Thanh Hoa, Vinh, Hue, and Da Nang due to population growth and the movement of citizens from rural areas to urban regions. Population growth is also accompanied by an increasing demand for water, domestic and industrial irrigation, and hydropower, resulting in the expansion of inland water surface. These changes may damage environmental sustainability, particularly by shoreline erosion, land loss, and salinity intrusion. The findings of land cover dynamics together with an interpretation of driving factors can provide policy-makers and scientific associations with appropriate input data for the further discussion of land environment management.

Chapter 5

Ensemble learning updating classifier for accurate land cover assessment in tropical cloudy areas

Abstract: Land use/cover information is fundamental for the sustainable management of resources. Notwithstanding the advancement of remote sensing, analysts daunt to generate sufficient-quality land use/cover products due to dense-cloud-contaminated and/or technical issues. This study proposes a novel approach (Ensemble Learning Updating Classifier/ELUC), which can be applied with various classification algorithms and data sets to simplistically generate new classifications or renew existing classifications with a remarkable accuracy improvement. Applying miscellaneous features of Landsat-8 images, the ELUC of a random-forest-based algorithm produces sequences of single-time classifications with a mean overall accuracy of 84%. Through the study period, these sequences of individual classifications were then joined to achieve a final classification that reaches an overall accuracy of 94%. Also, the ELUC of the random-forest-based algorithm outperforms that of Kernel-Density-Estimation with a 5% overall accuracy higher. These outcomes confirm the effectiveness of the ELUC for a remarkably consistent land use/cover estimation in a data-rich environment.

5.1 Introduction

Information on land use/land cover (LULC) dynamics is a key to the sustainable management of natural resources. The opening of data archives such as Landsat (Land Satellite) [156] and Sentinel-1, -2, -3 and -5P [157, 158, 159] provides a data-rich resource of remotely sensed imagery. The following is a considerable amount of studies on integrating multi-data sources both optical and Synthetic Aperture Radar (SAR) images such as Moderate Resolution Imaging Spectroradiometer (MODIS), Landsat, Sentinel, and Advanced Land Observing Satellite (ALOS) [160, 161, 16]. New algorithms have been developed for monitoring land surface changes, e.g., Spatial Temporal Adaptive Algorithm [162], Apply Change-vector Analysis in Posterior Probability Space (CVAPS)

[163], Continuous Change Detection and Classification (CCDC) [164], and Automatic Land Cover Classification Method (ALCC) [165]. Together with the development of new techniques, special programs have been designed. For example, Kempeneers et al. used a combination approach of spatio-temporal remote images to monitor the dynamics of global forests [166]. Hansen et al. produced a 12-year global forest cover assessment using millions of satellite images and complex algorithms [18]. The algorithms were applied to perform challenging tasks, for example, data integration, cloud and cloud-shadow masking, and atmospheric correction. In addition, numerous projects have been applied for broad-scale multi-category land cover assessments. At 10-m resolution, several attempts have been made to publish the 13-land-cover-category maps of Europe using a great set of Sentinel-2 images [167, 168] and the 10-m resolution global land cover map [169]. For a coarser spatial resolution (30 m), there are noticeable multi-class land cover products including the National Databases of the United States [170, 171, 155, 172] and the GlobeLand 30 global product with ten dominant land cover categories [173].

These products are essential to various environmental studies, but they may not meet the major requirement of numerous potential users. First, predefined land cover maps do not always meet the specific objectives of current projects. For instance, a classification of orchards can be significant to some scientific communities while it is not a distinguishing class in most existing land cover products. Also, the low consistency and frequency of the regional and global maps cannot accurately reflect local regions [16, 17]. Users, therefore, may wait for, or create and update more consistent land cover products for their interests [174, 175, 176]. As mentioned in the previous paragraph, such tasks require various advanced stages from careful image selections to appropriate reference sampling designs, proper image-processing technique selections, and specialized approaches to change detection, which might daunt potential analysts.

To date, there has been little agreement on what is the most appropriate approach to comprehensively mapping and monitoring land cover in a cloudy large area. A variety of deep neural networks with high performance has been proposed, but they are possibly not for large-scale land cover estimates. It is possible that deep learning requires a huge number of annotated information, which is normally restricted in the field of large-scale multi-class land cover assessments [177]. That is, deep learning is mainly applied for a single classification such as urban mapping [178], crop-type detection [179], and forest assessment [180]. Few studies have focused on a comprehensive land cover classification [175, 181]. Meanwhile, the blend of multiple data extracted from time-series information demonstrates its superiority over a single-time analysis [16, 17]. Time-series information can detect various temporal phenology to be fruitful for complex multi-class land cover analyses [182]. However, time-series approaches significantly suffer from dense cloud cover, especially in tropical cyclone climate regions.

A much-debated question is whether a method can be tolerant of dense cloud-contaminated problems, and thus considerably increase the data density. I, therefore, propose such an approach for less specialized users to map, update, and evaluate

land use/cover dynamics at a regional scale. I seek to examine the performance of an ensemble learning method, based on random forest (RF) theory that has played a significant function in the remotely sensed history [183]. This approach is named *"Ensemble Learning Updating Classifier"* (ELUC). Unlike the direct evolvement of various time-series data into a single-classification process, I used Landsat 8 OLI images to independently develop a stack of initial probabilities, before they are continuously updated to achieve another stack of posterior probabilities, and thus robustly increase the classification accuracy. The independent classification allows the ELUC to handle various spatio-temporal data sources while decreasing the serious impact of feature differences among the various data sources. The ELUC can also ingest images, which may be most contaminated by clouds but there is still valuable information, to critically increase the density of data availability over the assessment period. Also, the stack of probabilities is expected to be exploited for several applications. They can be utilized to estimate the classification confidence of a pixel in each classification step or to create a sequence of time-series classifications that may be applied for a rapid response to changes in land use/cover such as tracking fire areas.

The specific objective of this study was to lay out a detailed application of such an approach for a hotspot in the Vietnamese Mekong Delta (VMD). The main tasks are: (1) creating and updating a sequence of time-series classifications over the study period, (2) comparing the performance of the proposed approach based on the RF algorithm with the Kernel Density Estimation algorithm; (3) detecting land use/cover types over the study area, providing a typical and standard method for dealing with other operations; I also apply this method for the whole VMD to consistently compare our classification results and the previous land cover products, and also illustrate the potential of this method for a broader-scale application.

5.2 Materials and methods

5.2.1 Overview of the ELUC

Ensemble Learning Updating Classifier (ELUC) is developed for the continuous update of land cover categorizations over time, based on a RF algorithm (Fig. 5.1). For each pixel in a step, the ELUC generates an independent stack of prior probabilities corresponding to each of the land cover classes. If there is only one classification (here termed a *"moment classification"*), the highest prior probability of a matching land cover is utilized to label for that specified land cover at that pixel. Likewise, if there are many images in the study period, a time-series sequence of the prior probabilities is joined to produce another stack of posterior probabilities. Then, the highest posterior probability in the stack determines the corresponding land cover class. Finally, an ELUC classification is generated.

To begin this process, I present an example of continuous 10-moment integration in detail, demonstrating the potential accuracy improvement of the ELUC over time. Then, a comparison between the performance of the ELUC using the RF and Kernel Density

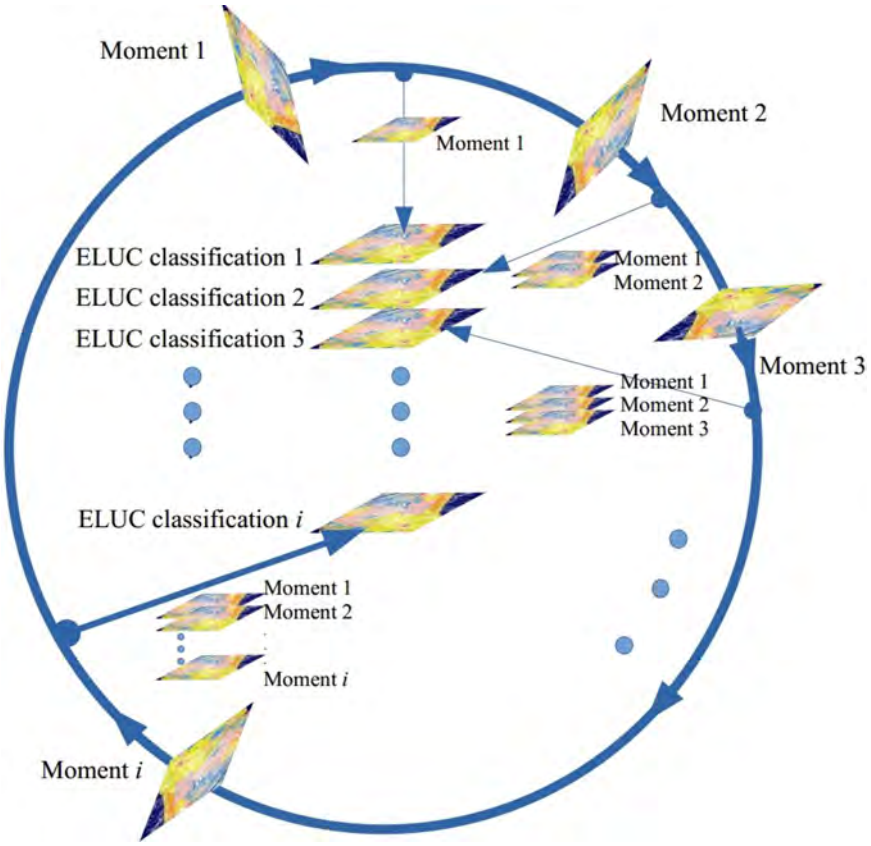


Figure 5.1: Overview flowchart of the ELUC approach. A moment i (i = 1, 2, . . . , and 10) is a single-time classification. The ELUC classification i is a constant integration classification of the moments 1, 2, . . . , and i.

Estimation (KDE) algorithms is conducted. Following this step, I produce a whole square-tile classification of 1° x 1° longitude and latitude, using all the informative Landsat 8 OLI images over the study period.

5.2.2 Study area

The study area spreads between N $9^o0'$ - $10^o0'$ and E $105^o0'$- $106^o0'$ (Fig. 5.2), including parts of Soc Trang, Bac Lieu, Ca Mau, Kien Giang, and Hau Giang provinces, five of 13 provinces in the VMD known as the third largest delta in the world. The area has low altitude and flat topography with higher in the North West and lower in the South East. Due to the East Sea being nearby, the area has a favorable subtropical climate condition with a rainy season from May to July and a dry season from August to April. Thanks to this perfectly natural environment, the region has become a biological treasure trove and the most productive area in agricultural and aquaculture production in Vietnam, resulting in a high dynamic change in LULC. In the North East, there is a huge area of paddy fields while most aquaculture farms are located in the South West close to the coastal areas. The following major land cover types are orchards alongside rivers whereas there are fragmented regions of inland wetlands and mangrove forests in the South. In summary, this region is dominated by agriculture, wetlands, orchards, built-

up areas, aquaculture, open water, and mangrove forests (See Appendix A.4 for further details). Recently, this area has suffered from serious problems associated with climate change, such as the inland expansions of saltwater intrusion due to sea-level rise and subsidence. There is a clear link between these issues and land use/cover changes in the region [184, 185]. However, the consistent assessment of the land use/cover dynamics is challenging due to several reasons, for example, dense cloud contamination, and dynamic, complex, and fragmented landscapes.

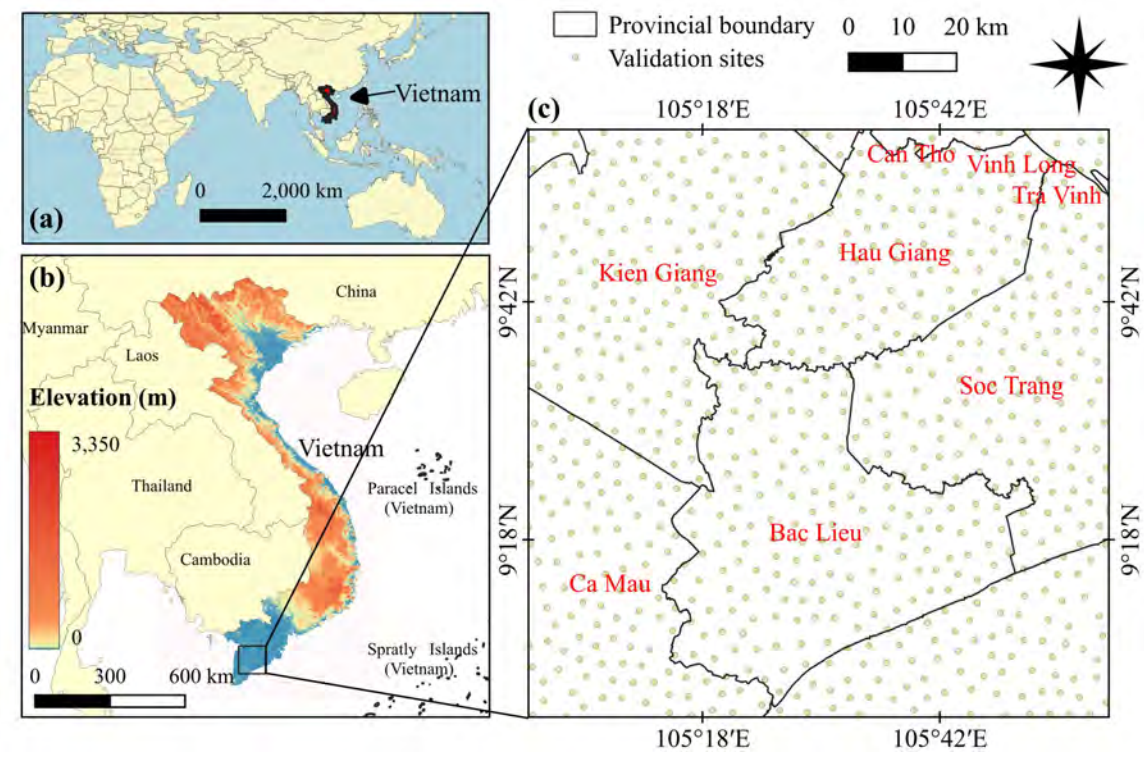


Figure 5.2: The location of Vietnam in the world (**a**); the location of study area within the boundary of Vietnam (**b**); the zoomed area and provincial boundaries in the Vietnamese Mekong Delta (**c**.

5.2.3 Image preprocessing

This seaside location has a dense cloud cover throughout a year, causing impossible to select a single-occasion cloud-free satellite image covering the entire area. I, therefore, collected all informative Landsat 8 OLI Level 1 products over the study period (Table 5.1). After collection, minimal preprocessing was applied, which aims to examine the limitations of the ELUC in this study. In other words, selected images were quickly handled for registration accuracy and atmospheric correction, but left solar-angle correction and topographic radiometric correction unprocessed. The following task is the simplistic masking of clouds and cloud shadows based on the Quality Assessment Band of Landsat 8 OLI Level 1 and the cloud-contaminated pixels were eliminated or masked as *'No data'*. Then, the images were re-projected to the Universal Transverse Mercator

(UTM) projection (Zone 48, WGS-84 datum) and a band stack was constructed from the original bands of each Landsat 8 image, except for Band 8 – Panchromatic and Band 9 – Cirrus. Eleven spectral indices were produced to enhance the ability to classify and build data-rich information. All data were automatically processed by using open source libraries (e.g., GDAL) and python programming.

Table 5.1: Landsat 8 OLI imagery, its original band stacks and spectral index stacks.

Sensor	Acquisition data (yyyymmdd)	Moment	Data stack
Landsat 8 OLI	20181031	1	Band
		2	Index
	20181202	3	Band
		4	Index
	20190204	5	Band
		6	Index
	20190308	7	Band
		8	Index
	20190409	9	Band
		10	Index

The band includes: Band 1 – Coastal aerosol, Band 2 – Blue, Band 3 – Green, Band 4 – Red, Band 5 – Near Infrared, Band 6 – SWIR 1, Band 7 – SWIR 2, Band 10 – Thermal Infrared, and Band 11 – Thermal Infrared 2.
The index includes: NDVI [186], SAVI [123], NDBI [118], NDBaI [120], NDWI [187], UI [119], EVI [188], EBBI [117], WRI [189], NDTI [190], and NDPI [191].

5.2.4 Prior probability calculation for each land cover

The ELUC was developed based on a random-forest-based method using a multitude of decision trees [192]. For each pixel, prior probability values belonging to each of the land cover categories were calculated for each satellite image. The input data were an n-dimensional vector of spectral bands or spectral indices. Specifically, the ELUC algorithm employed a bootstrap aggregation technique. It used a random subset of n samples from m samples of the total training data set ($n < m$) and a random subset of feature variables to train a decision tree to become a classifier. Each classifier then voted for a category label based on given training information while the RF algorithm had multiple classifiers. Subsequently, the number of votes was converted into prior probability values and the majority of votes or the highest prior probability value decided the category label for this step.

The most common parameters to run the RF includes the determination of the number of feature variables and the number of decision trees [193, 194, 195, 196]. Base on the high recommendation in extensive studies [197, 22], 100 tree classifiers are used

in this study.

To examine the performance of ELUC, the Kernel-Density-Estimation-based algorithm was also employed to estimate the prior probability values with the same approach and input data. The KDE [130] is one of the most effective procedures for monitoring multi-class land cover at a large scale [198, 16]. The detailed presentation of the KDE can refer to [130].

5.2.5 Updating moments

Unlike using only single-time classification, a practical method was used to automatically create time-series classifications. This approach took advantage of imperfect classifications and overcame missing data issues to improve classification performance. For each pixel, the ELUC estimated posterior probability values belonging to each of the particular land cover categories by multiplying the prior probability values of moment classifications over the given period (see Section 5.2.4. Then, the highest value of the posterior probability corresponding to a specified land cover category was used to label the predicted category at that time step. In order to minimize misclassification, an post-classification optimization was used. Specifically, the estimated prior probability of a land cover category, say $p(C_k)$, might achieve almost zero or zero due to poor-quality data or *'No data'* of the pixel at that moment. If it happens, the posterior probability of that land cover category will be equal to zero or close to zero. In other words, even though the prior probability of most moments equals 100% voting for the specified category, the updating probability product of this pixel might be zero, causing a misclassification. To handle this issue, the prior probability of a particular land cover category must not be extremely small. To this end, a careful experiment was involved to modulate prior probability values as Eq. (5.1) [131]while posterior probability values were calculated as Eq. (5.2).

$$p'\left(C_k\right) = a * p\left(C_k\right) + \frac{1-a}{L} \tag{5.1}$$

$$p_p\left(C_k\right) = \prod_{i=1}^{M} p_i'\left(C_k\right) \tag{5.2}$$

where $p^{'}(C_k)$ is the modification of prior probability value of land cover C_k; a is a constant value ($a = 0.7$) which was evaluated by numerous experiments; L is the total number of land cover categories ($L = 7$); $p_p(C_k)$ is the posterior probability value of category C_k; and M is the number of moments ($M_{\max} = 10$).

5.2.6 Accuracy assessment

Accuracy assessment was estimated following good guidance [63]. On-site intensive surveys with local and remote sensing experienced experts were conducted to complete

a definitive accuracy estimation for the individual classifications and time-series classifications. First, reference data were randomly obtained for each land cover class. I then used the spatiotemporal high-resolution images in Google Earth to interpret quickly land cover types. Finally, I confirmed the true land cover types by on-site surveys. However, the positions, that were removed due to the cover of clouds, cloud shadows, and other noise, were not estimated for that moment's classification. A confusion matrix was employed to estimate the performance of the classifications with different metrics, including user's accuracy, producer's accuracy, overall accuracy, standard error of the mean, and Kappa coefficient [199].

5.3 Results

5.3.1 Time series individual pixel

At each moment, a set of probability values was estimated for each pixel. In other words, a pixel reflected seven probability values corresponding to seven land cover categories. The highest probability of the set determined the land cover category to that moment. For visualization purposes, I introduced seven series of probability stacks corresponding to the seven ground-truth land types (Fig. 5.3). Results of the moment and ELUC classification are presented in Fig. 5.4.

As shown in Fig. 5.3, the probability values were subject to considerable fluctuations, ranging from zero to almost 100% over the moments. A zero value illustrated a *'No data'* or a data gap at that moment (e.g., moments 7 and 8 in Fig. 5.3a) while a high value (e.g., moments 10 in Fig. 5.3a) indicated the full confidence of the model to vote for the corresponding land cover. The data gap could be a result of the clouds, cloud shadows, and haze masking removals. Despite the same location, the model could either effectively perform or fail to detect the ground-truth land type, depending on the moment. For example, from Fig. 5.3c It can be seen that the model successfully detected orchard land in moment 1, but it failed to distinguish orchard from the other lands in moments 7 and 8. Likewise, the model showed more frequent signs of confusion in Figs. 5.3e and f. On the contrary, the outstanding performance of the model can be seen in Fig. 5.3b with a perfect classification of wetland from other lands over the period. The confusion might be caused by a low-quality pixel of that moment due to cloud-contaminated issues, whereas a cloud-free pixel allowed the model to immediately recognize the true land cover class.

5.3.2 Accuracy over time of the ElUC with RF algorithm

The results of moment and ELUC classification maps can be compared in Fig. 5.4. What stands out in the figure is the constant increase of the ELUC classifications over time. Although some moment classifications (e.g., moments 3 and 4) severely suffered from cloud contamination, they did not dilute the ELUC performance. Rather, the ELUC achieved noise-free classifications, especially the final map (ELUC 10). To

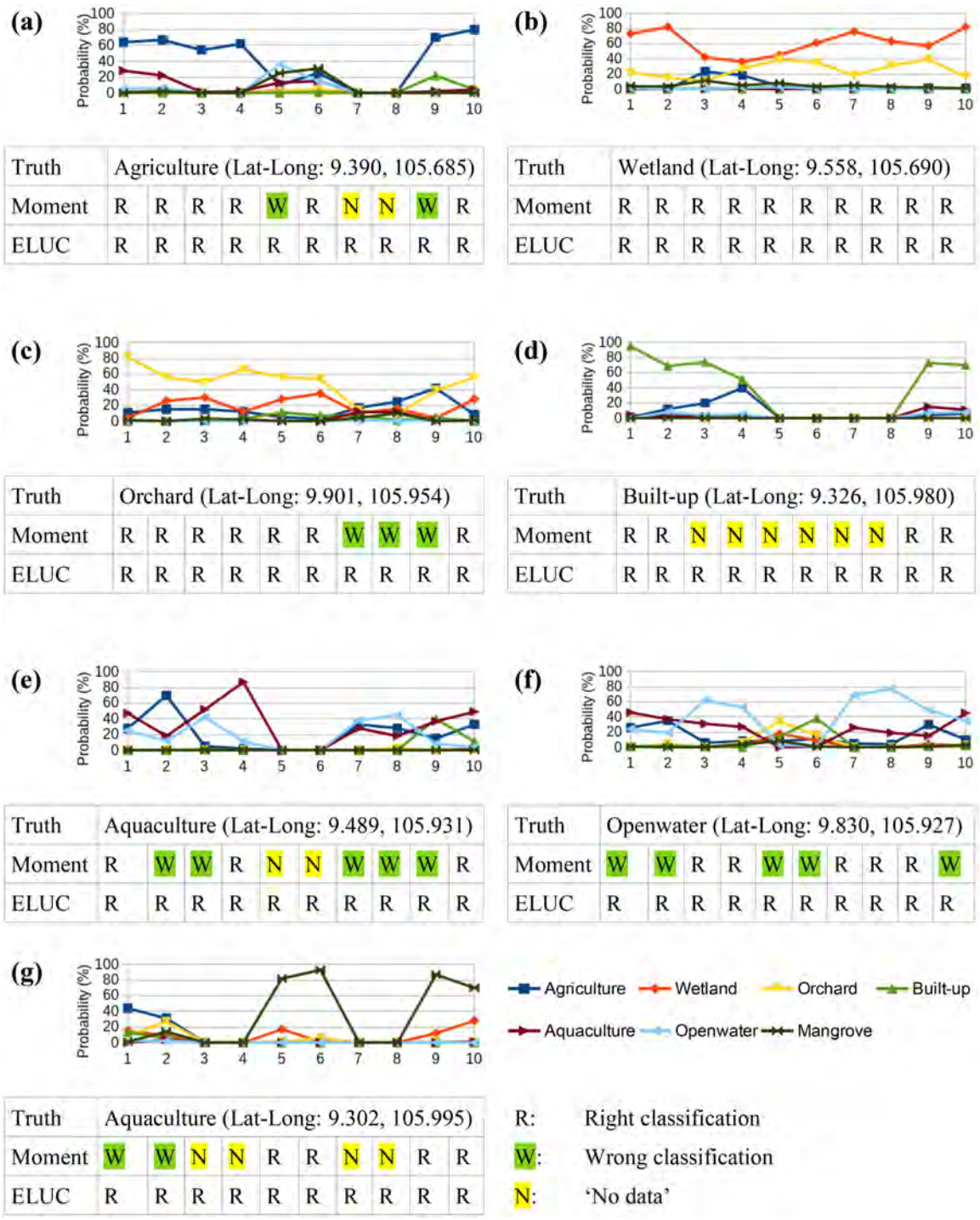


Figure 5.3: Probability assessments (%) of 10 moment classifications in seven given ground-truth sites corresponding to each land cover. The horizontal axis represents the order of moments which are classifications of given inputs, either band stacks (1, 3, 5, 7, and 9) or spectral index stacks (2, 4, 6, 8, and 10). These stacks are produced from Landsat 8 OLI images. The below are ground-truth land cover, the results of moment classifications and ELUC classifications, using RF algorithm.

quantify the improvement, I estimated the overall accuracy of both moment and ELUC

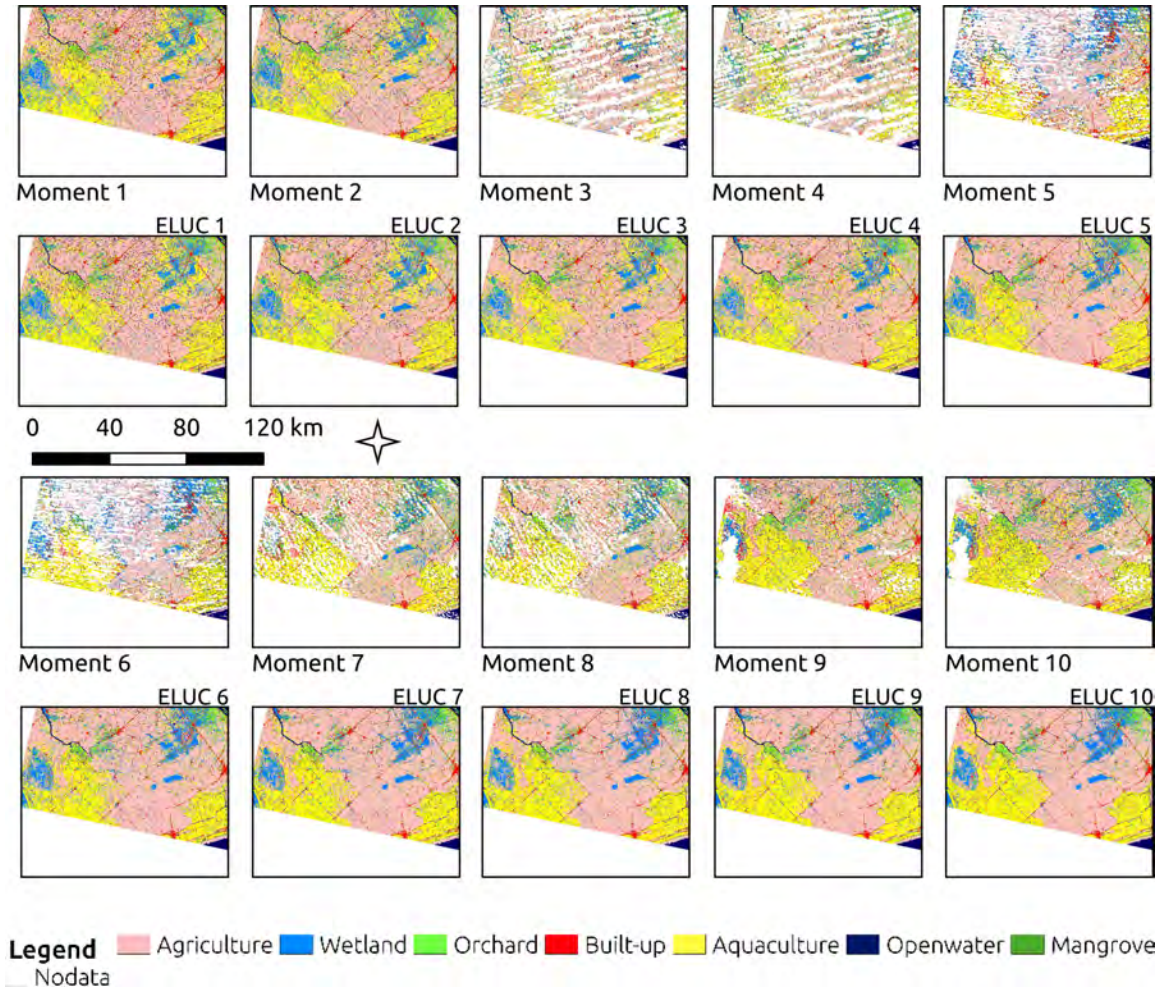


Figure 5.4: A sequence of ten moment classifications and ten ELUC classifications derived from Landsat 8 OLI images over a dry season (8/2018 – 4/2019) in a test area of the Vietnamese Mekong Delta.

classifications and provided the results in Fig. 5.5. The overall accuracy of the moment classifications fluctuated between 80% to 87% while a substantial increase from 87% to 94% was seen in the overall accuracy of ELUC classifications. It shows a rapid increase when a better moment was integrated, but a nearly steady trend when a worse moment was incorporated. Across ten-moment classifications, the ELUC achieved an average accuracy 5% higher than that of the moments in this research (90% compared with 85%).

5.3.3 Comparison between Random Forest and Kernel Density Estimate algorithms

To illustrate the outstanding of RF, I compared the performance of the Random Forest (RF) and Kernel Density Estimate (KDE) with the same approach. The results of these methods are shown in Fig. 5.6. Interestingly, the RF achievement was significantly better than the KDE algorithm with an average accuracy of 9% higher (85% compared

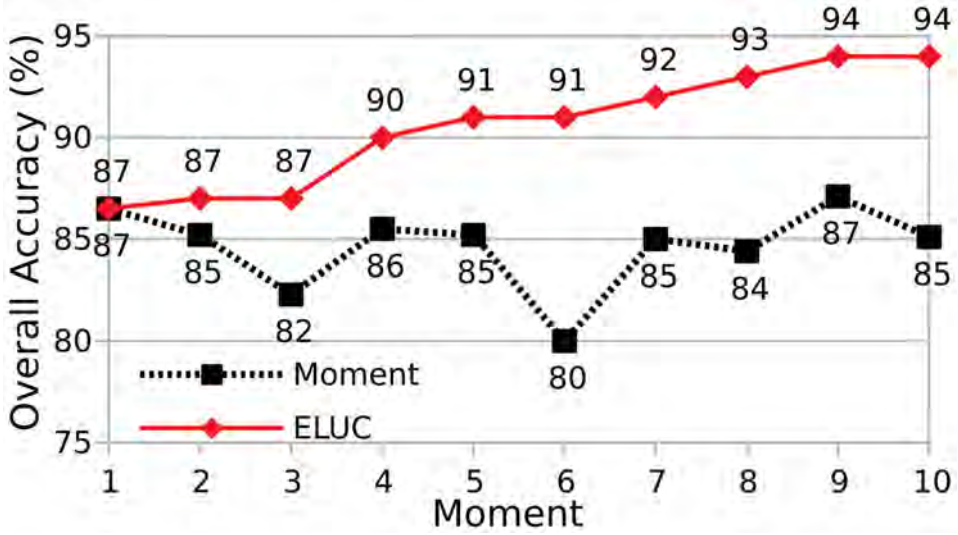


Figure 5.5: Comparison between the overall accuracy of moment classifications and ELUC classifications with RF algorithm over the study period. The moment is a single classification while the ELUC is a continuous update of all up-to-the-moment classifications. The horizontal axis represents the order of moment and ELUC classifications.

with 76%). What stands out in the figure is that spectral indices (moments 2, 4, 6, 8, and 10) played major importance in the KDE classification, but they were not likely important for the RF. For the KDE, the overall classification accuracy of the band stack obtained only 75% in moment 1, but that of spectral indices derived from the same image reached 82% in moment 2. In contrast, the RF obtained an accuracy of 87% and 85% in moments 1 and 2, respectively. This opposite result can be seen in the other index-band pairs.

5.3.4 Land cover classifications

Herein, I provide a classification of an entire square tile of 1° x 1° longitude and latitude from an update of 10-moment classifications, using both the RF-ELUC and KDE-ELUC. I present the preliminary classification of the first step in Fig. 5.7 and the final results in Fig. 5.8. The maps are clear and noise-free. Overall accuracy, user's accuracy, producer's accuracy, and Kappa coefficient are presented in Tables 5.2 and 5.3. Both the KDE-ELUC and RF-ELUC algorithms performed well and achieved an overall accuracy of 88% and 94%, respectively. In Table 5.2, the KDE-ELUC could perfectly classify mangrove and agriculture from other land classes (100% and 98% in user's accuracy), but it showed a high degree of misclassification in some land types, especially perennial crop/orchard (75% in user's accuracy). However, the RF-ELUC achieved a user's accuracy of over 90% with all classes.

As mentioned in the literature review, it is challenging to monitor land cover for a large-scale cloudy region, using remotely sensed data. I, therefore, applied the ELUC for the entire VMD to consistently compare our findings and the existing literature. It also illustrated the potential of the ELUC for a broader-scale application. The result of

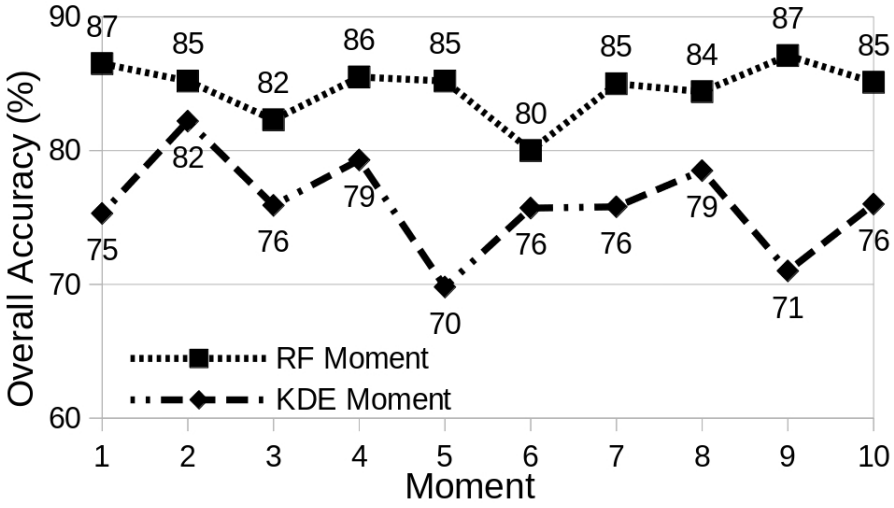


Figure 5.6: A comparison between the performance of the Random Forest and Kernel Density Estimate through moment classifications. The horizontal axis represents the order of moments which are classifications of given inputs, either original band stacks (1, 3, 5, 7, and 9) or spectral index stacks (2, 4, 6, 8, and 10). These stacks were produced from Landsat 8 OLI images.

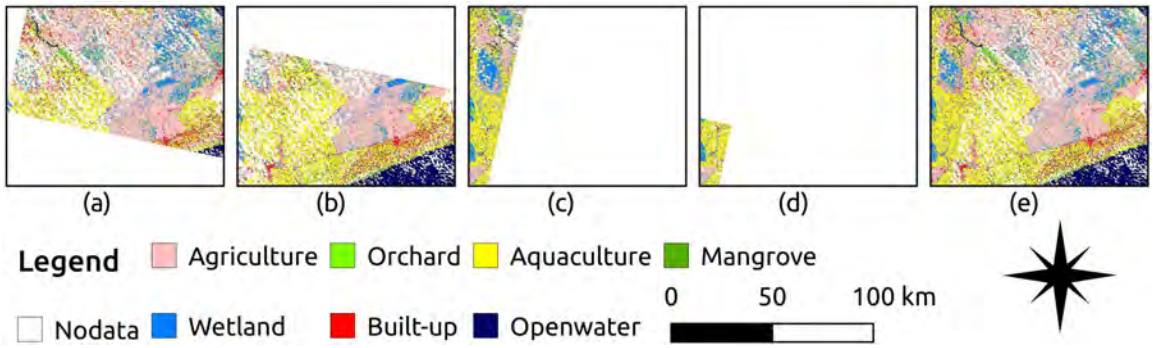


Figure 5.7: The preliminary classification of 4 moments extracted from the intersection of four Landsat 8 OLI images (**a**), (**b**), (**c**), and (**d**) within the study area square, and the mosaic classification of the four moments (**e**).

the VMD land cover product and the comparison of this product with the most updated land cover products in the VMD were presented in Text S1. I also compared this study with national statistical survey data. Interestingly, this study was relatively outstanding in comparison to the previous land cover products (with an overall accuracy of 95% compared to approximately 94%, 85%, and 79%). It was the fittest product to the national survey data (See Text S1 [20] for further details).

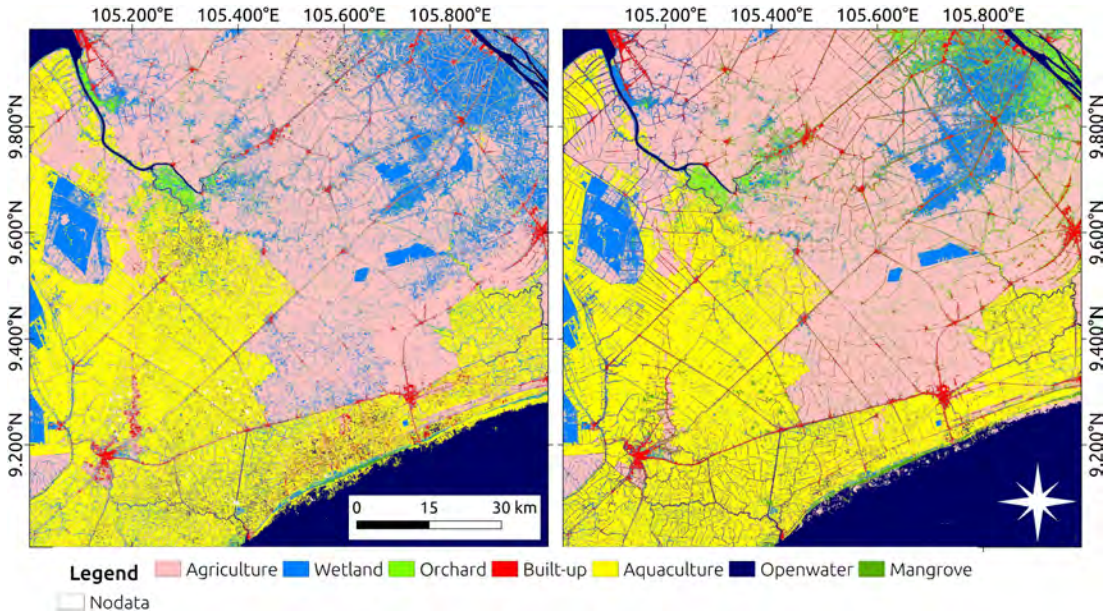


Figure 5.8: The final classification of ten-stage integration in the study area square using the Random Forest in the left and the Kernel Density Estimate in the right. These results were created from Landsat 8 OLI images.

Table 5.2: Accuracy estimate of the KDE-ELUC land cover classification using a confusion matrix.

		Land cover map							
		AG	WL	PC	BA	AQ	OW	MA	Total
Reference data	AG	322	10	5	6	9	0	0	352
	WL	2	252	14	0	9	0	0	277
	PC	3	26	69	0	0	0	0	98
	BA	1	1	1	92	2	1	0	98
	AQ	3	2	0	7	302	24	0	338
	OW	0	2	1	6	18	113	0	140
	MA	0	18	2	0	0	0	77	97
	Total	331	311	92	111	340	138	77	1400
	UA (%)	98	81	75	83	89	82	100	
	PA (%)	92	91	71	94	89	81	79	
	SE (%)	0.8	2.2	4.5	3.6	1.7	3.3	0.0	

UA: Users' accuracy (%); PA: Producers' accuracy (%); KC: Kappa coefficient (0.85); SE: Standard error of the mean of users' accuracy (%); OA: Overall accuracy (88%); S (OA): Standard error of the mean of overall accuracy (0.9%); AG: Agriculture; WL: Wetland; PC: Perennial crop/orchard; BA: Built-up area; OW: Open water; MA: Mangrove forest.

Table 5.3: Accuracy estimate of the RF-ELUC land cover classification using a confusion matrix.

		Land cover map							
		AG	WL	PC	BA	AQ	OW	MA	Total
Reference data	AG	341	2	2	1	6	0	0	352
	WL	2	268	0	0	4	0	3	277
	PC	3	17	78	0	0	0	0	98
	BA	5	1	4	86	1	1	0	98
	AQ	2	0	0	0	329	7	0	338
	OW	4	0	1	8	11	115	1	140
	MA	0	1	0	0	0	0	96	97
	Total	357	289	85	95	351	123	100	1400
	UA (%)	96	93	92	91	94	94	96	
	PA (%)	97	97	80	88	97	82	99	
	SE (%)	1.0	1.5	3.0	3.0	1.3	2.2	2.0	

UA: Users' accuracy (%); PA: Producers' accuracy (%); KC: Kappa coefficient (0.92); SE: Standard error of the mean of users' accuracy %); OA: Overall accuracy (94%); S (OA): Standard error of the mean of overall accuracy (0.1%); AG: Agriculture; WL: Wetland; PC: Perennial crop/orchard; BA: Built-up area; OW: Open water; MA: Mangrove forest.

5.4 Discussion

5.4.1 Moment classifications and classification probability

Moment classifications may be used for a rapid response to changes in land cover whereas the sequence of classification probability over time could be useful information for several applications. That is, high-quality images without cloud-contaminated pixels and missing data can aid to acquire highly accurate moment classifications. These results may effectively present the accelerated land cover change such as tracking fire areas and fluctuations in the water level of a reservoir. The sequence of classification probabilities supports understanding a given ELUC classification. It means the probability values are able to generate confident thresholds, whether a low or high confident one. High-confident pixels might be considered as reference data for the next moment. Low-confident pixels might be misclassified and need further handling.

5.4.2 Minimizing misclassification

The ELUC can automatically minimize misclassification due to the noise problems of low-quality images. Although the ELUC obtained correct estimates at any moment for a clear and stable area (Fig. 5.3b), it misclassified in a few moments in other areas (Fig. 5.3a). Such moment misclassification does probably not impact the final ELUC classification since the algorithm joins the major votes or probabilities throughout time.

In Fig. 5.3e, the aquaculture area has confused the ELUC; that is, it was classified as open water or agriculture according to moments. A possible explanation for this might be that agriculture, aquaculture, and open water are likely to have similar spectral reflectance and geographical characteristics. In this research region, there are several mixed areas of rice and aquatic farming [200].

5.4.3 Reduction of data preprocessing

The ELUC reduces demands for the initial processing of satellite images in some manners. Firstly, because each moment is classified independently, the pixel values of different images are not necessary to be comparable to each other in the preparation steps. Thanks to this, the algorithm is not seriously impacted by the problems of difference in haze and sun angle, which are normally calibrated by various complicated methods [164]. In addition, clouds and cloud shadows are a tough challenge for any land cover classification algorithms because non-flagged clouds may be classified as false land cover categories or false changes [180]. The ELUC may be no exception when producing classifications for individual moments. The preliminary result owing to cloud-contaminated issues is presented in Fig. 5.5. Hence, masking clouds and their shadows, a time-consuming and difficult task, is an essential preprocessing stage to exclude them, if analysts desire to use moment classification for a time-series change detection analysis. Interestingly, the final classification of the ELUC by updating new information over time is tolerant of cloud contamination. Such results are somewhat counter-intuitive as shown in Fig. 5.8.

In this research, I minimized preprocessing steps and kept the variety of classification issues to demonstrate the limitations of ELUC. The common standard of initial processing, includes georeferencing and co-registration, atmospheric correction, radiance conversion, relative correction, and topographic correction [172]. Herein, I only masked clouds and their shadows based on the availability of the Quality Assessment Band (BQA) of Landsat 8 OLI Level 1. However, in our experience, the classifications of ELUC can be much improved when obvious inaccuracy is removed. Importantly, when the sequence of moment classifications is applied for rapid change tracking, it is highly recommended to employ a full preprocessing step.

5.4.4 Adaptable and flexible applications

The ELUC can widely work together with any data and algorithms. First, the ELUC can update classified images that are produced from different approaches, various seasons, and at any spatio-temporal scale. For instance, changes in the water surface of a reservoir between a dry and rainy season may be identified by ELUC classifications with a data-rich environment. The change information might be of special interest to various scientists or practitioners in sustainable water resources management [201]. In addition, it is straightforward to build moments from different inputs such as a stack of bands or spectral indices. With the agreement of previous studies, I found

that spectral indices of the same image play a beneficial role for multiple land-cover classifications [174]. It can be seen from the data in Fig. 5.6 that the moment created from spectral indices obtained more accurately than the moment produced from the band stack for the KDE algorithm. In addition, experts can employ distinct algorithms to establish accurate moment classifications. Different algorithms are likely to contain unique characteristics, which may benefit the ELUC since it joins the major probabilities throughout time.

Alongside this diversity of classification approaches, the ELUC can operate a large combination of various data sets, both optical and radar sensors such as Landsat, ALOS/AVNIR-2, ALOS-2/PALSAR-2, SRTM, SPOT, and Sentinel. Such a great stack of images has been used successfully by the KDE algorithm [16, 17]. Multi-resource incorporation should be examined while assembly algorithms are essential in the field of land-cover change detection analysis [166, 161].

5.4.5 Potential of ELUC for large-scale use

Several attempts have been made to monitor land cover in the entire VMD, which is covered by clouds up to 93% in peak seasons [202, 203, 204]. Truong et al. and Hoang et al. have successfully quantified comprehensive land cover in the delta, using the fusion of multiple remotely sensed sources [17, 52]. However, these results contain relatively a few misclassifications among non-forest classes such as barren land and open water. This inconsistency may be due to clouds and cloud shadows [205]. Another great effort has been performed by Liu et al., which obtained an overall accuracy of 93.5% [206]. Although most of the land cover measures of this product are close to the national statistical data (the most accurate data), there are a few gaps in the area of built-up land and forest. This discrepancy might be attributed to the different definitions of land cover classification systems. Another possible explanation for this is that the product was validated using visual interpretation from Google Earth images. This process may not avoid errors, although they have carefully selected the validation data [174]. The result of this study, however, comparatively outperforms its rivals for the VMD. It may be that my approach can be tolerant of the cloud-contaminated issues, which may be the most challenging in this region. That is, the ELUC can take advantage of the informative pixel, even though the image is mostly contaminated by clouds and cloud shadows. Thin clouds or haze may be impossible to handle and thus lead to misclassifications, but the ELUC can continuously update with the new moment classification to deal with this gap. Further details about the comparison of my results with the most updated findings are presented in Text S1. It is worth noting that our approach may not outperform the state-of-the-art deep learning models, which require more training data sets. The present method is, however, outstandingly applied for the broader-scale region with simple preprocessing.

5.4.6 Limitations and future work

The ELUC seems to face a challenge in a data-poor environment when moments are used for tracking rapid changes. The results of this study show that many pixels across the study area were missing data or *'No data'* (Fig. 5.7e), causing uncertainty in moment classifications. Also, the accuracy estimation of all moments throughout time may be complex, time-consuming, and incomparable between them. Because of the data gaps at cloud-contaminated pixels, it is impossible to collect random samples of ground reference data for the accuracy assessment of the moment. I, therefore, used all usable reference data to validate moments in this research. Since the present research was designed to accurately estimate land cover categories throughout the entire study area instead of rapid change tracking, developing a full picture of accuracy assessment was not solved here. However, it is an important issue for future studies. Also, the combination of different data sources should be explored to recognize the full potential of the ELUC for accurate multi-land-cover assessment at a regional or global scale. It might also be possible to use a different theory in which the probability of a pixel is more accurately estimated for each land cover category.

5.5 Conclusions

This study set out to develop a novel approach (ELUC) to estimate large-scale land use/cover for a dense cloud-contaminated area with minimal preprocessing. Although the proposed ELUC is based on the random forest classifier, it can be operated with various algorithms to increase the performance of the model. In this work, the model leverages the information of time-series classifications to enhance the classification accuracy, say 5% improvement compared with the single-time application. In addition, this study was designed for the continuous update for more accurate land cover mapping, but the ELUC can be used to track land use/cover dynamics via moment classifications over the study periods. Meanwhile, the sequence of classification probabilities over time is expected to be useful information for several applications such as the estimate of a confident threshold for further studies. Finally, the ELUC can be tolerant of dense cloud-contaminated issues, thus significantly increasing the data density.

More importantly, this approach permits general users to simply create more consistent land use/cover products, instead of using a predefined land cover map, which may not match their specific goals. Also, this algorithm may be of interest to scientists, analysts, and others in the Earth-observing community, providing them a deeper insight into the value of a rich-data environment. With the rapid development of the current technology and remote sensing data, this study lays the groundwork for future research into Earth's surface exploration.

Chapter 6

First comprehensive quantification of annual land use/cover from 1990 to 2020 across mainland Vietnam

Abstract: Extensive studies have highlighted a need for frequently consistent land cover information for interdisciplinary studies. This Chapter proposes a comprehensive framework for the automatic production of the first Vietnam-wide annual land use/land cover (LULC) data sets (VLUCDs) from 1990 to 2020, using available remotely sensed and inventory data. Classification accuracy ranged from 85.7 $\pm$ 1.3% to 92.0 $\pm$ 1.2% with the primary dominant LULC and 77.6 $\pm$ 1.2% to 84.7 $\pm$ 1.1% with the secondary dominant LULC. This confirmed the potential of the proposed framework for systematically long-term monitoring LULC in Vietnam. Results reveal that despite slight recoveries in 2000 and 2010, the net loss of forests (19,940 km^2) mainly transformed into croplands over 30 years. Meanwhile, productive croplands were converted to urban areas, which increased approximately ten times. A threefold increase in aquaculture was a major driver of the wetland loss (1,914 km^2). The spatial–temporal changes varied, but the most dynamic regions were the western north, the southern center, and the south. These findings can provide evidence-based information on formulating and implementing coherent land management policies. The explicitly spatio-temporal VLUCDs can be benchmarks for global LULC validation and utilized for a variety of applications in the research of environmental changes toward the Sustainable Development Goals.

6.1 Introduction

Information about land use/land cover (LULC) and its dynamic changes are fundamental to a variety of studies on environmental issues [207] such as climate change [208], drought and food [209], and carbon emissions [87]. That is, frequently updated accurate LULC products provide policymakers with a profound understanding of the complex interplay between land use/land cover change (LULCC) and its risk, which

helps to inform coherent policies for the sustainable management of land resources [210, 211, 212].

The ready availability of remote sensing data and computing technologies opens a great era in cost-effective mapping LULC at a broad scale. Numerous algorithms have been developed to improve LULC classification, e.g. Spatial Temporal Adaptive Algorithm [162], Automatic Land Cover Classification Method [165], and Apply Change-vector Analysis in Posterior Probability Space [213]. Together with the development of these complex algorithms, special projects have been designed for large-scale land cover assessment. For example, at the 10-m spatial resolution, several attempts have been made to publish 13-category LULC maps of Europe [167] and global LULC maps [169] using a great set of Sentinel MSI images. For a coarser spatial resolution (30 m), there are quality multi-category LULC products, including the National Databases of the United States [170, 171, 155, 172], and the GlobeLand30 global product of 10-category LULC [173]. Nonetheless, owing to the computational restriction and the limitations of representative reference data to train and test classifiers, these products have not reflected consistently and frequently the detailed patterns and characteristics of LULC at local or national scales [52, 16]. Also, due to the predefined research periods and differences in the land cover classification systems (LCCSs), these products seldom meet the prime requirement of projects' specific objectives.

Recently, a remarkable performance in cloud computing has advanced LULC observation sciences. For example, the National Aeronautics and Space Administration Earth Exchange (NASA-NEX) and Amazon Web Service (AWS) allow analysts to access and process the NASA Earth Observation (EO) data on the cloud [214]. More importantly, Google Earth Engine (GEE) provides an outstanding cloud computing platform with open access to a variety of EO data. Thanks to the potential of big data processing of these platforms, researchers have completed extensive studies to a greater extent, for example, on urban change monitoring [215], cultivated land mapping [216], and forest disturbance detection [213]. Multi-category land cover products were also produced such as a 13-category land cover map of Southeast Asia covering 11 nations [217]. Although the overall accuracy of such products reaches up to 86%, the authors identify limitations regarding the insufficiency of high-quality reference data for time-series analyses [218]. Therefore, very few studies have been conducted for multi-temporal LULC mapping at a broad scale.

The quantity and quality of training data play an essential role in the production of LULC maps. Yet, collecting sufficient and precise training data requires considerable effort, especially at large scales and multiple periods [219]. Several attempts have been proposed methods that allow for to collection of cost-effectively high-quality training data. Bagan et al. (2019) extracted training data from a previous land cover map and utilized them for mapping new LULC products [220]. While the authors applied a bi-temporal spectral measurement to decrease the bias of extracted training data, the accuracy of these data may not be ensured due to the inherent classification errors of the previous maps [221]. To enhance the effectiveness of training data collection, Huang et al. (2020) used spectral similarity and distance indicators to detect the changed and

unchanged training sites, and thus kept the unchanged ones as migrated training data [219]. The measurement was applied for the availability of Landsat TM images. Results showed that the accuracy of the migrated training data obtained over 92.98% and the classification map which used the migrated training data had a similar overall accuracy of 71.42% to that used ground-truth data in 2010. Nevertheless, these results were validated by outdated maps, namely the ESA global CCI land cover data sets, which may contain inherent classification errors. In addition, using the Sentinel MSI images, Ghorbanian et al. (2020) employed the same approach to migrate Iran-wide training data from 2017 to 2019 [218]. The classified map that utilized the migrated training data obtained a great accuracy of 91.35%. Despite the potential of the automatic training migration method, it is still not known whether this method can be applied for multi-sensor data sources such as the different Landsat sensors or the harmonized Landsat and Sentinel images [219]. Therefore, an exploration of the potential training migration method for multi-sensor remote sensing data is integral for a time-series assessment of multi-category LULC dynamics at a large scale.

Given the ideas, this research aim is to explore the potential training migration method for multi-sensor remote sensing data and then produce the first Vietnam-wide annual land use/cover data sets (VLUCDs) from 1990 to 2020 as a case study. In Vietnam, remotely sensed data have been utilized to produce quality LULC products, but most products cover a small area of the country or a few predefined periods [15, 222, 223]. The previous inter-provincial LULC data sets were seven-category LULC maps for central and southern Vietnam in 2007 and 2017, and northern Vietnam in 2007 and 2015 [16]. More recently, Vietnam-wide maps were produced to map annual forest cover from 2015 to 2019 [52, 224]. Despite the potential of these products, due to the primary focus on forest monitoring, the classification accuracy of non-forest LULC categories may be insufficient for other applications. Meanwhile, there has been a highly dynamic LULCC which varies among different regions in Vietnam. Despite the report of continuous net forest gain by the Ministry of Agricultural and Rural Development (MARD), a systematically comprehensive review has reported forest loss in Vietnam [225, 18]. The rates and patterns of changes on the national scale may remain unknown completely. Hence, timely, accurate, and comprehensive LULC products can provide a profound understanding of LULCC patterns and processes. This information can support policymakers in forming crucial decisions on sustainable development and resource management. The maps may be benchmarks for quantifying regional and global land cover products.

The central novelty of this paper is to propose a new framework for the automatic nationwide annual LULC monitoring and provide the results of the first VLUCDs and LULCC over the recent three decades. There are major tasks: (1) Data preparation; (2) Design a proper LCCS and reference data; (2) Proposing a consistent framework for the automatic production of the VLUCDs; (3) Creating and validating the VLUCDs, and; (4) Detecting profound changes in LULC since 1990. I developed a new random-forest-based classification approach to classify the wide availability of Landsat Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+) and Operational Land Imager

(OLI), and Sentinel C-band Synthetic Aperture Radar Ground Range Detected (SAR GRD) and MultiSpectral Instrument (MSI) time-series images over the study period. The implementation is described fully in the method section.

6.2 Materials and methods

The overall method is presented in Fig. 6.1 with major steps: (1) Data preparation; (2) Defining a proper LULC classification system and reference data; (3) Proposing a consistent framework for the automatic production of Vietnam-wide annual LULC data sets from 1990 to 2020; (4) Creating and validating the VLUCDs, and; (5) Generating major change pattern and processes of LULC over the past three decades.

6.2.1 Study area

The study area is mainland Vietnam with a population of 97 million people (2018; Fig. 6.2). The country covers an area of over 300,000 km^2 including the Red River Delta, and the Mekong River Delta which is the third-largest delta in the world. The topography of Vietnam is diverse (up to 3,300 m altitude) with over 75% of the total area being hills and mountains. These areas are covered by mainly tropical rainforests. Climate is changeable but dominated by a tropical monsoon type with mean annual humidity of 84%, mean annual rainfall from 1,200 to 3,000 mm, and mean annual temperature from 21 to 27 °C [226]. The complex patterns of climate and topography create the rich biodiversity and landscape heterogeneity of Vietnam's LULC. Nonetheless, there are identifying characteristics of LULC in different regions. While the southern region is principally occupied with rice, orchards, and aquaculture lands, the northern region is primarily covered by forests and plantations, except for the Red River Delta. Dominant LULC types in the northern center are evergreen broadleaf forests and annual croplands whereas woody crops, deciduous broadleaf forests, and evergreen needle-leaf forests dominate the southern center. In this study, to reduce the complexity of the landscape information, I divided the whole country into five main regions and separately classified each region. These regions are presented in Fig. 6.2.

6.2.2 Land cover classification system

Defining a standard land cover classification system (LCCS) is a crucial step in the practical land cover assessment. It should be delineated precisely depending on the objectives of users and the availability of mapping resources. Most LULC maps employ the theory and framework of the International Geosphere-Biosphere Programme (IGBP) [54], the Land Cover Classification System (LCCS), and the Coastal Change Analysis Program (C-CAP) Land Cover Classifications [55]. Meanwhile, the most updated LCCS of previous LULC products, covering the entire Vietnam, includes 18 land cover categories [227]. However, some categories are inappropriate for Vietnam's LULC. For example, snow and ice do not exist, while one cropland category does not represent

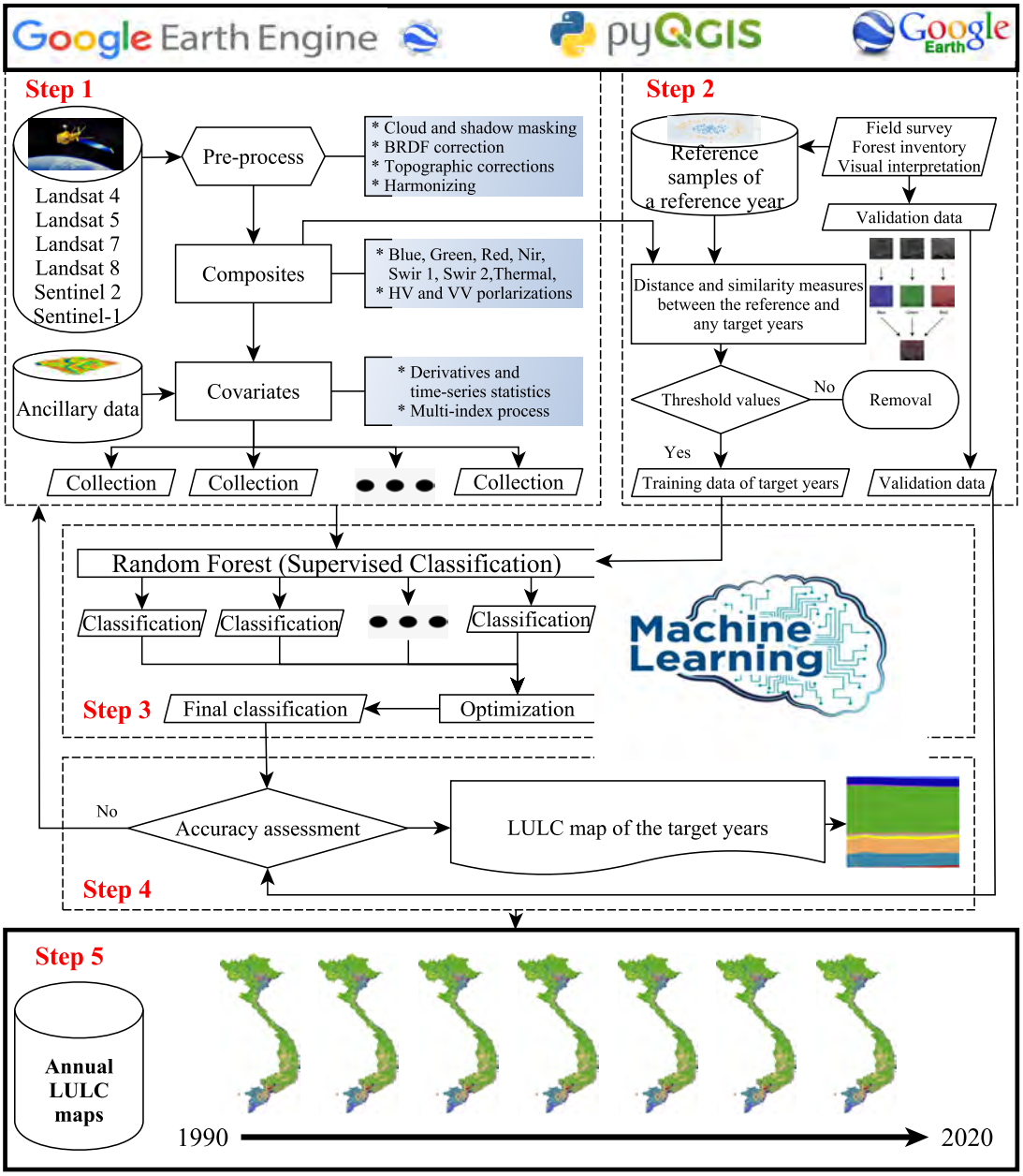


Figure 6.1: The overall workflow for automatic Vietnam-wide annual land use/cover mapping and monitoring, using Landsat TM, ETM+ and OLI, and Sentinel SAR GRD and MSI images with the random-forest-based algorithm. This figure is generated using yEd Graph Editor.

the diverse croplands in Vietnam. Although detailed classifications of high and low developed built-up areas play a fundamental role in urban planning and management for the rapid urbanization of Vietnam, they are not in the previous LULC products. In this study, therefore, a new LULC classification system or topology was developed by remaining the appropriate categories of the Food and Agriculture Organization (FAO) LCCS and adding new proper categories based on the local biophysical environment

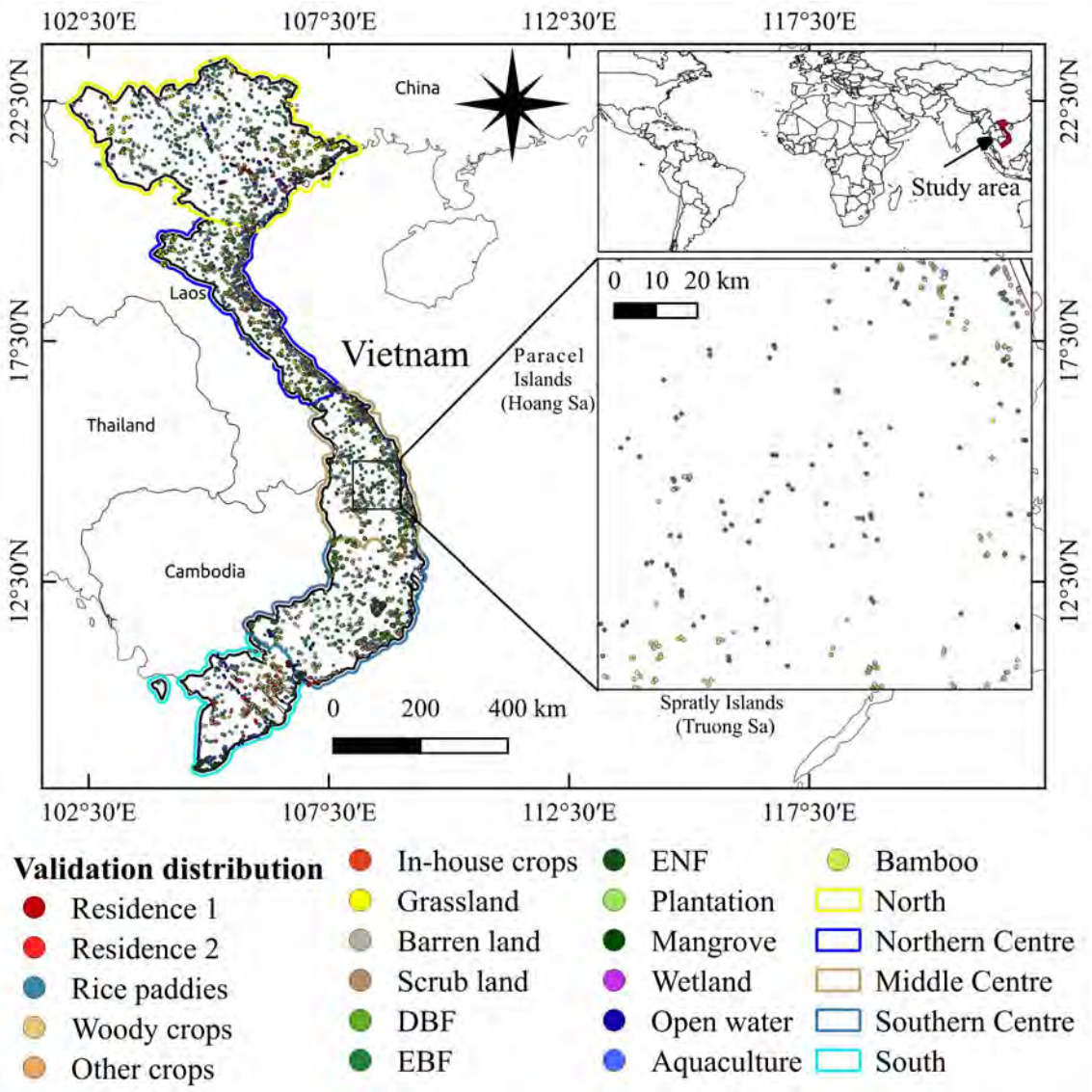


Figure 6.2: Location of mainland Vietnam in the world: major division zones (bold lines), distribution of validation data points across the country. These points are independent from the training data. This figure is generated using QGIS 3.18.0-Zurich while the country boundary is extracted from the GADM.

and end-users recommendations in Vietnam. First, I classified a 10-category system of primary dominant land use/cover (PDLC/Level-1). The PDLC was then separated into more detailed land types to generate an 18-category system of secondary dominant land use/cover (SDLC/Level-2). I found that this system is appropriate for practically mapping and applications. The categories and descriptions of the system are presented in Table A.4.

6.2.3 Remote sensing data

Multi-sensor remote sensing data were used in this study. The data were preprocessed and derived from the GEE. Te data were re-projected to Universal Transverse Mercator

(UTM) projection (Zone 47-49 N and WGS-84 datum) and then resembled into a 30-m spatial resolution using a bi-cubic interpolation method [228]. The Geospatial Data Abstraction Library (GDAL), the Geographic Resources Analysis Support System, and Python were utilized for these processing tasks. Specifically, the data included the United States Geological Survey (USGS) Landsat TM, ETM+, and OLI Surface Reluctance Tier 1 with a 30-m spatial resolution, Sentinel MSI Level-2A, and SAR GRD with a 10-m spatial resolution. The Landsat and Sentinel MSI have been atmospherically corrected while each scene of Sentinel SAR GRD was preprocessed using Sentinel-1 Toolbox for thermal noise removal, radiometric calibration, and terrain correction using the Shuttle Radar Topography Mission (SRTM) [229], and then converting to decibels. Landsat ETM+ images, after the Scan Line Corrector failure in 2003, were removed from this study since the failure may result in inconsistently time-series comparison. Over 99% of the data sets from the GEE archive are reported to have high geometric accuracy with the error being less than half a pixel [219]. Otherwise, the images were eliminated from the image collection to reduce the obvious bias of further analysis.

For reliable and consistent time-series analysis, further processing is essential. For the optical data, to reduce illumination impacts from elevation, aspect and slope, the topographic correction was performed using the Modified Sun-Canopy-Sensor Topographic Correction algorithm [230]. While the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) [231] was applied to perform atmospheric correction for Landsat TM and ETM+, the Land Surface Reflectance Code (LaSRC) [232] was adopted for Landsat OLI. All Landsat images were masked and removed clouds, cloud shadows and saturation pixels utilizing the Function of Mask (CFMASK) [233]. Sen2Cor was adopted to correct atmospheric issues and mask clouds for Sentinel MSI [234]. Finally, because of the different solar and view angles of Landsat OLI and Sentinel MSI, normalizing the bidirectional reflectance distribution function (BRDF) was applied for the data. Although numerous approaches have developed for BRDF correction, the recent technique generated by Roy et al. [235] is frequently utilized due to its reliability and effective implementation [236]. This method, therefore, was employed for the BRDF correction of all selected optical images in this study. For the Sentinel SAR GRD data, a further process was speckle filtering. The filtering was done using Lee filter, which is superior due to its capacity of maintaining point targets, edge, linear spaces and texture information [237].

To increase the availability of cloud-free composite data, the harmonization of different Landsat satellite sensor images, and the Landsat OLI with Sentinel MSI into a congruent time series was desirable for a cloudy region such as Vietnam. The harmonization allows to accurately compare across all years and to measure the spectral similarity and spectral distance between different years. The measurement of spectral similarity and spectral distance was applied for an automatic training migration model, which was described in the following sections. A linear transformation with band-respective coefficients was applied for the harmonization of Landsat TM and ETM+ spectral feature to OLI spectral feature [238]. In the meantime, the harmonized

Landsat OLI and Sentinel MSI images were processed by employing a method developed by Claverie et al. [239]. The band-respective coefficients with slope and intercept image constants are presented in Table 6.1. After that, I generated composites of seven bands including blue, green, red, nir, swir 1, swir 2, and thermal bands for two seasons, the dry season from April to September and the wet season from October to March of the following year. These composites were adopted to measure a variety of covariates, which are represented in the following paragraphs. In addition, seasonal composites of VV and VH polarization in ascending and descending orbits of Sentinel SAR GRD were handled in this research.

6.2.4 Satellite-based covariate calculation

In this section, a series of covariates were calculated from the band composites. For the optical data, I calculated the medoid [62] and the standard deviation for the six bands. Following the successful application of numerous features extracted from original satellite image bands, this study also added the medoid of the 20th and 80th percentile [227] of the six bands into the seasonal composites to detect the seasonal changes in the biophysical environment. The ratios between spectral bands were calculated; they are blue/green, red/blue, red/green, red/nir, and nir/(red*swir 1). Besides, a great number of spectral indices were also measured from Landsat TM, ETM+, OLI, and Sentinel MSI images (Eq.(6.1) - Eq.(6.17)). harmonization of these satellite data was created (Table 6.1). In addition, I calculated the seasonal mean of VH, VV, and the normalized difference between VH and VV polarization from the Sentinel SAR GRD images. Finally, I generated seasonal composite collections of covariates.

Table 6.1: Band-respective coefficients are defined with slope and intercept image constants and used for the harmonized Landsat OLI and Sentinel MSI images.

Respective bands		Blue	Green	Red	NIR	SWIR1	SWIR2
Landsat TM,	Intercept	0.0003	0.0088	0.0061	0.0412	0.0254	0.0172
ETM+ & OLI	Slope	0.8474	0.8483	0.9047	0.8462	0.8937	0.9071
Landsat OLI &	Intercept	-0.0107	0.0026	-0.0015	0.0033	0.0065	0.0046
Sentinel MSI	Slope	1.0946	1.0043	1.0524	0.8954	1.0049	1.0002

Atmospherically Resistant Vegetation Index (ARVI) [240].

$$ARVI = \frac{B_{\text{nir}} - 2B_{\text{red}} + B_{\text{blue}}}{B_{\text{nir}} + 2B_{\text{red}} + B_{\text{blue}}} \tag{6.1}$$

Difference Vegetation Index (DVI) [241].

$$DVI = B_{\text{nir}} - B_{\text{red}} \tag{6.2}$$

Enhanced Built-Up and Bareness Index (EBBI) [117].

$$EBBI = \frac{B_{\text{swir}} - B_{\text{nir}}}{10\sqrt{B_{\text{swir}} - B_{\text{nir}}}} \tag{6.3}$$

Enhanced Vegetation Index (EVI) [242].

$$EVI = 2.5\frac{B_{\text{nir}} - B_{\text{red}}}{B_{\text{nir}} + 6B_{\text{red}} - 7.5B_{\text{blue}} + 1} \tag{6.4}$$

Green Chlorophyll Index (GCI) [243].

$$GCI = \frac{B_{\text{nir}}}{B_{\text{green}}} - 1 \tag{6.5}$$

Mangrove Vegetation Index (MVI) [244].

$$MVI = \frac{B_{\text{nir}} - B_{\text{green}}}{B_{\text{swir}} - B_{\text{green}}} \tag{6.6}$$

Normalized Burn Ratio (NBR) [245].

$$NBR = \frac{B_{\text{nir}} - B_{\text{swir}}}{B_{\text{nir}} + B_{\text{swir}}} \tag{6.7}$$

Normalized Different Bareness Index (NDBaI) [120].

$$NDBaI = \frac{B_{\text{swir}} - B_{\text{tir}}}{B_{\text{swir}} + B_{\text{tir}}} \tag{6.8}$$

Normalized Difference Built-Up Index (NDBI) [118].

$$NDBI = \frac{B_{\text{swir}} - B_{\text{nir}}}{B_{\text{swir}} + B_{\text{nir}}} \tag{6.9}$$

Normalised Difference Pond Index (NDPI) [246].

$$NDPI = \frac{B_{\text{green}} - B_{\text{swir}}}{B_{\text{green}} + B_{\text{swir}}} \tag{6.10}$$

Normalized Difference Turbidity Index (NDTI) [246].

$$NDTI = \frac{B_{\text{green}} - B_{\text{swir}}}{B_{\text{green}} + B_{\text{swir}}} \tag{6.11}$$

Normalized Difference Vegetation Index (NDVI) [121].

$$NDVI = \frac{B_{\text{nir}} - B_{\text{red}}}{B_{\text{nir}} + B_{\text{red}}} \tag{6.12}$$

Normalized Difference Water Index (NDWI) [187].

$$NDWI = \frac{B_{\text{green}} - B_{\text{nir}}}{B_{\text{green}} + B_{\text{nir}}} \tag{6.13}$$

Soil Adjusted Vegetation Index (SAVI) [123].

$$SAVI = 1.5\frac{B_{\text{nir}} - B_{\text{red}}}{B_{\text{nir}} + B_{\text{red}} + 0.5} \tag{6.14}$$

Structure Insensitive Pigment Index (SIPI) [247].

$$SIPI = \frac{B_{\text{nir}} - B_{\text{blue}}}{B_{\text{nir}} - B_{\text{red}}} \tag{6.15}$$

Urban Index (UI) [119].

$$UI = \frac{B_{\text{swir}} - B_{\text{nir}}}{B_{\text{swir}} + B_{\text{nir}}} \tag{6.16}$$

Water Ratio Index (WRI) [248].

$$WRI = \frac{B_{\text{green}} - B_{\text{red}} + B_{\text{nir}}}{B_{\text{blue}}} \tag{6.17}$$

6.2.5 Ancillary data sets

Extensive research has shown that ancillary information can improve the accurate performance of LULC classifications [16, 249, 250]. In this study, I first added terrain indices including slope, aspect, and elevation. These indices were computed from ALOS Global Digital Surface Model or "ALOS World 3D-30 m (AW2D30)" [68]. Also, distance to rivers, coastlines, transport systems and buildings, and soil types were included in the covariate collections. The buildings and transport systems were generated from the OpenStreetMap, while soil types and river networks were extracted from the OpenDevelopmentMekong.

6.2.6 Reference data

Reference data of 18 LULC categories (see Appendix A.4 for further details) was created from field surveys, provincial LULC statistics, and visual interpretations. I conducted nationwide comprehensive surveys in 2015, 2016, 2018, 2019, and 2020 to collect 3,078; 2,659; 10,550; 41,986; and 32,853 reference samples, respectively. Along with these ground-observed data, previous outdated LULC maps [52, 16, 17], provincial LULC statistics and high-resolution satellite images available on Google Earth were also considered. Herein, I generated approximately 9,360 polygons of single homogeneous LULC types (Fig. 11) throughout the country for each year from 2015 to 2020. From these polygons, I extracted up to 120,000 reference pixels (points) for each of the years. Meanwhile, due to the non-availability of ground-truth data, reference data from 1990 to 2014 were collected using provincial LULC statistics, the natural-color images of Landsat TM, ETM+ and OLI, Sentinel MSI, and high-resolution satellite images available on Google Earth. For each year, I randomly extracted 1,050 points per LULC category for validating the classification models and the others were used for training the classification model. The reference data for the year 2020 were utilized not only for creating and validating the LULC map of the year 2020 but also for implementing the automatic training migration model from this reference year to any target years. The migration model is described in more detail in the following section.

6.2.7 Automatic training migration model

Training data is tremendously essential in mapping LULC; however, collecting sufficiently accurate training samples is challenging, especially for large-scale areas, long-term history analyses, and data-scarce environments such as Vietnam [14]. If training data are not collected consistently, it can result in misclassification or low accuracies [251]. Thus, it is paramount to propose a practical approach for training data collections. In this study, I utilized an automatic model to migrate from the reference data of a reference year to target years. The method had three essential steps. First, I created a set of training data from a reference year (2020). Then, for each pixel, I computed its surface reflectance values from its corresponding Landsat TM, ETM+, and OLI images of the reference year and target years. The surface reflectance values of six bands (Table 6.1) were utilized for the measurement of Euclidean distance (ED) [252] and spectral angle distance (SAD) [253]. Finally, with the ED (Eq. 18) and SAD (Eq. 19), I distinguished changed pixels and unchanged pixels by running a trial and error model to determine thresholds. Although the thresholds can be estimated by analysts, the experimented thresholds of ED and SAD in this study were 0.05 and 0.95, respectively. These thresholds were successfully applied to migrate the training data of the year 2020 to the target years. The unchanged pixels were preserved and utilized as training data for the target years.

$$ED = \sqrt{\sum_{i=1}^{N} (X_i - Y_i)^2} \tag{6.18}$$

$$SAD = arccos\left(\frac{\sum_{i=1}^{N} X_i Y_i}{\sqrt{\sum_{i=1}^{N} (X_i)^2 \sum_{i=1}^{N} (Y_i)^2}}\right) \tag{6.19}$$

where X is spectral signature vector of an image pixel in the reference year; Y is spectral signature vector of an image pixel in the target year; N is the number of image bands ($N = 6$).

6.2.8 Machine learning modelling

After completing the data preprocessing task, I generated covariate collections, including (1) optical-image-based covariates in dry seasons and (2) in wet seasons, and (3) SAR-based covariates in dry seasons and (4) in wet seasons. The ancillary information was also added to these covariate collections. It is worth noting that some of these covariates or features may not significantly contribute to the enhancement of classification performance while overabundant features can affect the performance speed or run out of the computing capacity of the classification model. Hence, random forest algorithm [192] was employed to estimate important features. I removed some less important features and kept essential features which were represented in Table A.5.

For classification, I applied a random forest algorithm for several reasons. First, it has previously been observed that the random forest algorithm can handle principal drawbacks that a single-tree-based method may face such as an over-fitting and non-optimal solution [254]. Also, the random forest shows the out-performance of its rivals such as fuzzy adaptive resonance theory-supervised predictive mapping (Fuzzy ARTMAP), support vector machine (SVM), artificial neural network (ANN), Mahalanobis distance (MD), and spectral angle mapper (SAM) [22].

A new random-forest-based approach was developed in this study. Unlike the common use of single-time classification, for each pixel, I independently estimated prior probability values belonging to each of the specified land covers for each of the covariate collections. These prior probability values were then joined to create a set of posterior probability values. The largest value of the posterior probabilities corresponding to a specific land cover was utilized to label the predicted land cover. However, the predicted prior probability of a pixel, for example, $p(C_k)$, might reach almost zero or zero because of *'No data'* of that pixel at that covariate collection. If this occurs, the posterior probability of that pixel will be nearly zero or zero. That is, although the prior probability of most other collections equals 100% voting for a specified land cover, the probability product of this pixel might be almost zero, causing misclassification. Hence, the prior probability of a pixel corresponding to a particular land cover must not be extremely tiny. To this end, Eq. 6.20 was developed to adjust the prior probability values while posterior probability values were calculated as Eq. 6.21.

$$p'(C_k) = c * p(C_k) + \frac{1-c}{N} \tag{6.20}$$

$$p_c(C_k) = \prod_{i=1}^{I} p'_i(C_k) \tag{6.21}$$

where $p^{'}(C_k)$ is the adjustment of prior probability value of a land cover C_k; c is a constant value (c = 0.7) which was evaluated by trial and error experiments in this study; N is the number of land cover categories (N = 18); $p_c(C_k)$ is the posterior probability value of category C_k; and I is the number of covariate collections (I = 4).

The random-forest-based model was performed using Scikit-Learn 0.22 and Python 3.8.5. Since the input data of multi-sensor image bands and covariates dramatically varied over the 30 years, it could not be optimized all the parameters of the random forest algorithm. The number of trees ($n_{estimators}$) in the forest and the size of the random subsets of features ($max_{features}$), however, are highly recommended to be adjusted [192]. Using RandomizedSearchCV in the Scikit-Learn, I found that $n_{estimators}$ = 200 (trees) and $max_{features}$ = 8 were optimal in this work. The other parameters were set as the default values.

6.2.9 Accuracy assessment

Following the wide-ranging recommendations of instruction manuals [63, 255] a statistic-based testing data set (Section 6.2.6) was independently generated to estimate the accuracy of final LULC products. I utilized a stratified sampling (1,050 points/LULC category) method and a confusion matrix to assess Vietnam-wide annual LULC products from 1990 to 2020. The matrix produced profound accuracy metrics, namely overall accuracy (OA), user accuracy (UA), standard error (SE), and kappa coefficient (KC). The uncertainty of accuracy was measured with a 95% confidence interval. These metrics are fully described in Tables 6.2 and 6.3.

6.2.10 Change analysis

The analysis of changes in LULC is to measure the differences including spatio-temporal dynamic patterns, the magnitude, and rate of variations observed over the study period. First, I estimated the diversity of LULC within each 30-m pixel width by counting the number of times that LULC changes over 30 years (Fig. 6.3c). I then estimated the area of each LULC within a five-year interval from 1990 to 2020 to observe the trend of LULC change (Fig. 6.7). I also computed the percentage of net change (Eq. 6.22) and then rescaled the percentage to a rank between 0 and 100% to monitor the most dynamic LULC (Fig. 6.8). Finally, I employed a Sankey diagram to emphasize the major transfers of LULC [256].

$$p = \left(\frac{A_{t2} - A_{t1}}{A_{t1} * (t_2 - t_1)} \right) * 100 \tag{6.22}$$

where p (year^{-1}) is the percentage of net change; and A_{t1} and A_{t2} (km^2) are the area of the LULC type in the observation years t_1 and t_2, respectively ($t_1 < t_2$).

6.3 Results

6.3.1 The accuracy of the first VLUCDs

Utilizing ground-based data and all the freely available remotely sensed images, I have provided a coherent method and the results of the first VLUCDs. The proposed method generated consistently spatio-temporal LULC maps, using a definitive LCCS designed regarding end users' recommendations and a standard LCCS. For a visual presentation, level-1 VLUCDs of the year 1990 and 2020 are presented in Fig. 6.3. The 5-year-interval maps (1990, 1995, 2000, 2005, 2010, 2015, and 2020) of level-2 VLUCDs are presented in Fig. 6.4. The level-1 and level-2 VLUCDs included ten categories of primary dominant land use/cover (PDLC) and eighteen categories of secondary dominant land use/cover (SDLC), respectively (see detail in Appendix A.4).

The reliability of the VLUCDs was evaluated by using both visual interpretations and statistical approaches. Based on high-resolution satellite images in Google Earth,

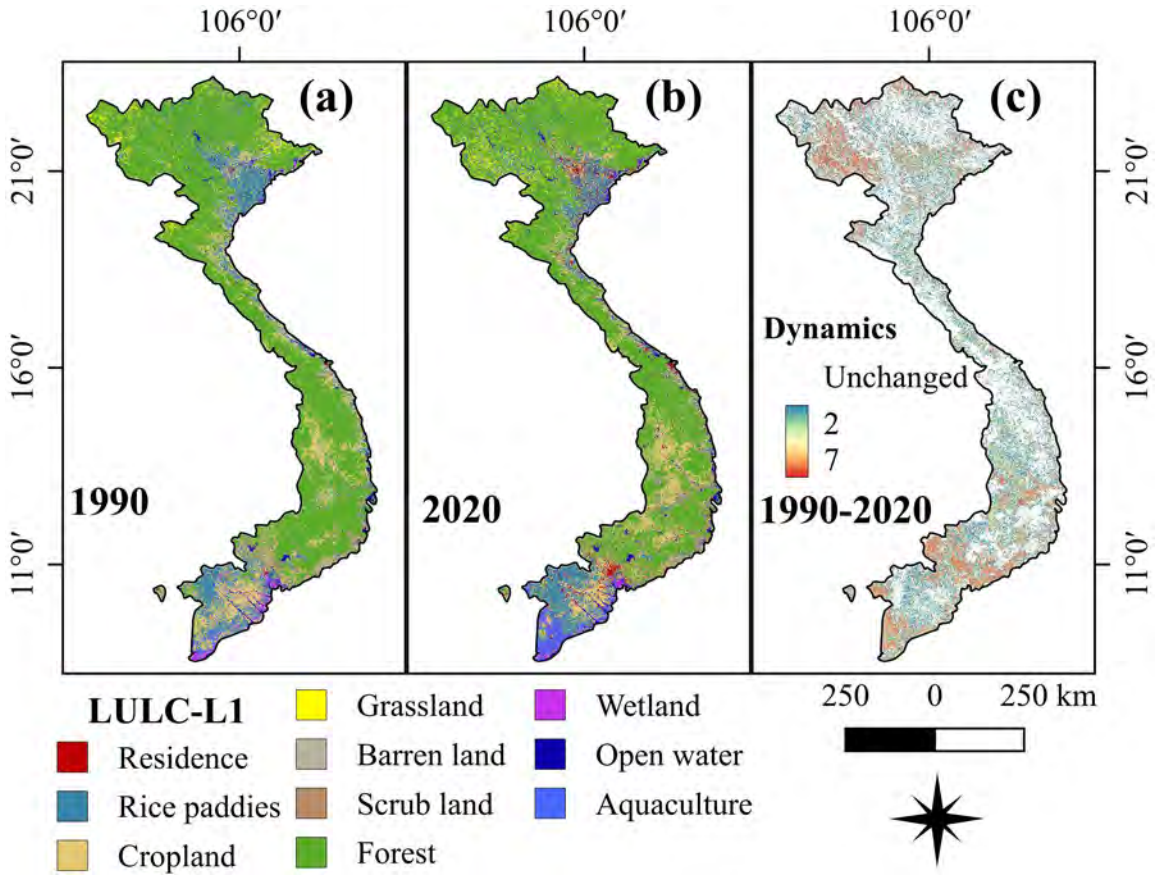


Figure 6.3: (**a** and **b**) show the level-1 Vietnam-wide LULC maps in 1990 and 2020 produced from a fusion of Landsat TM, ETM+ and OLI, and Sentinel SAR GRD and MSI images with the random-forest-based algorithm. (**c**) presents a spatial–temporal dynamic change in LULC from 1990 to 2020 in Vietnam.

I found that the VLUCDs were clear and noise-free. A confusion matrix method with stratified random sampling (1,050 points/LULC category) was utilized to independently validate classification accuracy. Statistical metrics were measured, namely producer accuracy (PA), user accuracy (UA), F1 score, overall accuracy (OA), standard error (SE), and kappa coefficient (KC). These metrics of the level-1 and level-2 LULC maps of the year 2020 are fully described in Tables 6.2 to 6.4. Meanwhile, overall obtained OA, KC, and uncertainty of the level-1 and level-2 VLUCDs are presented briefly in Figs. 6.5 and 6.6. Uncertainty of measurement was estimated with a 95% confidence interval. Specifically, the OA of the level-1 and level-2 LULC maps ranged from 85.7$\pm$1.3% to 92.0$\pm$1.2% and 77.6$\pm$1.2% to 84.7$\pm$1.1% over the study period, respectively.

With the numerous LULC types and long-term observation, these results constituted an outstanding achievement [23]. For the detailed LULC products (Level 2), open water and mangrove had the highest levels of accuracy, accounting for over 96% in both the PA and UA. This successful classification may be explained by the benefits of using multiple spectral indices (Section 6.2.4) such as the NDWI, WRI, and NDPI, which could distinguish open water from land, aquaculture ponds, and others, whereas mangrove can be accurately identified with the MVI [22]. This was followed by rice paddies,

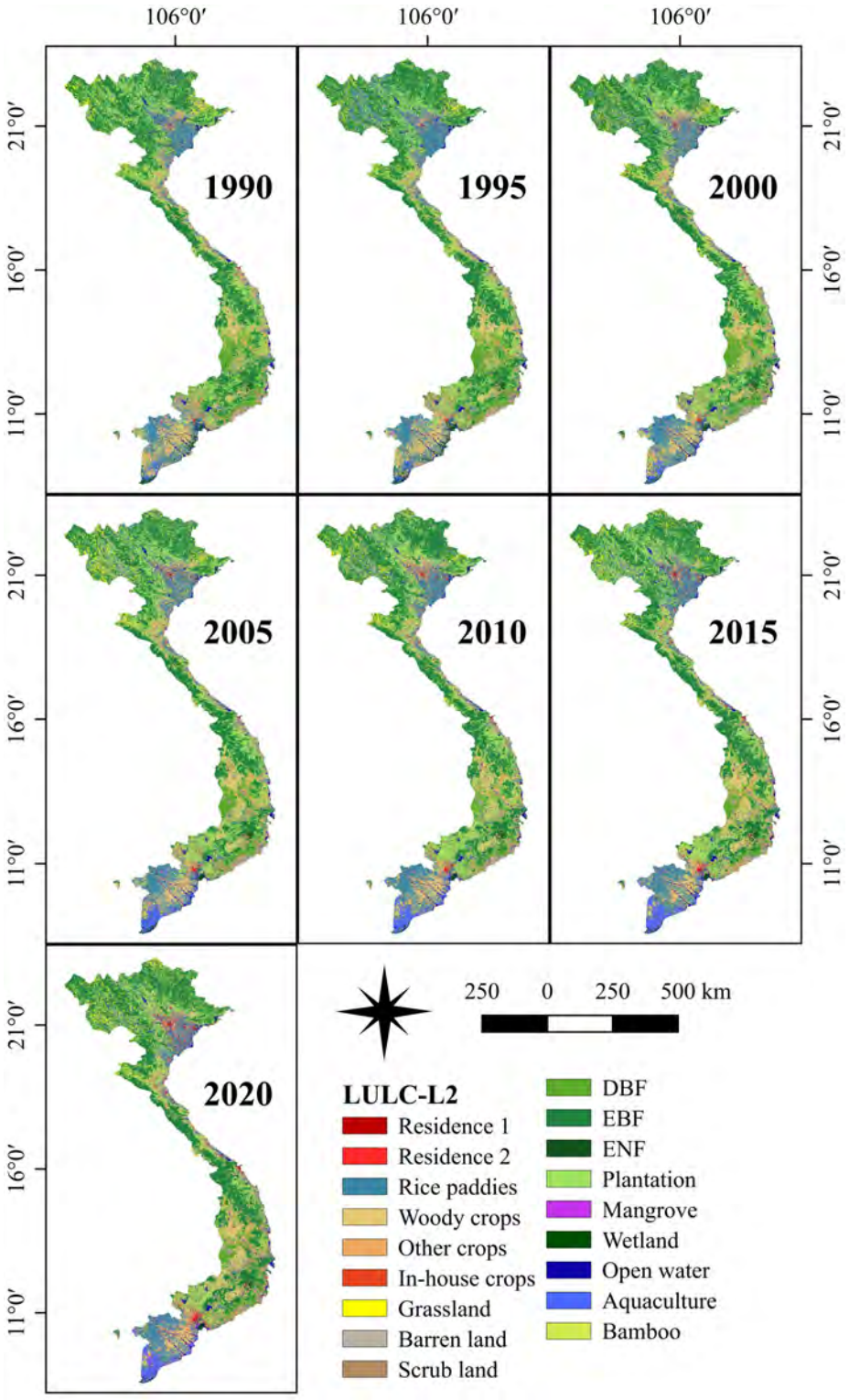


Figure 6.4: The level-2 Vietnam-wide LULC maps in 1990, 1995, 2000, 2005, 2010, 2015 and 2020 produced from a fusion of Landsat TM, ETM+ and OLI, and Sentinel SAR GRD and MSI images with the random-forest-based algorithm.

which had an accuracy of above 90%. It seems possible that rice is frequently cultivated in flat terrain, where is not be affected by topographic problems such as the shadows of

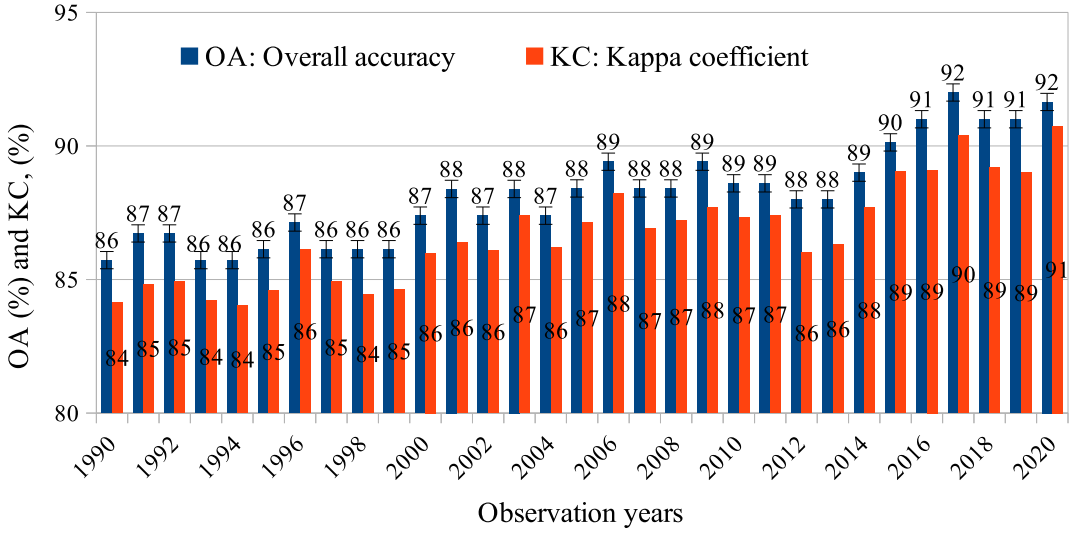


Figure 6.5: The overall accuracy (OA) and kappa coefficient (KC) of the level-1 Vietnam-wide annual LULC maps produced from the all freely available Landsat TM, ETM+ and OLI, and Sentinel SAR GRD and MSI images with the random-forest-based algorithm. The OA and KC are obtained by using a confusion matrix and a stratified validation method with independent samples (1,050 points/LULC category). The bars indicate uncertainties of OA measured with a 95% confidence interval.

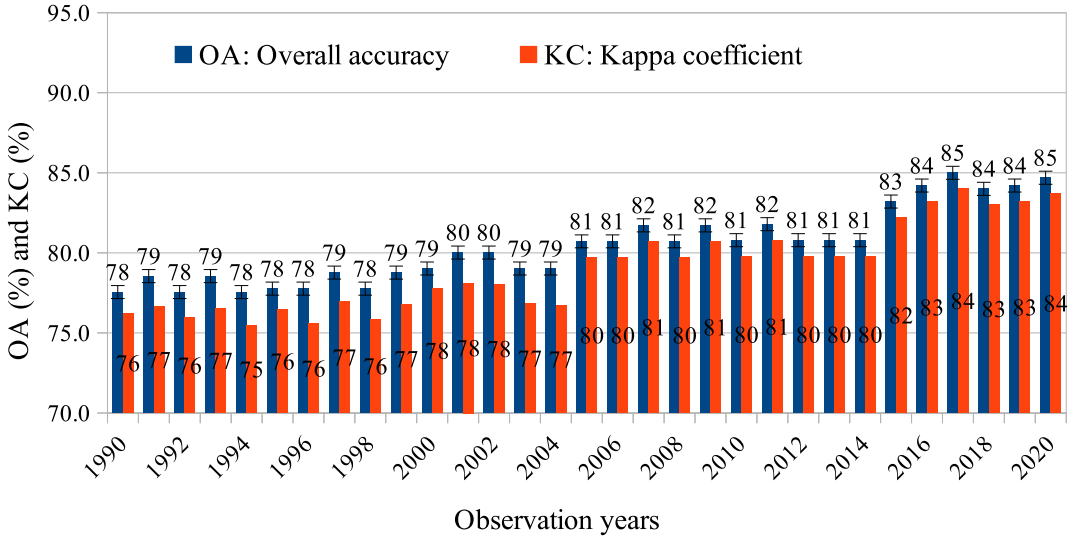


Figure 6.6: The overall accuracy (OA) and kappa coefficient (KC) of the level-2 Vietnam-wide annual LULC maps produced from the all freely available Landsat TM, ETM+ and OLI, and Sentinel SAR GRD and MSI images with the random-forest-based algorithm. The OA and KC are obtained by using a confusion matrix and a stratified validation method with independent samples (1,050 points/LULC category). The bars indicate uncertainties of OA measured with a 95% confidence interval.

mountainsides. The spectral reflectance of rice is also stable [257]. Although the model could separate forests from others, it tended to misclassify different forests. Another

Table 6.2: Confusion matrix of the 2020 Vietnam-wide land use/cover map (Level 1) created from the integration of Landsat OLI, Sentinel SAR GRD and MSI satellite images with the random-forest-based algorithm. PA: Producer accuracy (%); PA: Producer accuracy (%); UA: User accuracy (%); SEM: Standard error of the mean for UA; F1: F1 score; Overall accuracy: 91.6%, and Kappa coefficient: 90.7%. RL: Residence; RP: Rice paddies; CL: Cropland; GL: Grassland; BL: Barren land; SL: Scrubland; FL: Forest land; WL: Wetland; OW: Open water; AC: Aquaculture.

		Land cover map									
		RL	RP	CL	GL	BL	SL	FL	WL	OW	AC
Reference data	RL	988	3	21	1	6	4	0	2	0	2
	RP	1	988	26	1	4	1	8	0	0	1
	CL	27	18	772	1	28	38	17	1	0	0
	GL	1	15	57	990	15	8	64	0	0	0
	BL	25	8	14	44	980	11	11	0	0	1
	SL	0	4	23	13	17	986	44	1	0	3
	FL	1	2	14	0	0	2	853	0	0	0
	WL	3	2	119	0	0	0	53	1040	0	44
	OW	0	3	0	0	0	0	0	1	1031	4
	AC	4	7	4	0	0	0	0	5	19	995
	PA	96.2	95.9	85.6	86.1	89.6	90.4	97.8	82.5	99.2	96.2
	UA	94.1	94.1	73.5	94.3	93.3	93.9	81.2	99.0	98.2	94.8
	SEM	0.7	0.7	1.4	0.7	0.8	0.7	1.2	0.3	0.4	0.7
	F1	0.95	0.95	0.79	0.90	0.91	0.92	0.89	0.90	0.99	0.96

limitation is to classify plantation forests from woody crops, which is also found by numerous studies [258, 259]. Likewise, the model could not entirely divide the different types of residential areas, but it showed a clear separation of the residential areas from others. To increase the accuracy of the maps for further analyses, I combined these mixed categories. This combination obtained an increase in accuracy of approximately 6% with a few losses of detail in LULC types.

6.3.2 Distribution and trend of land use/land cover changes

Change detection was conducted to comprehend LULCC patterns and processes. To this end, the level-1 (PDLC) Vietnam-wide annual LULC data sets (L1-VLUCDs) were utilized for further analysis in this study. Although the annual maps are integral to obtaining the process of LULC dynamic changes in Vietnam, the five-year-interval land cover products in 1990, 1995, 2000, 2005, 2010, 2015, and 2020 were utilized to acquire a more profound change visualization. A post-classification analysis method was employed to measure the spatio-temporal LULCC and the percentage of changes.

The spatial distributions and the patterns of Vietnam LULC are shown in Fig. 6.3. The temporal distribution of the net changes in LULC from 1990 to 2020 is presented

Table 6.3: Confusion matrix of the 2020 Vietnam-wide LULC map (Level 2) created from the integration of Landsat OLI, and Sentinel SAR GRD and MSI satellite images with the random-forest-based algorithm. PA: Producer accuracy (%); UA: User accuracy (%); SEM: Standard error of the mean for UA; F1: F1 score; Overall accuracy: 84.7%, and Kappa coefficient: 83.8%. R1: Residence 1; R2: Residence 2; RP: Rice paddies; WC: Woody crops; OC: Other crops; IC: In-house crops; GL: Grassland; BL: Barren land; SL: Scrubland; DBF: Deciduous broadleaf forest; EBR: Evergreen broadleaf forest; ENF: Evergreen needleleaf forest; PL: Plantation land; MF: Mangrove forest; IW: Inland wetland; OW: Open water; AC: Aquaculture; BA: Bamboo areas (**to be continued on Table 6.4**).

		Land cover map								
		R1	R2	RP	WC	OC	IC	GL	BL	SL
Reference data	R1	804	137	3	0	10	0	2	6	1
	R2	212	848	4	17	12	3	1	7	8
	RP	0	1	932	5	27	1	3	5	0
	WC	0	2	12	674	45	0	6	3	6
	OC	0	2	27	26	801	2	1	5	11
	IC	15	34	9	15	26	1044	0	26	84
	GL	0	0	14	39	45	0	975	17	9
	BL	16	14	11	8	12	0	40	960	10
	SL	0	1	3	13	8	0	17	16	898
	DBF	1	0	1	21	5	0	0	2	17
	EBF	0	0	1	8	1	0	2	1	1
	ENF	0	0	0	37	0	0	0	0	5
	PL	0	0	12	12	16	0	1	1	0
	MF	0	6	2	7	2	0	0	1	0
	IW	0	3	7	160	35	0	0	0	0
	OW	1	0	2	0	0	0	0	0	0
	AC	1	2	10	0	3	0	0	0	0
	BA	0	0	0	8	2	0	2	0	0
	PA	83.2	76.1	95.0	84.6	89.4	82.9	83.4	88.2	88.4
	UA	76.6	80.8	88.8	64.2	76.3	99.4	92.9	91.4	85.5
	SEM	1.3	1.2	1.0	1.5	1.3	0.2	0.8	0.9	1.1
	F1	0.80	0.78	0.92	0.73	0.82	0.90	0.88	0.90	0.87

in Fig. 6.7. The most dominant LULC was forest, accounting for an approximately half area of the entire country. This was followed by croplands (16.3%), rice fields (14.2%), and open water (including parts of saltwater, 8.1%). Grassland and scrubland occupied a relatively similar proportion (2.8%) while the smallest LULC was residential areas (1.3%).

Table 6.4: (**– continued from Table 6.3**). Confusion matrix of the 2020 Vietnam-wide LULC map (Level 2) created from the integration of Landsat OLI, and Sentinel SAR GRD and MSI satellite images with the random-forest-based algorithm. PA: Producer accuracy (%); UA: User accuracy (%); SEM: Standard error of the mean for UA; F1: F1 score; Overall accuracy: 84.7%, and Kappa coefficient: 83.8%. R1: Residence 1; R2: Residence 2; RP: Rice paddies; WC: Woody crops; OC: Other crops; IC: In-house crops; GL: Grassland; BL: Barren land; SL: Scrubland; DBF: Deciduous broadleaf forest; EBR: Evergreen broadleaf forest; ENF: Evergreen needleleaf forest; PL: Plantation land; MF: Mangrove forest; IW: Inland wetland; OW: Open water; AC: Aquaculture; BA: Bamboo areas.

		Land cover map								
		DBF	EBF	ENF	PL	MF	IW	OW	AC	BA
Reference data	R1	0	0	0	0	0	0	0	3	0
	R2	0	0	0	0	0	2	0	1	0
	RP	2	0	0	5	0	0	0	0	0
	WC	5	7	0	35	2	0	0	0	0
	OC	3	2	0	15	0	1	0	0	0
	IC	6	0	0	0	0	0	0	0	0
	GL	2	24	0	43	0	0	0	0	1
	BL	6	1	0	5	0	0	1	4	0
	SL	33	12	0	13	0	0	0	2	0
	DBF	980	30	4	35	0	0	0	0	0
	EBF	2	355	1	41	0	0	0	0	6
	ENF	9	202	1045	19	0	0	0	0	0
	PL	2	39	0	603	2	3	0	0	0
	MF	0	2	0	17	1036	1	1	3	0
	IW	0	0	0	100	2	1042	0	95	0
	OW	0	0	0	0	1	0	1028	3	0
	AC	0	0	0	0	6	1	20	939	0
	BA	0	376	0	119	1	0	0	0	1043
	PA	89.4	84.7	79.3	87.3	96.1	72.2	99.3	95.6	67.2
	UA	93.3	33.8	99.5	57.4	98.7	99.2	97.9	89.4	99.3
	SEM	0.8	1.5	0.2	1.5	0.4	0.3	0.4	0.9	0.3
	F1	0.91	0.48	0.88	0.69	0.97	0.84	0.99	0.92	0.80

What can be clearly seen in Fig. 6.7 is the steady decline of forest area from 170,458 km^2 in 1990 to 150,517 km^2 in 2020. The area of wetlands experienced a slight increase from 4,404 km^2 in 1900 to 5,138 km^2 in 1995, followed by a continual decrease to 2,490 km^2 in 2020. In contrast, there was a sharp increase in the area of aquaculture and residential land, accounting for approximately three and ten times over the three decades. The area of open water showed a slight rise while there was a small fluctuation

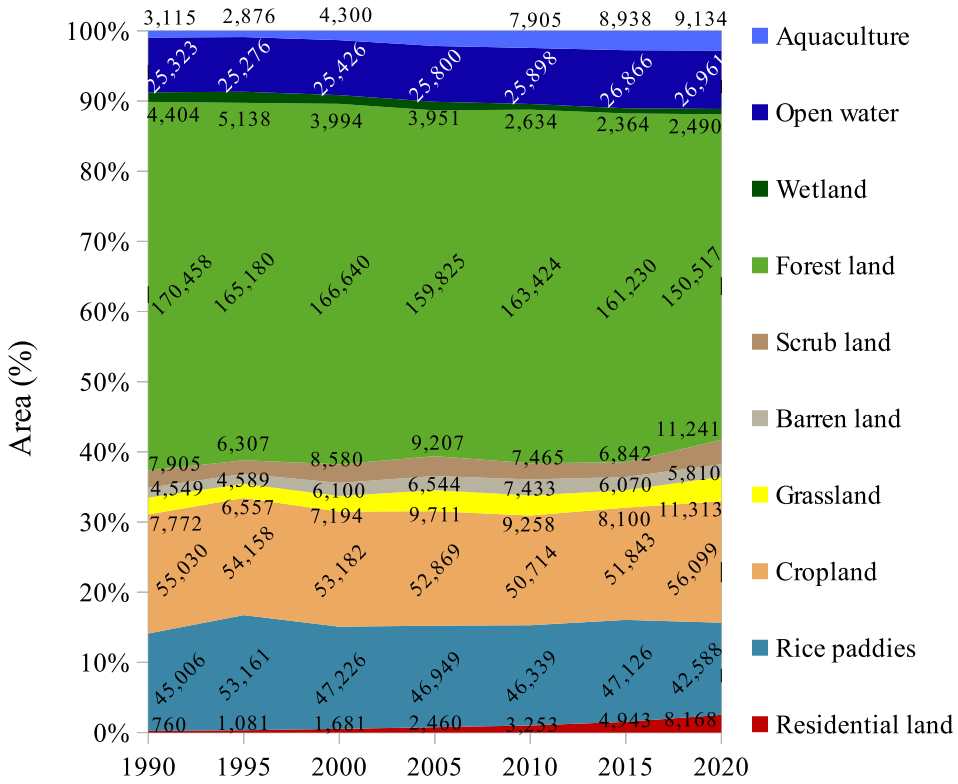


Figure 6.7: Temporal distribution of LULC across Vietnam extracted from the level-1 Vietnam-wide annual LULC data sets. The data labels represent the area of each LULC category (km^2) in the year 1990, 1995, 2000, 2010, 2015, and 2020.

in the area of the other land types.

Meanwhile, Fig. 6.8 reveals the highly dynamic change in LULC in Vietnam. The most considerable change was the area of urban land, which increased by about 50% over each five-year interval. Interestingly, the graph shows substantially opposite trends between aquaculture and wetlands. The area of aquaculture decreased 8% whereas the wetlands expanded 17% from 1990 to 1995, followed by a 50% increase in aquaculture but a 22% decrease in wetlands by 2000. Forest cover had undergone an up-and-down variation by 2015, but it has presented a remarkable drop until now. It is noticed that the percentage of change in the forest cover was insignificant, but its dynamic areas were remarkable (Fig. 6.9).

6.3.3 Major spatio-temporal land use/land cover dynamics

Vietnam's LULC has experienced a considerable change over the past 30 years. Fig. 6.3c shows the spatio-temporal dynamic changes. The north and south were the most dynamic areas, especially the western north and the south. The dynamic conversions among different LULC can be seen in Fig. 6.9. There was a fundamentally dynamic conversion between forests and croplands. Forest areas remarkably reduced while residential and aquaculture land significantly increased. To easily visualize the LULC

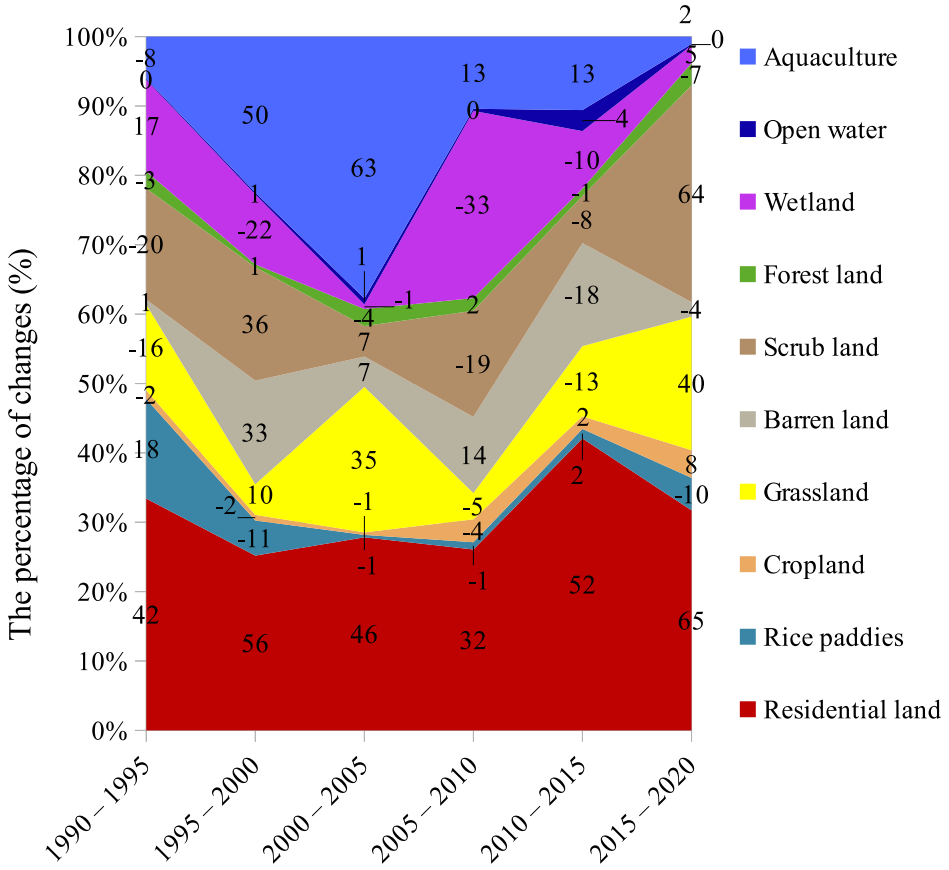


Figure 6.8: Temporal dynamics of net changes in LULC across Vietnam, extracted from the level-1 Vietnam-wide annual LULC data sets in the years 1990, 1995, 2000, 2010, 2015, and 2020. The data labels represent the percentage of changes (%) within five-year intervals. The positive and negative values indicate an increase and a decrease, respectively.

transformation, I created additional data in Figs. 6.10 and 6.11. It is noticed that a considerable proportion of forests was converted into croplands while a major driver of wetland loss resulted in the expansion of aquaculture. Residential lands mainly expanded on the areas of rice, croplands, and barren lands, which are located nearby coastlines. To acquire a more detailed visualization of change patterns, a few hotspot regions were extracted throughout the country to discuss the change pattern and processes.

6.4 Discussion

Large-scale annual LULC information is integral for understanding the land dynamic process, thus supporting the strategies of land management. In reviewing the literature, little consistent multi-spatio-temporal LULC data was found on a national or regional scale. In this study, a comprehensive framework is developed to produce consistently

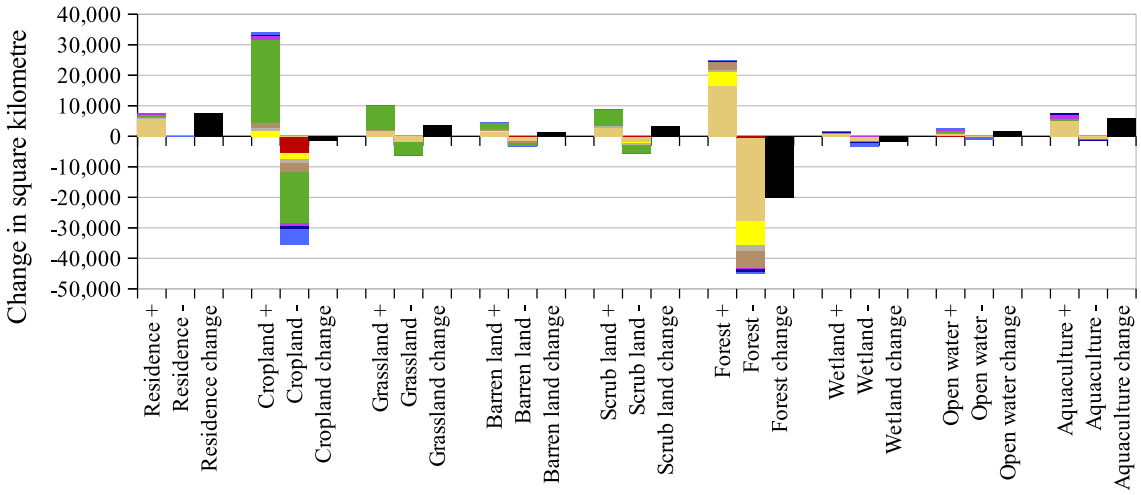


Figure 6.9: LULC gain/loss and conversions between 1990 and 2020; "+" means gain and "-" means loss in area (km^2).

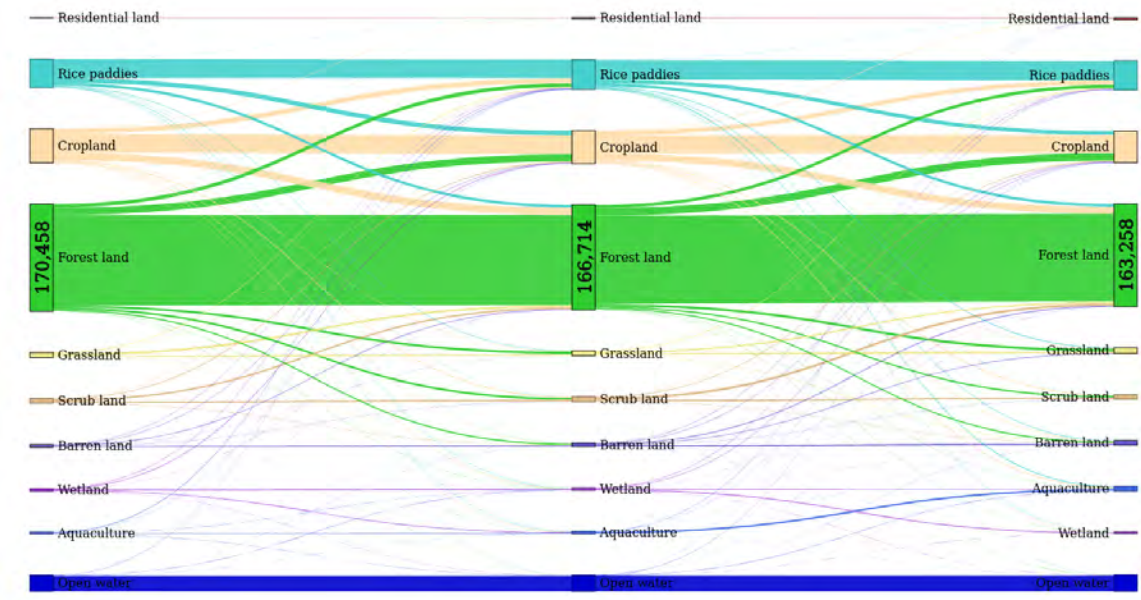


Figure 6.10: Transitions among different land types in Vietnam between 1990 and 2010. The numbers indicate the areas of forests, which is the most dominant land in Vietnam (km^2).

Vietnam-wide annual LULC data sets, using remote sensing and ground-based data. Results show that surface reflectance images can provide a coherent time-series data set as long as they are atmospherically corrected. Surprisingly, although all available Landsat images of the entire year are utilized, there are data gaps due to cloud and shadow masking areas. These gaps might affect the accuracy of classified maps even though gaps have been filled by ancillary data such as terrain indices. This issue may be explained by the fact that Vietnam is one of the cloudiest countries in the world [207].

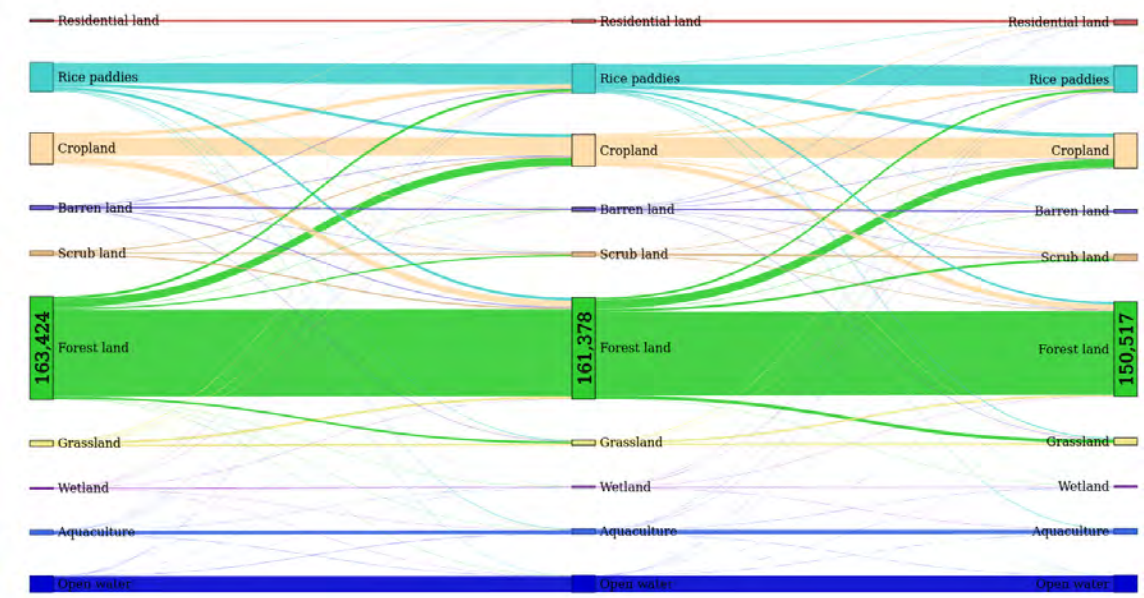


Figure 6.11: Transitions among different land types in Vietnam between 2010 and 2020. The numbers indicate the areas of forests, which is the most dominant land in Vietnam (km^2).

However, the harmonious blend of Landsat OLI, Sentinel SAR GRD and MSI has filled such missing-data gaps since 2015, which can also improve the accuracy of mapping (Figs. 6.5 and 6.6). Besides, the training migration model significantly reduced the cost and efforts in collecting training data.

Regarding change patterns, the rapid development of urbanization is considered an essential interest in Vietnam. The expansion of urban areas has frequently occurred in the capital and regional capitals, namely Hanoi, Hai Phong, Da Nang, Ho Chi Minh (HCM), and Can Tho cities. Herein, I analyse the process of change in HCM as a typical example. As shown in Fig. 6.12a, the growth of urban land has remarkably increased since the 1990s. This may be explained by the fact that the introduction of new policies known as "Renovation" (1986), which has promoted the development of socio-economic factors, followed by a massive population migration to cities [260]. The urbanization has primarily taken place on croplands, which agrees with the findings of previous studies [15, 260, 261]. The development is predicted to accelerate over developing regions, which causes the loss of croplands, and thus may threaten sustainability and livelihoods [262].

Another considerable change is the uncontrolled development of agricultural and aquaculture land, especially in the Vietnamese Mekong Delta (VMD). As part of the third-largest basin in the world, the VMD plays an integral role in the contribution of agricultural products, due to its favorable natural condition of 700 km coastline and a dense network of rivers. Rice and shrimp have contributed to a vital position in Vietnam's economic development for decades [263, 224]. However, the intensification of uncontrolled aquatic farming has caused significant changes in LULC across the

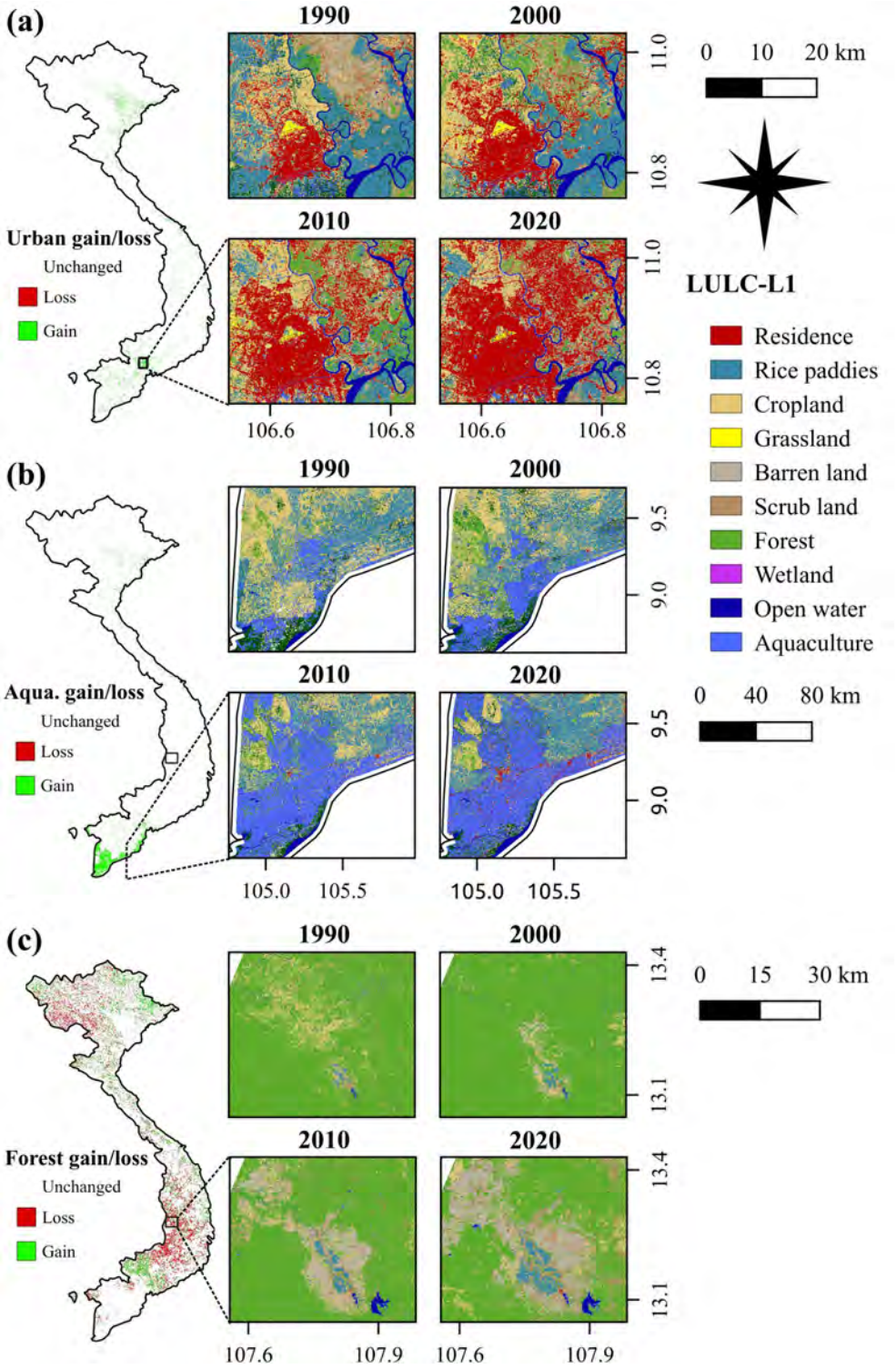


Figure 6.12: Spatial–temporal dynamics (left) and change pattern (right) of LULC in (**a**) residential land, (**b**) aquaculture land, and (**c**) forests land in Vietnam. This figure is generated using QGIS 3.18.0-Zurich while the country boundary is extracted from the GADM.

region, especially along coastal zones since 2000 (Fig. 6.12b). There are several possible explanations for this finding. In 2000 and 2001, the Vietnamese government proposed

resolutions 09/ND-CP and 1116, which replaced low-value (e.g. rice) to high-value (e.g. shrimp and fruit) agricultural production, encouraging farmers to transform certain coastal areas into aquatic production. Also, due to the high profits of shrimp (200,000 VND/kg) in comparison to traditional crops such as rice (5,000 VND/kg) [15], numerous inland areas were converted into aquatic farming. Local people illegally cut mangrove forests to expand aquaculture in several coastal regions. Since 2010, there was not only an increase in aquaculture, but also a significant conversion of other croplands into rice paddies (Fig. 6.12b). These changes resulted in the formulation of another policy aiming at the increase of intensive rice and fish farming in 2012 [264, 265]. These findings indicate that the development of socio-economic policies is considered the primary reason driving LULCC. Land policies, therefore, should be formed and implemented in serious consideration of regional socio-economic and environmental development.

Contrary to the expectations, this study indicates a net area loss of forests instead of the constant increase reported by the Vietnamese Ministry of Agriculture and Rural Development (MARD). Despite the net forest regrowth in 2000 and 2010, the forest cover has undergone a decrease in recent years. Also, the area of forests in this study is greater than the data reported by the MARD. These inconsistencies are due to several reasons, especially the difference in forest definitions. The MARD excluded agricultural (e.g., rubber), aqua-cultural ecosystems, scattered trees, bamboos, palms, etc. from forests34. These non-forest lands, covering a relatively large area of the country (e.g., 10 km^2 of rubber only; 2017), were highly dynamic [52] but not fully reported by the MARD. In 2008, the revised definition of forests set a minimum of 10% tree cover as forests, instead of 30% tree cover in the previous definition. Since 2016, they started to include certain agricultural lands managed by the Vietnam Administration of Forestry in the forest lands but without forest covering [265]. These revisions likely increased the reporting data of forests. Furthermore, the results show that deforestation occurred in numerous regions. Fig. 6.12c presents a representative example of forest loss in the central highlands. There has been a constant decrease in forest cover due to the expansion of rice paddies, barren lands, and croplands. This finding of forest loss corroborates the discoveries of a great deal of previous work in LULC observation covering Vietnam [266, 267].

Regarding limitations of this study, I could not estimate the benefit of the individual sensor's characteristics although the harmonious blend of the Landsat OLI, Sentinel SAR GRD and MSI images can fill data gaps and improve the accuracy of mapping classification. Moreover, instead of using ground-truth data, I validated the annual maps of the year 1990 to 2014 with the data collected by visual interpretation. Although great and careful efforts were applied in the collection procedure, errors might not be inevitable due to the restricted high-resolution images in Google Earth, especially before 2000. In addition, I have utilized the random forest algorithm to eliminate the less important input features, but there are relatively numerous remaining features, resulting in a high computational cost. Herein, the nationwide annual multi-category LULC maps and overall change detection were successfully developed, but the drivers

of LULC changes should be measured in details for individual land cover. Finally, deep learning neural networks are expected to be applied for large-scale LULC mapping.

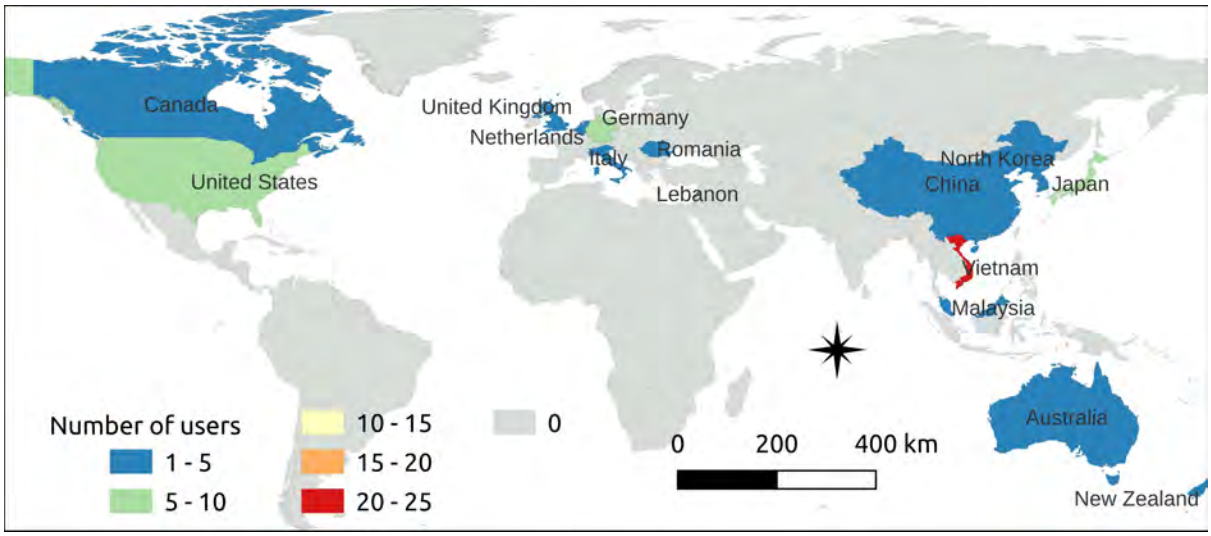


Figure 6.13: Geographic distribution of worldwide users of the JAXA HRLULC Map Products of Vietnam.

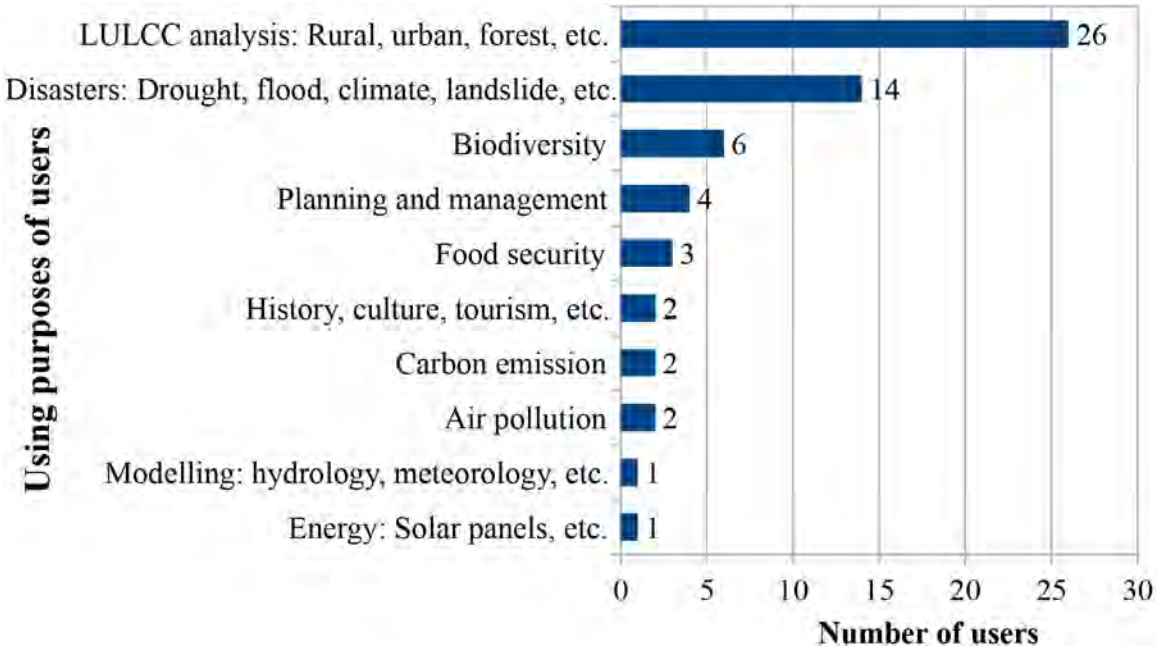


Figure 6.14: Purposes of worldwide users of the JAXA HRLULC Map Products of Vietnam.

Notwithstanding the above-mentioned limitations, the national database of annual LULC from 1990 to 2020 in Vietnam is essential for various applications. The database has recently published (Sep. 2021) for free use on the website of JAXA as a part of High-Resolution Land Use and Land Cover (HRLULC) Map Products, including several versions; my present database is HRLULC 30m resolution map of Vietnam

[1990-2020, each year] (ver.21.09:10 & 18 categories).Herein, some statistical information of the HRLULC Map Products is provided. Figure 6.13 shows geographic distribution (nations) of the users and figure 6.14 shows the purposes of the users from Mar. 2018 to Mar 2022. Among the purposes the majority are LULCC analyses, drought and flooding studies, climate change and adaptation, water resources, and natural resource management. This fact indicates the utility and importance of the free high resolution time-series land use/land cover maps.

6.5 Conclusions

Regularly updated and accurate LULC information is fundamental to interdisciplinary studies. The recent advancement of remote sensing and computational science has improved the mapping capacity of LULC. This study set out to develop a new framework for automatically monitoring nationwide annual LULC and provide the first VLUCDs over the past 30 years. To this end, I utilized ground-based data, the informative Landsat TM, ETM+ and OLI, and Sentinel SAR GRD and MSI images, after comprehensively assembling and preprocessing on the GEE platform. Then, I developed a new random-forest-based method and an automatic training migration model (ATMM) to map Vietnam LULC. The obtained VLUCDs had overall accuracies ranging from 85.7$\pm$1.3% to 92.0$\pm$1.2% with the ten primary dominant land use/cover and 77.6$\pm$1.2% to 84.7$\pm$1.1% with the eighteen secondary dominant land use/cover. This confirms the potential of the proposed framework for the systematically long-term monitoring of LULC in Vietnam. Results reveal that there was a decrease in the area of forests (19,940 km^2) and wetlands (1,914 km^2) whereas the area of aquaculture and urban increased approximately three and ten times over the three decades, respectively. The deforestation was mainly due to the expansion of croplands, which were in return replaced by numerous built-up areas. The rapid growth of aquaculture was considered the main driver of wetland loss. The explicit spatio-temporal benchmark of the VLUCDs can be utilized for a tremendous variety of applications in the research of environmental changes toward the Sustainable Development Goals. In addition, the ATMM allows analysts to remarkably save time, cost, and labor for collecting sufficient and representative training data. This proposed method is possible to apply for a multi-temporal LULC assessment at a broader scale.

Chapter 7

Connecting geospatial to local socio-economic and biophysical variables for quantifying the dynamics and determinants of forests changes across Vietnam over recent three decades.

Abstract:In the recent few decades, Vietnam has experienced a considerable change in land use/land cover (LULC), especially forest land. However, there is not a comprehensive analysis of the dynamics and drivers at the nationwide spatial scale over a long-term period. In this research, the socioeconomic and biophysical drivers of forest changes at the commune scale were estimated reliably. Utilizing the results of the Vietnam-wide annual LULC database available in the Japan Aerospace Exploration Agency (JAXA), the dynamic changes in forest land from 1990 to 2020 were first estimated reliably. To decide the major drivers of the changes, a synthesis of case studies working on the analysis of the forest changes in Vietnam at various spatial levels was conducted successfully. Subsequently, a machine learning technique was adopted to measure the drivers of the forest changes. The results indicate that although the forest area has increased from 2005 to 2010, it has undergone a decrease over the full study period. There is a dramatic conversion between forest and agricultural land, especially in the North-West and Central Highlands. This conversion is mainly driven by agricultural expansion/shifting, topographic position index, accessibility/infrastructure, population growth/migration, and distance to systems such as irrigation, drainage, and mining/industry. The identification of the drivers in this study is likely to help enhance the accuracy of the land use/land cover change prediction. These findings provide coherent evidence-based information about the dynamics and drivers of forest changes at the nationwide spatial and decadal temporal scales and thus can support informing land policies in Vietnam.

7.1 Introduction

A change in land cover refers to the alteration of the biophysical surface of the Earth while land-use changes can be defined as variations in the way of using and managing a specific land area. Due to a direct connection between these two terms, land use/land cover change (LULCC) is generally used in the literature. One of the greatest changes in LULC is forest transition, especially forest degradation and deforestation. The unexpected change of forests impacts adversely on regional climate extremes [268], biodiversity [269], and ecosystem services [270]. As a result, proper estimations of forest changes and drivers is a central topic in international discussions and projects such as Paris Agreement and Reducing Emissions from Deforestation and Forest Degradation (REDD+) [13]. Long-term monitoring and interpreting of forest changes and drivers at a large scale and a fine spatial level provide evidence-based information for these discussions and projects. This information is also essential to making decisions in informing policy, planning, and managing forest resources.

Despite its importance, the detailed long-term estimation of forest transition is challenging, especially in developing regions, which frequently have a lack of information about the accurate datasets of forests for a long period. Recent evidence suggests that many developing countries are short of professional competence to generate fine spatiotemporal and accurate forest mapping resources [14]. For example, a careful synthesis of over 200 case studies focusing on forest and agricultural land in South and Southeast Asia shows that although numerous studies have been conducted in certain countries such as India and Indonesia, few studies have been done in the other countries of the regions such as Cambodia, Laos, and Vietnam [225].

Vietnam has a major resource of tropical forests, but it has undergone highly dynamic changes over the recent three decades [271, 1]. These changes are affecting adversely the long-term balance of ecosystem services [272] and climate [273]. Unfortunately, very few accurate and consistent datasets of forests have been conducted for a long-term period across the country. According to the report of the general statistics office on forest management, available five-year-interval national-scale forest maps are frequently utilized by local policymakers. These maps are not always meet the requirement of all users and are not frequently published to various institutes, organizations, and analysts. Up to date, some data sources have developed, but their differences, for example, production methods, producers, and input data usage result in inconsistencies in the resultant data products [14]. Likewise, regional and global products have limitations when they are applied for analyses at a local-level scale. They cannot reflect accurately at a local and a national scale due to their coarse spatial resolution and low accuracy [274, 1, 18]. These limitations challenge the effective usage of the existing forest data sets to provide scientific information for decision-makers to inform coherent policies, do long-term strategic planning, and offer corporate management.

Recently, a database for the Vietnam-wide annual land use/cover data sets (VNLCD) have been first created from 1990 to 2020 [20]. Although the database does not focus only on forests, it is beneficial for the estimations of forest changes and drivers. First,

the accuracy of the forest class is high. The database contains variously different LULC, which allows for measuring the transition between forests and other land types.

Although there is a manuscript talking details about how the database has been conducted [20], a brief summary of the generated process is provided herein for a general understanding of the LULC data. Specifically, a state-of-the-art framework was established utilizing various ground-based and satellite-based data sources, including Landsat and Sentinel sensors. The framework contained an automatic training migration model and an optimal approach for post classification optimization. The established approach could address issues such as cloud-contaminated problems by employing the availability of multiple spatiotemporal data, thus improving the accuracy of mapping. The satellite data were separated into wet seasons (May – November) and dry seasons (December to the next April) to identify the changes in LULC phenology, including forest and agriculture. The automatic training migration model offered a practical mode for collecting reference data over the recent three decades. Regarding the land cover classification system (LCCS), I applied the standard land cover scheme, viz. Land Cover Classification System with necessary modifications according to the local biophysical environment and end-users' suggestions in Vietnam. That is, ten land cover types (Fig. 7.4) were decided. Applying a benchmark [63], the quality of the VNLCD was statistically validated with ground-based reference data derived from extensive field surveys (from 2015 to 2020) and inventory data over the country. Different metrics were provided, including overall accuracy, kappa coefficients, and standard error of the mean. The VNLCD was published on the Japan Aerospace Exploration Agency website.

This paper aims to explore the dynamics and drivers of forest changes across mainland Vietnam. To this end, I first analyze the forest change from 1990 to 2020 over the study area, utilizing the VNLCD. I then conduct a synthesis of the case studies to identify major drivers of the forest changes before deciding the selected major drivers for this study. The remaining main part of the paper is to quantify the drivers using a logistic regression model. This is the first study to undertake a longitudinal analysis of the long-term dynamics and drivers of the forest changes in Vietnam. It provides an important opportunity to advance the understanding of the pattern, rate, and process of forest changes at a commune level, which can provide insights into informing more coherent policy to enhance the management of forest in Vietnam and expecting to apply in other tropical regions.

7.2 Materials and methods

7.2.1 Forest

Forests provide with different valuable services such as soil erosion protection, habitat provision for animals and humans. In the field of ecology, various definitions of forests are found. In this study, I adopted the definition of forests provided by the Food and Agriculture Organization of the United Nation (FAO). Herein, forest is defined as "land

spanning more than 0.5 hectares with trees higher than 5 meters and a canopy cover of more than 10 percent, or trees able to reach these thresholds in situ. It does not include land that is predominantly under agricultural or urban land use". Also, mangrove is not defined as forests in this study.

7.2.2 Forest diversity and change quantification

I first estimated and removed possible errors in the VNLCD. Although the VNLCD had a high level of accuracy, there might be misclassifications due to the effects of different climate situations and changes in plant phenology. The misclassifications would happen a few times in the time-series maps. To eliminate the errors, I created occurrence maps derived from the 1990 - 2020 time-series maps with arbitrary quantiles, including the first quartile, third quartile, and ninetieth percentile. For each land type, I overlaid the occurrence map on the fine-resolution images in Google Earth to identify misclassifications. With a trial-and-error technique, I found that the third quartile was the best threshold to eliminate possible misclassifications. Subsequently, I quantified the dynamics of forestland, including gain, loss, and major conversion from 1990 to 2020. I also estimate the diversity of forest changes by computing the number of changes between forest and non-forest land at a commune scale over the recent three decades (Fig. 7.2).

7.2.3 Hot spot analyses

Changes in LULC are interests, but the change might not be statistically significant, biasing the estimation of main drivers. To increase the reliability of the estimation, I employed a hot spot analysis to identify the statistically significant changes. Although several approaches have been established to estimate hot spots, Getis-Ord Gi* [275] has been practically applied in the geographic research [276]. Hence, I employed the Getis-Ord Gi* method to compute z-score and p-value for each region of interest (ROI). An ROI that has a high z-score and a small p-value is a significant hot spot while a cold spot has a low negative z-score and a small p-value. Herein, the ROI that has a z-score greater than three was defined as a hot spot. The hot spot analysis was performed at the commune level known as the ROI. I computed Getis-Ord Gi* values as follows (Eq. (7.1); Eq. (7.2); Eq. (7.3)). The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{n\sum_{j}^{n} w_{i,j}^2 - (\sum_{j=1}^{n} w_{i,j})^2}{n-1}}} \tag{7.1}$$

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{7.2}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \tag{7.3}$$

where G_i^* is Getis-Ord local statistics; x_j is the attribute value for feature j; $w_{i,j}$ is the spatial weight between feature i and j; and n is equal to the total number of features.

The hot spot analysis was done by using ArcGIS Desktop 10.8.2. First, all the proxy driver data were projected to WGS 84/UTM zone 48N, using Data Management Tools. Then, the Incremental Spatial Autocorrelation tool was utilized to select an appropriate distance threshold. The Incremental Spatial Autocorrelation tool measured spatial autocorrelation for a series of distance increments and reported each distance increment. Finally, Hot Spot Analysis (Getis-Ord Gi*) tool was employed to identify the statistically significant spatial cluster of high values (hot spots) and low values (cold spots). The tool created a new output feature class with $z-$scores, $p-$values, and confident level bins of Getis-Ord Gi*. How Hot Spot Analysis (Getis-Ord Gi*) works can be referred from ArcGIS Pro Tool Reference (here).

7.2.4 Case-study synthesis

I conducted a review of 47 publications to detect the drivers of spatio-temporal changes in forest land in Vietnam. I used the query expression "TI = (drivers OR drivers OR causes OR dynamics AND land*) AND TS = (Vietnam OR "Viet Nam" AND land*) AND TS = (*fores* OR defor* OR refor* OR degrad*)" to search the publications in Web of Science and Google Scholar databases. Then, I eliminated duplicates before gathering 456 publications in total, of which about 357 publications worked on LULC, including forests. Reading the titles and abstracts of the 357 publications, I eliminated 189 publications, which did not discuss forest topics. Among 168 publications remained, the third-fourth mainly worked on the general LULC and LULCC assessment. Finally, I selected 47 publications, which worked on estimating the drivers of forest changes. After the careful scanning of the full publications, I noted study topics, main drivers, research places, research periods, and approaches applied by the selected publications. If a major driver was noted in the publications, I defined it as a major driver. Otherwise, the mentioned drivers in the publications could be seen as the main drivers. Similar drivers were merged to generalize the drivers. They are presented in Table A.6.

7.2.5 Proxy data drivers

The main biophysical and socio-economic drivers (Section 7.2.4 were derived from different sources. Most of the data were extracted annually from the General Statistics Office of Vietnam (1990 - 2020; household surveys). They were originally presented in various formats including vector, tabular, and raster formats with different spatial resolutions. After that, considerable efforts were made to interpret the data before they were used for the coming analysis. Since the data had different spatial resolution, I kept the original spatial resolution and computed mean values at the commune scale. Then, for the time-series data (e.g., climate) I computed mean, rate of change, and standard deviation values over the study period (1990 - 2020). For the constant data (e.g., elevation and slope) which were insignificantly alter over the study period, I

computed mean values in the area of interest. Details about the proxy drivers are presented in Table A.7.

7.2.6 Handling the multicollinearity of the driver proxy data and modelling selection

Multicollinearity causes the inaccurate assessment of the regression coefficients. Herein, I first detected the high level of multicollinearity between the driver variables by computing the variance inflation factors (VIF) in Python (version 3.8.5), using the variance_inflation_factor() function from the statsmodels (version 0.12.0) library. I then excluded the driver proxy variables which have a high level of multicollinearity or VIF > 10 A.1 [277]. I acknowledged this limitation to reduce the analysis bias of the regression model.

It is noted that even though several models have been established to estimate the drivers of LULCC, there is not the best model. The performance of a model varies according to certain research cases. When a model is selected, it should be considered major factors, including reliability, generalizability, computational cost, and stability [278]. Recently, a logistic regression model has been commonly applied for identifying the drivers of LULCC, due to its clear understanding and impressive performance [225]. A logistic regression model, therefore, was applied in this study. The logistic regression model was developed in Python (version 3.8.5) and scikit-learn (version 0.23.2) library.

7.3 Results and discussion

7.3.1 Synthesis of the case study

The detailed drivers of the forest gain and loss from the synthesis of 47 case studies are presented in Table A.6. Only eight studies worked on a national scale (including a regional scale covering Vietnam) while the others were conducted at smaller areas from a village to provincial scale. Most studies quantified the drivers utilizing quantitative methods such as logistic regression, multiple correlation analysis, and linear regression whereas ten studies interpret the drivers based on interviews, fieldwork surveys, and historical data. There was a larger number of work on the drivers of forest loss (deforestation) in comparison to forest gain (reforestation/afforestation). The frequent distribution of the main drivers is presented in Fig. 7.1. They include socioeconomic drivers such as policy and income, and biophysical drivers such as climate and topography. As can be seen from the figure, the socioeconomic drivers were mentioned more frequently than the biophysical drivers (47 against 33, respectively). Regarding the forest loss, frequently mentioned drivers were agricultural expansion/shifting, policy/tenure, and accessibility/infrastructure. Meanwhile, policy/tenure and plantation were the most common drivers of deforestation. The most interesting aspect of this table is that tenure/policy decisions were involved actively in both the forest gain and

loss. In contrast, plantation solely contributed to reforestation/afforestation while fuelwood/logging/charcoal, land size, climate, urbanization, etc. were only mentioned in the deforestation studies.

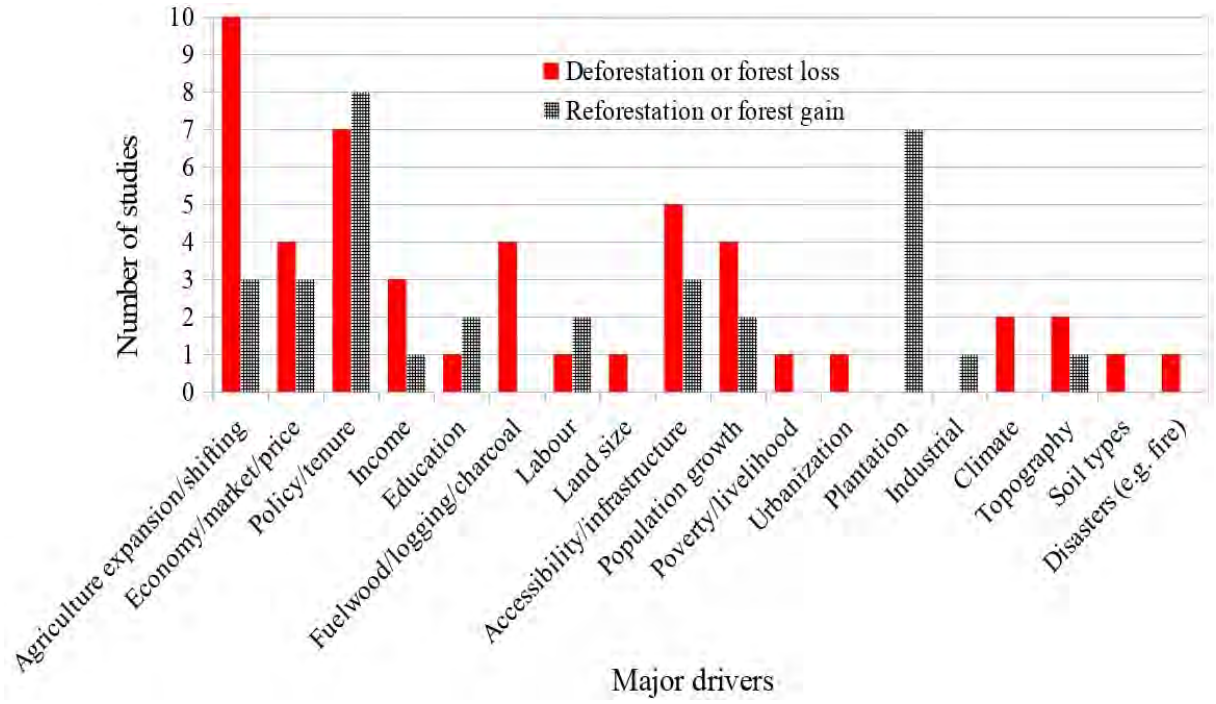


Figure 7.1: Frequency distribution of the main drivers extracted from the synthesis of case studies in Vietnam. More details about specific drivers can be found in Supplementary Text S1.

7.3.2 Land use/land cover and forest dynamic changes

The conversion of different land use/land cover types between 1990 and 2020 is presented in Fig. 7.2. As can be seen from the figure, major LULC changes were forests, residential areas, and aquaculture land. While the residential and aquaculture land experienced increased trends, the area of forest significantly shrank over the study period. This finding was also reported by Hansen et al. and Leinenkugel et al. [18, 279]. This result is, however, likely contrary to that of Meyfroidt et al., Jadin et al., and the report of the Vietnamese Ministry of Agriculture and Rural Development (MARD) who found that the area of forests increased [150, 280]. These differences are due to several rationales. First, Meyfroidt et al. and Jadin et al. only estimated within a shorter period from 1990 to 2010. Over this period, our previous results also showed an increase in forests in Vietnam [20]. Second, there was a slight difference in the definition of forests provided by this study and the MARD who did not consider rubber, scattered trees, etc. as forests. These LULC types changed dynamically but were not fully included in the report of forests by the MARD [52]. Interestingly, the MARD included certain areas administrated by the forestry division into the forest land although there may be no forests on the land [266].

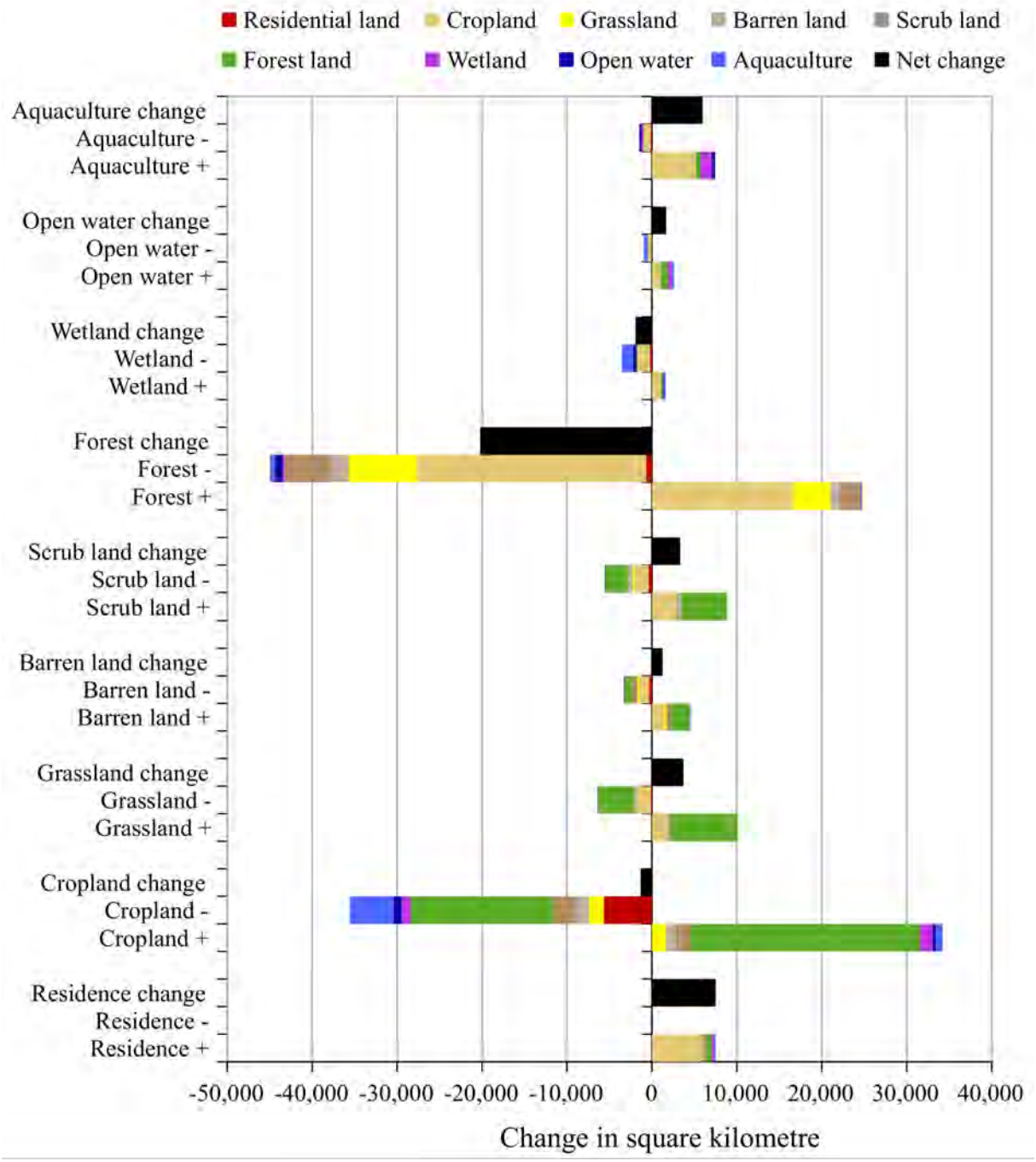


Figure 7.2: Gross losses, gross gains, and net change in land use/cover at Vietnam-wide scale between 1990 and 2020. Land use/cover types are showed in different colors. Areas are measured in square kilometre. The left represents losses while the right represents gains. Black bars present the net changes (gross gains - gross losses) of different land use/cover types.

Spatial distributions and dynamics of forests in 1990 and 2020 across mainland Vietnam are presented in Fig. 7.3a and Fig. 7.3b, respectively. As can be seen from the figures, forests are mainly distributed in the northern and west-center regions that had high altitudes/elevation. It is apparent from Fig. 7.3a and Fig. 7.3b that very few areas

of forests were detected in delta areas, especially the Red River Delta and Vietnamese Mekong Delta/the South. Fig. 7.4a and Fig. 7.4b present apparently the fractional distribution of forest losses and gains at the commune-level scale. Meanwhile, the hot spots and cold spots of forest gains and losses are shown in Fig. 7.4c and Fig. 7.4d. What stands out in the figures is the most dynamics and losses of forests in the Central Highlands and the North-West. This finding was also reported by Imai et al., and Kissinger et al. [281, 282]. This result may be explained by several reasons. As can be seen in Fig. 7.2, there is a major conversion between forests and agricultural land. The central Highlands provide for commercial agriculture with reasonable conditions such as adequate precipitation, moderate temperatures, and good soils. In this region, dominant commercial agriculture lands are rubber and coffee, which have significantly expanded between 2005 and 2015 [282] .

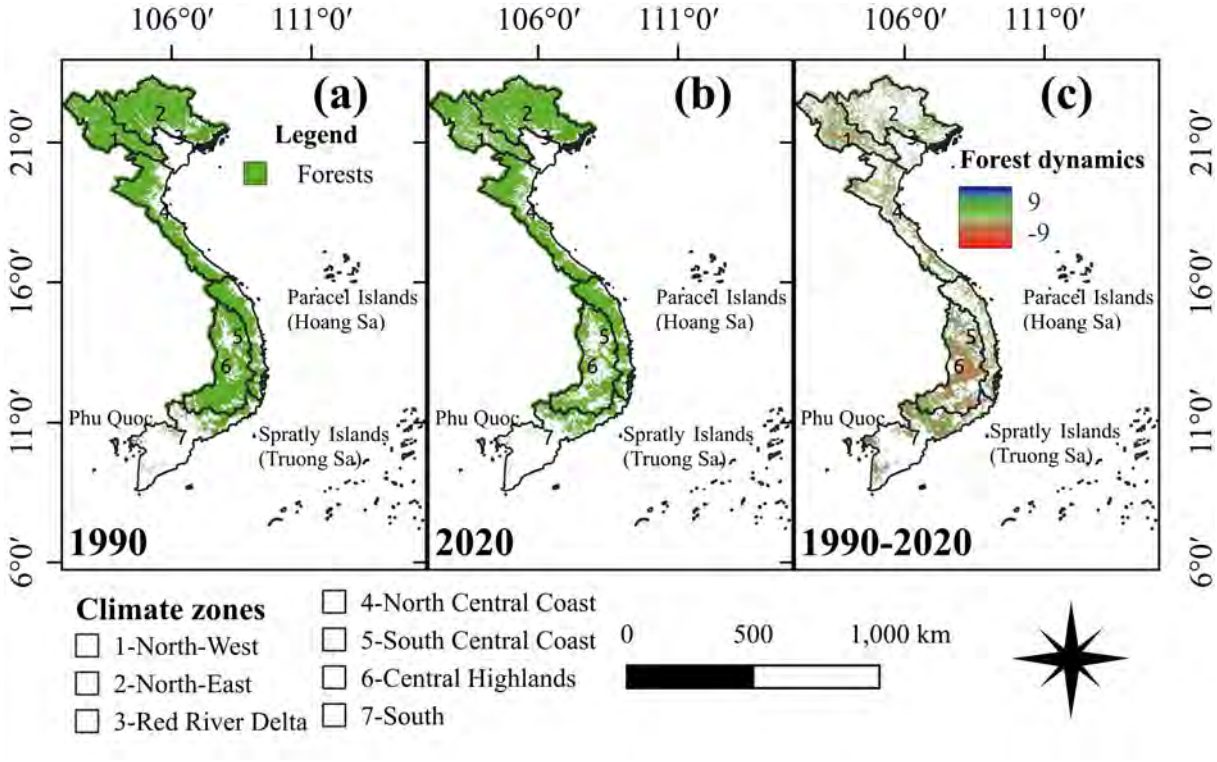


Figure 7.3: Spatial distribution of forests in mainland Vietnam in 1990 (**a**), in 2020 (**b**), and the dynamic change of forests over recent three decades (**c**). The red color represents a gain while blue color represents a loss over the study period. The data were analysed using the Vietnam-wide annual land use/cover data sets [1]

7.3.3 Measurement of the major drivers

Standardized coefficients are used to compare the relative effect of each individual driver variable to the forest change area. Specifically, before running the multiple least squares model, I standardized the driver variables to z-scores. Herein, the coefficients that I obtain from the model are standardized coefficients, which have standard de-

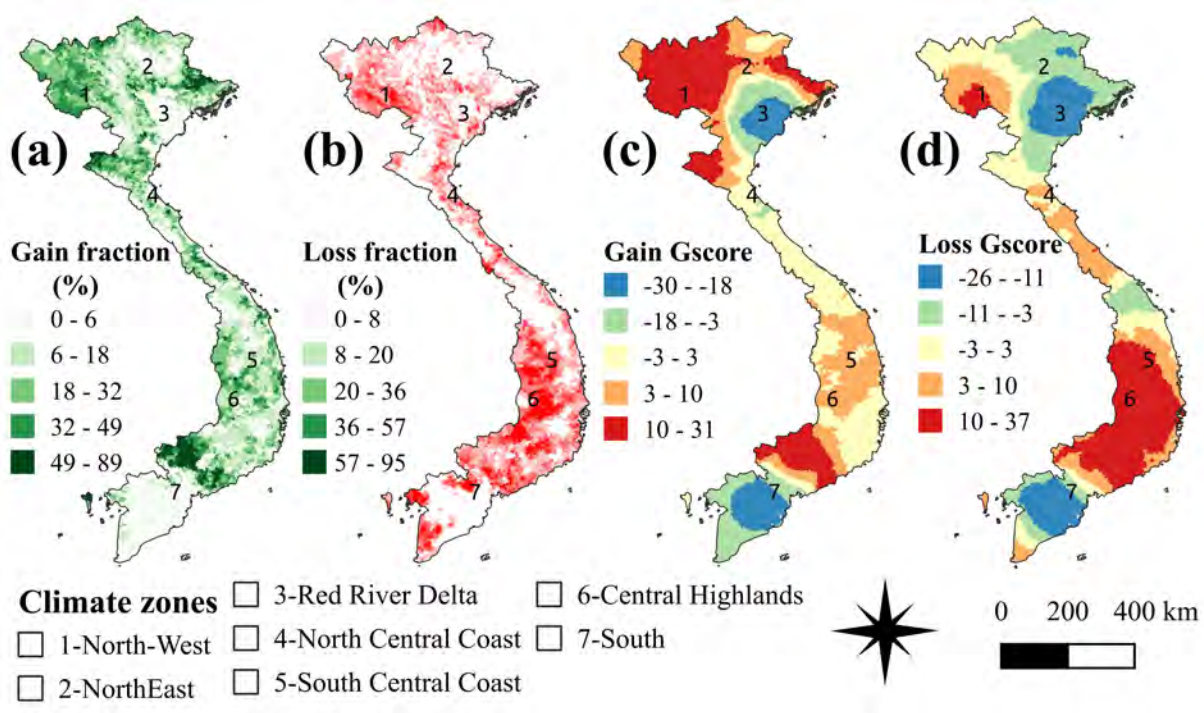
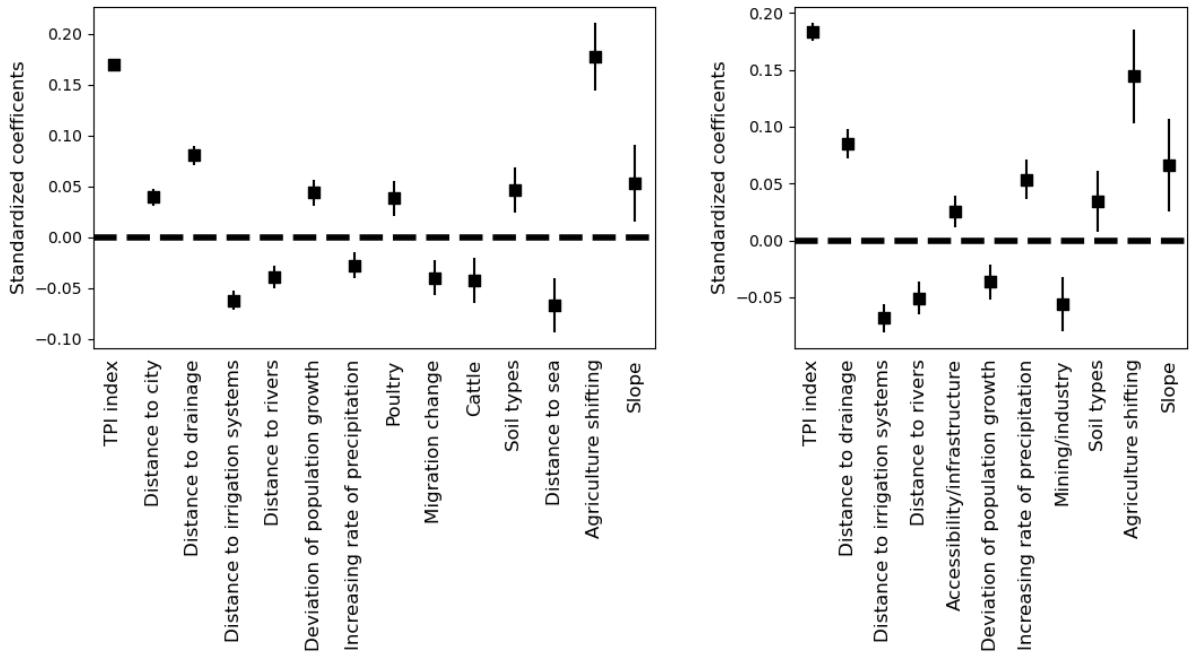


Figure 7.4: Spatial pattern of forest change in mainland Vietnam between 1990 and 2020. (**a**) the fraction of forest gain (%) and (**b**) the fraction of forest loss (%). (**c**) presents hot spot analysis of forest gain and (**d**) shows hot spot analysis of forest loss. Red color represents hot spots while blue color represents cold spots.

viations as their units. This means the driver variables can be precisely compared to each other, although they are measured on different units and scales. The higher the absolute value of the coefficient, the stronger the impact. A positive value of the coefficient means that a higher value of a driver variable causes a greater change in the forest area. In contrast, a negative value of the coefficient means a higher value of a driver variable has a lower change in the forest area. Herein, the standardized coefficients are presented in Fig. 7.5.

Forest gain

Fourteen major drivers significantly relate to the forest gain, of which a half is biophysical drivers and the others are socioeconomic drivers (Fig. 7.5a and Table A.8). Eight major drivers have positive relationships with the forest gain; of which the most important driver is agricultural shifting. This result can be explained by several possible reasons. As can be seen from Fig 7.2, there is a considerable conversion between forests and agricultural land. In other words, agricultural shifting can significantly cause changes in forest areas. This result also corroborates the findings of a great deal of the previous work on forest transition in Vietnam, which are described in detail in section 7.3.1. For example, Ashraf et al. has identified a net increase of forests in certain areas of Vietnam due to the major shrinkage of agricultural land [283]. What is surprising is that the topographic position index (TPI) and slope positively have substantial impacts



(a) Forest gain

(b) Forest loss

Figure 7.5: Most prominent drivers of forest changes: Forest gain (left) and forest loss (right). Square dots indicate the coefficients while the bars indicate the uncertainties at the 95% confidence interval.

on the forest gain, although my synthesis shows few studies that have mentioned these drivers. A possible explanation for this might be that forests are mainly located in mountainous areas that have high values of slope and TPI indexes [150, 284]. Other positive relationships between the forest gain and the driver variables are distance to the city, distance to drainage, deviation of population growth rate, poultry, and soil types. Among these drivers, the distance to drainage is the strongest factor correlating to the forest gain. This finding was also reported by Nguyen et al. [284]. It seems possible that drainage systems are frequently constructed in flat regions such as cities and densely populated residential areas that are non-forest land [285]. In terms of negative relationships, there are six major drivers, of which distance to the sea is the most considerable driver. This means forest gains occur in regions close to the sea. This result may be explained by the fact that planting forests are considered an effective solution to prevent natural hazards and disasters. Several reports have shown that forests are planted in coastal areas for protecting erosion and typhoons [286, 287].

Forest loss

Standardized coefficients corresponding to forest loss are shown in Fig 7.5b and Table A.9 with seven positive relationships and four negative relationships. The most interesting finding was that the topographic position index (TPI) also has the most positive relationship with forest loss, indicating that forest loss frequently occurs in regions that

have high values of the TPI. However, I have not found studies that identify the relationship between deforestation and TPI while slope has been utilized instead [222, 78]. In this study, the slope has a strong positive correlation with forest loss.

It is not surprising that agricultural shifting was found to cause a significant loss of forests. This finding broadly supports the work of other studies in this area linking agricultural expansion/shifting and deforestation [281, 266, 280, 288]. There are several types of agricultural expansion/shifting. For example, in the Central Highlands (e.g. Dak Lak, Dak Nong, and Binh Phuoc), a large number of forests areas have been converted into coffee and rubber since the 1990s [267]. Specifically, Dak Lak has relatively high elevation, fertile soil, and cool weather which offer a suited environment to coffee crops, especially Robusta coffee [289]. In contrast, with a lower elevation and a higher temperature, Binh Phuoc offers a favorable condition to rubber plantations [290]. Also, the quickly growing global demand for coffee and rubber resulted in the expansion of these crops. It is also supported by governmental policy to convert poor forests into coffee and rubber plantations [267, 290].

Other major positive drivers correlating to forest loss are accessibility/infrastructure and increasing rate of precipitation. These results support evidence from previous observation [291, 288, 292, 222]. Particularly, Nguyen et al. found that forest loss occurs more frequently in remote access areas, for example, far from roads, drainage, and a densely populated residence [284]. Meanwhile, the positive relationship between deforestation and the increasing rate of precipitation indicates that wetter areas are favorable conditions for forest transition. A possible explanation for this might be that increasing rate of precipitation might introduce extreme climate conditions, such as typhoons and floods, which may reduce forest productivity and degradation, and thus may cause deforestation [67, 225]. Another possible explanation for this is that the major degradation or loss of forests results in changes in precipitation [293, 294, 295]. The literature also reported an increase in extreme climate events, especially in the Central Highlands which is a hot spot of deforestation [296, 297].

Regarding the negative relationships between drivers and forest loss, there are four major drivers, namely distance to irrigation, distance to rivers, deviation of population growth, and mining/industry. It means that the loss of forest frequently happens in regions closer to rivers, irrigation systems, mining/industry zones, and diffusely populated areas. Although these drivers have not been mentioned in my synthesis of case studies on forest transition in Vietnam, these results match those observed in earlier studies in other regions [195, 298].

7.3.4 Significance of field surveys to land use/land cover changes and the influencing drivers

Data from a five-year survey is fundamental to the results of this study. First, it provided ground-truth sites to ensure the quality of my LULC maps produced from the remote sensing data. It had a pivotal role in validating maps from 1990 to 2014. In this period, although I could not conduct surveys, I could interpret historical somehow

LULC in the fields by local knowledge and perception. Specifically, during the recent surveys, I had first observed the LULC and then conducted interviews with the elderly and key land management informants. They gave us the convincing historical stories of how LULCC in the regions. Without these stories, I could not ensure the accuracy of my maps, the change of LULC, and the drivers. Take the Vietnamese Mekong Delta as a typical example. I have a separate paper discussing the transformation of rural landscapes between 1990 and 2019 in details [299]. Herein, I brief local knowledge and perception on land changes in specific regions that I have conducted the surveys. For example, in the coastal zones such as Ca Mau province, the considerable expansion of agriculture and aquaculture was mainly due to local and national policy implementation. Specifically, high demands on timber and charcoal products had caused a significant loss of mangrove forests since 1975. Then, the introduction of new irrigation systems and increase in the price market of aquaculture resulted in the substantial conversion wetlands/mangrove forests into fishery farming practices in 1980. From the 1990s, the government introduced a co-system cultivation allowing the locals to convert (20 - 40%) mangrove forests to other land types. However, some locals converted more than 40% of mangrove forests to other lands. For economic development, the government introduced policies (i.e., Resolution 09/NQ-CP (2000) and Resolution 1116/QĐ-CTUB (2001)) allowing locals to convert rice to aquaculture (e.g., shrimp) which had a higher economic value. Recently, due to adapt climate change impacts, new policies have informed, which accelerating the complexities of LULC change in the region. This interesting information is only a few stories that I have learned for my field surveys. It means field surveys are essential for understanding the influencing drivers of LULCC, which may not accurately measure, such as policy.

7.3.5 Limitations and future work

The generalisability of these results is subject to certain limitations. First, although I have collected the most essential drivers mentioned in my synthesis of the case studies and added new drivers, I could not quantify some important drivers such as forest management policy/tenure, globalization, disasters, and economy/market/price. These drivers may have profound effect the transition of forests [284, 300, 206, 150, 222]. In this study, the data of these drivers are not available at a proper spatiotemporal scale; or some have high multicollinearity, causing difficulties in the precise measurement of the impact of individual drivers; we, therefore, have excluded them from the analysis model. The lack of these drivers might impact the quantitative results. Secondly, the spatiotemporal difference of some collected drivers may cause uncertainties in the analysis of the drivers. Specifically, most of the drivers are derived from the General Statistics Office of Vietnam, which have been calibrated by the office; some are extracted from the WorldClim 2 database [301] or the satellite-based information (e.g. SRTM) [302]. These data may have relatively low quality and added further uncertainties in this work. However, the current shortage of the properly available data does not permit us to overcome this limitation. This would be a fruitful area for further work.

7.4 Conclusion

The present research aimed to analyze the changes and drivers of the land use/land cover changes (LULCC) focusing on forests in mainland Vietnam. To our best knowledge, this is the first work on the analysis of inter-annual dynamics of forests at the national scale from 1990 to 2020. Results have shown that Vietnam experiences a considerable change in LULC over the recent three decades. There is a significant transition between forests land and other lands, especially agriculture, grassland, and shrub-land. One of the more significant findings to emerge from this research is that the forest change has occurred in differences in space and time. The North-West and Central Highlands are the most dynamic change areas, which are considered hot spots. The study has also shown that although a net gain of forests has been reported in a short period (e.g. 2005 - 2010), Vietnam has undergone a net loss of forests from 1990 to 2020. The drivers of forest changes are various. The synthesis of case studies on forest transition has indicated that previous studies mainly focused on socioeconomic drivers. However, my results show that biophysical drivers also have significantly contributed to forest changes. In general, both biophysical and socioeconomic drivers are highlighted in this study. The development of the economy requires more food and agricultural products, forcing the extensive conversion of forests into other land types if biophysical conditions are reasonable for production. The second major contribution of my synthesis was that it broadens a more general knowledge of the drivers of the forest changes. These findings have significant implications for the understanding of how effectively to manage forest resources in Vietnam. They may be of interest to other countries. The identification of socioeconomic and biophysical drivers in this study is likely to help enhance the accuracy of the land use/land cover change prediction.

Chapter 8

Conclusions and Remarks

8.1 Summary of key findings and conclusions

The purpose of the current research was to analyze the dynamic changes of LULC in mainland Vietnam. To this end, I first built a national dataset of reference data derived from multiple sources, including intensive surveys over five year period (2015-2020), statistical inventory data, and super-high-resolution images from Google Earth. Together with our systematic literature review, this dataset played a critical role in revealing the limitations of the previously available LULC databases in mainland Vietnam. The review has shown that very few studies have conducted a comprehensive analysis of LULCC on a Vietnam-wide scale. Some focus on specific LULC types such as forests, urbanization, and croplands. The widely available LULC maps covering the whole country are regional or global products. These products have limitations such as low spatiotemporal resolution, differences in LULC schemes, and low accuracy/without complete accuracy assessment. As a result, these products do not reflect accurately at the local and national scales. Let's take a typical example, a classification of orchards can be significant for scientific communities in developing world while it is not a distinguishing class in the existing land cover products, which are widely used currently.

To overcome the above-mentioned major limitations, one of the more significant outcomes to emerge from this study is that an innovative comprehensive framework was established for an accurate LULC analysis across mainland Vietnam. First, based on the previously popular LCCSs and indigenous knowledge obtained from the intensive surveys, a new classification scheme was introduced for the case of Vietnam, which may match the major demand on the basic project requirement of the LULC end-users and map suitability. Second, a new classification approach (named "ELUC") was successfully developed to map accurately pilot areas in Vietnam. This approach can be applied with various classification algorithms and data sets to simplistically generate new classifications or renew existing classifications with a remarkable accuracy improvement. Specifically, applying miscellaneous features of Landsat-8 images, the ELUC of a random-forest-based algorithm produced a sequence of single-time classifications with a mean overall accuracy of 84%. Through the study period, this sequence

was then joined to achieve a final classification that reached an overall accuracy of 94%. In compared with an previous approach, which used Kernel-Density-Estimation method, I found that the ELUC of the random-forest-based algorithm had a 5% overall accuracy higher. These outcomes confirm the effectiveness of the ELUC for a remarkably consistent land use/cover estimation in a data-rich environment. Third, incorporating the ELUC, multiple geospatial data sources, and an automatic training migration model (ATMM), I quantified annual LULC for the whole country over the recent three decades. Obtained LULC maps had overall accuracy ranging from 85.7$\pm$1.3% to 92.0$\pm$1.2% with the 10-class primary dominant land use/cover and 77.6$\pm$1.2% to 84.7$\pm$1.1% with the 18-class secondary dominant land use/cover. These results demonstrated the potential of the proposed framework for the systematically long-term monitoring of LULC in Vietnam. Particularly, it successfully offered an explicit spatiotemporal database (named "VLUCDs") of LULC in Vietnam, which may be utilized for a tremendous variety of applications in the research of environmental changes toward the Sustainable Development Goals. In addition, the ATMM allows analysts to remarkably save time, cost, and labor for collecting sufficient and representative training data.

This study has also identified the patterns and processes of LULCC in Vietnam. The most dominant LULC was forest, accounting for an approximately half area of the entire country. This was followed by croplands (16.3%), rice fields (14.2%), and open water (including parts of saltwater, 8.1%). Grassland and scrubland occupied a relatively similar proportion (2.8%) while the smallest LULC was residential areas (1.3%). Interestingly, there was a steady decline in forest areas from 170,458 km^2 in 1990 to 150,517 km^2 in 2020. The area of wetlands experienced a slight increase from 4,404 km^2 in 1990 to 5,138 km^2 in 1995, followed by a continual decrease to 2,490 km^2 in 2020. In contrast, there was a sharp increase in the area of aquaculture and residential land, accounting for approximately three and ten times over the three decades. The area of open water showed a slight rise while there was a small fluctuation in the area of the other land types. Interestingly, LULCC in Vietnam varied according to time, space, and land types. Regarding spatial dynamic changes, the western North, the Central Highlands, and South were the most dynamic areas. There was a fundamentally dynamic conversion between different land types.

The most obvious finding to emerge from this study is that I quantified changes and drivers of the major land use/land cover changes (LULCC) with a focus on forests. To our best knowledge, this is the first work on the analysis of inter-annual dynamics of forests at the national scale from 1990 to 2020. Results have shown that there was a considerable conversion between forest lands and other lands, especially agriculture, grassland, and shrub-land. The forest change varied according to space and time. The western North and Central Highlands are the most dynamic change areas, which are considered hot spots. Although a net gain of forests had been reported in a short period (e.g. 2005 - 2010), Vietnam underwent a net loss of forests from 1990 to 2020. The drivers of forest changes were various. Our synthesis of case studies on forest transition drivers indicated that previous research primarily focused on socioeconomic drivers. Our results, however, showed that biophysical drivers also have significantly

contributed to forest changes. In general, both biophysical and socioeconomic drivers were emphasized in this study. Our synthesis broadens a more general knowledge of the drivers of the forest changes. The development of the economy are required more food and agricultural products, forcing the extensive conversion of forests into other land types if biophysical conditions are reasonable for production. These findings may have significant implications for the understanding of how effectively to manage forest resources in Vietnam.

8.2 Limitations and recommendation for further research work

The generalisability of results is subject to certain limitations in this study. For instance, the ELUC approach seems to face a challenge in a data-poor environment if it is applied for tracking rapid changes. A data-poor environment means the existence of data gaps due to reasons such as cloud masks or non-available data. These gaps may cause uncertainty in single-time classifications. Because of these gaps, it is impossible to collect random samples for the accuracy assessment of the single-time classifications. However, since the present research was designed to accurately estimate land cover categories yearly instead of tracking a rapid change within a year, developing a full approach of accuracy assessment for single-time classifications was not solved here. It is an important issue for future studies. A further consideration is that a combination of other data sources should be explored to recognize the full potential of the ELUC for a more accurate assessment of multiple LULC types at a regional or global scale. Also, the ELUC have been tested with two classification algorithms, it should examine the potential of the ELUC with other algorithms.

Another limitation of this study is that the benefit of the individual sensor's characteristics could not been estimated accurately although various data sources have been used, including Landsat OLI, Sentinel SAR GRD, and MSI images. Regarding reference data, the annual maps of the year 1990 to 2014 were validated with the data collected by visual interpretation instead of utilizing ground-truth data. Although great and careful efforts were applied in the collection procedure, errors might not be inevitable due to the limit of super-high-resolution images in Google Earth, especially before the 2000s. A more detailed validation with a more comprehensive set of ground-truth reference data should be considered for additional work. Also, to improve dramatically the accuracy of mapping, deep learning neural networks are expected to be applied for large-scale LULC mapping projects.

Regarding the quantification of the drivers of forest changes, this study accepted disadvantages. First, although the major drivers mentioned in our synthesis of the case studies have been collected and new drivers were added, some drivers were not measured quantitatively, such as forest management policy/tenure, globalization, disasters, and economy/market/price. These drivers may have a profound effect on the transition of forests. Nevertheless, they are not available at a proper spatiotemporal

scale; or some have high multicollinearity, causing difficulties in the precise measurement of the impact of individual drivers; we, therefore, have excluded them from the analysis model. The lack of these drivers might impact our quantitative results. Also, the spatiotemporal difference of some collected drivers may cause uncertainties in the analysis of the drivers. Most of the drivers were derived from the General Statistics Office of Vietnam, which were calibrated by the office; some were extracted from the WorldClim-2 database or the satellite-based information (e.g. SRTM). These data might have relatively low quality and added further uncertainties. However, the current shortage of properly available data does not permit us to overcome this limitation. This would be a fruitful area for further work. Further research should be undertaken to investigate the drivers of other major LULCCs in more details such as urbanization, agriculture, aquaculture, and wetland.

8.3 Significance and contribution of the current study

Developing an innovative approach to achieving timely, accurate, consistent, and large-scale data sets for assessing the rate, patterns, and processes of LULCC is highlighted as beneficial to sustainable development goals. This study certainly adds to our understanding of dynamic LULCC, which is linked to a range of fields including ecosystem and climate change. Herein, a new mapping framework and quantitatively estimating national-scale LULCC trends in the recent three decades is clearly established throughout the paper. In particular, I produced an annual LULC database of 31 maps. To our best knowledge, this is one of the initial comprehensive national databases on three-decade annual LULC monitoring in the developing world. The database has recently published (Sep. 2021) for free use on the website of JAXA as a part of High-Resolution Land Use and Land Cover Map Products. The products have been used by a variety of users for various purposes. Among the purposes the majority are LULCC analyses, drought and flooding studies, climate change and adaptation, water resources, and natural resource management. This fact indicates the utility and importance of the free high resolution time-series land use/land cover maps.

Our approach can be replicated in other regions with similar biophysical characteristics and environmental conditions. It is noted that it is benefited from a rich-data environment, which is in line with current trends in remote sensing development. Hence, it might be promising for estimating LULCC on a broader scale such as regional and global levels. Regarding the quantification of forest changes, previous studies mainly focused on socioeconomic drivers. However, I emphasized the importance of both biophysical and socioeconomic drivers in this study. They may have significant implications for the understanding of how effectively to manage forest resources. Also, they may enhance the accuracy of the land use/land cover change prediction in Vietnam. Other countries, especially the developing world, may be of interest to our findings.

8.4 A proposal for my future work

We intend to estimate the relationship between the time-series dynamic change of LULC and flash flood in Vietnam, using machine learning techniques, remote sensing, and field survey data.

8.4.1 Research context

One of the most devastating natural disasters is flooding, especially flash floods. By the end of the twentieth century, it was globally estimated a loss of approximately 400 billion USD due to flash flooding damage [303]. The most considerable loss occurred in the tropical cyclone climate zones of the Southeast Asia, which accounted for almost a half of the global amount from 1995 to 2015 [304]. Recently, losses associated with flash floods have been more serious owing to an increase in level and intensity of flooding influencing factors such as the higher frequency of extreme rain events and unsustainable land change [297, 105]. It is predicted that the loss may reach one trillion USA by 2050 [305]. Hence, to mitigate damage associated with flash floods is highly paid attention by many governments in the world.

One of the important factors influencing the occurrence of floods is LULC information. Numerous studies of flood susceptibility mapping show the importance of the LULC factor [306, 307]. A much-debated question is whether the time-series change of LULC significantly contributes to a flash flood occurrence. To our best knowledge, no studies have taken into account the long-term annual change of LULC as an important factor in flood susceptibility mapping. It might be the lack of reliable LULC data in a long history. Recently, an annual time-series LULC database (1990 - 2020) has been published on the JAXA web page. The data were comprehensively validated by numerous field survey sites, which confirmed certainly the reliability of the LULC products [20]. It is time to examine the impact of long-term annual LULCC on the occurrence of flash floods.

Detecting susceptible areas of floods are fundamental for flooding mitigation and minimizing the loss of flooding damage. However, it is challenging the accurate estimation of flood susceptible areas due to the sophisticated interaction between a flood occurrence and its influencing factors, which are various. A considerable amount of literature has been worked on flood susceptibility prediction. Numerical models have been traditionally utilized. However, they face considerable challenges due to the requirement of various input parameters, which are frequently inadequate in the developing world. Recently, alternative approaches have developed, including multiple criteria decision-making models, statistical analyses, and machine learning techniques incorporating geospatial data. These approaches, however, have their advantages and disadvantages [308]. To make use of the advantages of the different approaches, an ensemble modeling technique may be essential for flood susceptibility prediction [309]. Also, deep learning with a deeper structure of algorithms may provide analysts with more detailed information to understand the occurrence mechanisms of foods [310].

Currently, although a large number of models have been adopted or flood susceptibility prediction, there is not a principal recommendation on selecting a proper model.

8.4.2 Research questions

The essential aim of proposal is to develop an effective method to improve the accuracy of flood susceptibility prediction in Vietnam, with an integration of machine learning techniques, remote sensing, geospatial, and field survey data. Also, we will examine the relationship between flood susceptibility areas and its influencing factors with a focus on the annual time-series LULCC. To this end, we are going to answer the following questions.

1. What have most effective machine learning methods been used for flood susceptibility mapping?
2. What are the advantages and disadvantages of these methods?
3. Can we develop a new method making use the advantages of the existing methods?
4. Are the time-series information of influencing factors more important than single-time ones in mapping flood susceptibility areas?

8.4.3 Significance of the proposed research

Although it is challenging to completely control floods, the cost of flood damage can be mitigated by successfully predicting flood susceptibility areas and their influential factors. Herein, we intend to develop a new method for improving flood susceptibility mapping with an integration of machine learning techniques, remote sensing and field survey data. Specifically, we will investigate the influential input factors, which help to better understand the mechanism of flood occurrence. This is the first attempt to estimate the influence of long-term time-series input variables with a focus on LULCC. Also, in line of the current remote sensing and computational science development era, we expect to accurately and timely forecast the area that is susceptible to floods in this study. It can be a promising method supporting for different applications such as flood risk controlling and flood hazard mitigation in the world, especially the developing nations which are frequently facing the lack of reliable data for flood forecasting.

Bibliography

[1] Duong Cao Phan, Ta Hoang Trung, Van Thinh Truong, and Kenlo Nishida Nasahara. Ensemble learning updating classifier for accurate land cover assessment in tropical cloudy areas. *Geocarto International*, pages 1–18, feb 2021.

[2] R. S. DeFries, C. B. Field, I. Fung, G. J. Collatz, and L. Bounoua. Combining satellite data and biogeochemical models to estimate global effects of human-induced land cover change on carbon emissions and primary productivity. *Global Biogeochemical Cycles*, 13(3):803–815, 1999.

[3] Belinda Reyers, Patrick J. O'Farrell, Richard M. Cowling, Benis N. Egoh, David C. le Maitre, and Jan H.J. Vlok. Ecosystem services, land-cover change, and stakeholders: Finding a sustainable foothold for a semiarid biodiversity hotspot. *Ecology and Society*, 14(1), 2009.

[4] Roger A. Pielke, Andy Pitman, Dev Niyogi, Rezaul Mahmood, Clive McAlpine, Faisal Hossain, Kees Klein Goldewijk, Udaysankar Nair, Richard Betts, Souleymane Fall, Markus Reichstein, Pavel Kabat, and Nathalie de Noblet. Land use/land cover changes and climate: Modeling analysis and observational evidence, 2011.

[5] Joseph Millard, Charlotte L. Outhwaite, Robyn Kinnersley, Robin Freeman, Richard D. Gregory, Opeyemi Adedoja, Sabrina Gavini, Esther Kioko, Michael Kuhlmann, Jeff Ollerton, Zong Xin Ren, and Tim Newbold. Global effects of land-use intensity on local pollinator biodiversity. *Nature Communications*, 12(1), 2021.

[6] David W. Kicklighter, Jerry M. Melillo, Erwan Monier, Andrei P. Sokolov, and Qianlai Zhuang. Future nitrogen availability and its effect on carbon sequestration in Northern Eurasia. *Nature Communications*, 10(1), 2019.

[7] Tamsin L. Edwards, Sophie Nowicki, Ben Marzeion, Regine Hock, Heiko Goelzer, Hélène Seroussi, Nicolas C. Jourdain, Donald A. Slater, Fiona E. Turner, Christopher J. Smith, Christine M. McKenna, Erika Simon, Ayako Abe-Ouchi, Jonathan M. Gregory, Eric Larour, William H. Lipscomb, Antony J. Payne, Andrew Shepherd, Cécile Agosta, Patrick Alexander, Torsten Albrecht, Brian Anderson, Xylar Asay-Davis, Andy Aschwanden, Alice Barthel, Andrew Bliss, Reinhard Calov, Christopher Chambers, Nicolas Champollion, Youngmin Choi, Richard

Cullather, Joshua Cuzzone, Christophe Dumas, Denis Felikson, Xavier Fettweis, Koji Fujita, Benjamin K. Galton-Fenzi, Rupert Gladstone, Nicholas R. Golledge, Ralf Greve, Tore Hattermann, Matthew J. Hoffman, Angelika Humbert, Matthias Huss, Philippe Huybrechts, Walter Immerzeel, Thomas Kleiner, Philip Kraaijenbrink, Sébastien Le clec'h, Victoria Lee, Gunter R. Leguy, Christopher M. Little, Daniel P. Lowry, Jan Hendrik Malles, Daniel F. Martin, Fabien Maussion, Mathieu Morlighem, James F. O'Neill, Isabel Nias, Frank Pattyn, Tyler Pelle, Stephen F. Price, Aurélien Quiquet, Valentina Radić, Ronja Reese, David R. Rounce, Martin Rückamp, Akiko Sakai, Courtney Shafer, Nicole Jeanne Schlegel, Sarah Shannon, Robin S. Smith, Fiammetta Straneo, Sainan Sun, Lev Tarasov, Luke D. Trusel, Jonas Van Breedam, Roderik van de Wal, Michiel van den Broeke, Ricarda Winkelmann, Harry Zekollari, Chen Zhao, Tong Zhang, and Thomas Zwinger. Projected land ice contributions to twenty-first-century sea level rise. *Nature*, 593(7857):74–82, 2021.

[8] Anita D. Bayer, Mats Lindeskog, Thomas A.M. Pugh, Peter M. Anthoni, Richard Fuchs, and Almut Arneth. Uncertainties in the land-use flux resulting from land-use change reconstructions and gross land transitions. *Earth System Dynamics*, 8(1):91–111, 2017.

[9] R. Fuchs, M. Herold, P. H. Verburg, and J. G.P.W. Clevers. A high-resolution and harmonized model approach for reconstructing and analysing historic land changes in Europe. *Biogeosciences*, 10(3):1543–1559, 2013.

[10] George C. Hurtt, Louise Chini, Ritvik Sahajpal, Steve Frolking, Benjamin L. Bodirsky, Katherine Calvin, Jonathan C. Doelman, Justin Fisk, Shinichiro Fujimori, Kees Klein Goldewijk, Tomoko Hasegawa, Peter Havlik, Andreas Heinimann, Florian Humpenöder, Johan Jungclaus, Jed O. Kaplan, Jennifer Kennedy, Tamás Krisztin, David Lawrence, Peter Lawrence, Lei Ma, Ole Mertz, Julia Pongratz, Alexander Popp, Benjamin Poulter, Keywan Riahi, Elena Shevliakova, Elke Stehfest, Peter Thornton, Francesco N. Tubiello, Detlef P. van Vuuren, and Xin Zhang. Harmonization of global land use change and management for the period 850-2100 (LUH2) for CMIP6. *Geoscientific Model Development*, 13(11):5425–5464, 2020.

[11] Han Liu, Peng Gong, Jie Wang, Nicholas Clinton, Yuqi Bai, and Shunlin Liang. Annual dynamics of global land cover and its long-term changes from 1982 to 2015. *Earth System Science Data*, 12(2):1217–1243, 2020.

[12] N. Tsendbazar, M. Herold, L. Li, A. Tarko, S. de Bruin, D. Masiliunas, M. Lesiv, S. Fritz, M. Buchhorn, B. Smets, R. Van De Kerchove, and M. Duerauer. Towards operational validation of annual global land cover maps. *Remote Sensing of Environment*, 266, 2021.

[13] Karina Winkler, Richard Fuchs, Mark Rounsevell, and Martin Herold. Global land use changes are four times greater than previously estimated. *Nature Communications*, 12(1), 2021.

[14] David Saah, Karis Tenneson, Mir Matin, Kabir Uddin, Peter Cutter, Ate Poortinga, Quyen H. Nguyen, Matthew Patterson, Gary Johnson, Kel Markert, Africa Flores, Eric Anderson, Amanda Weigel, Walter L. Ellenberg, Radhika Bhargava, Aekkapol Aekakkararungroj, Biplov Bhandari, Nishanta Khanal, Ian W. Housman, Peter Potapov, Alexandra Tyukavina, Paul Maus, David Ganz, Nicholas Clinton, and Farrukh Chishtie. Land Cover Mapping in Data Scarce Environments: Challenges and Opportunities. *Frontiers in Environmental Science*, 7, 2019.

[15] Hanh Tran, Thuc Tran, and Matthieu Kervyn. Dynamics of land cover/land use changes in the Mekong Delta, 1973-2011: A Remote sensing analysis of the Tran Van Thoi District, Ca Mau Province, Vietnam. *Remote Sensing*, 7(3):2899–2925, 2015.

[16] Cao Duong Phan, Ta Hoang Trung, Kenlo Nishida Nasahara, and Takeo Tadono. JAXA high-resolution land use/land cover map for Central Vietnam in 2007 and 2017. *Remote Sensing*, 10(9), 2018.

[17] Van Thinh Truong, Thanh Tung Hoang, Cao Duong Phan, Masato Hayashi, Takeo Tadono, and Kenlo Nishida Nasahara. JAXA annual forest cover maps for Vietnam during 2015-2018 Using ALOS-2/PALSAR-2 and auxiliary data. *Remote Sensing*, 11(20), 2019.

[18] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R.G. Townshend. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850–853, 2013.

[19] Han Quang Hanh, Hossein Azadi, Thomas Dogot, Vu Dinh Ton, and Philippe Lebailly. Dynamics of Agrarian Systems and Land Use Change in North Vietnam. *Land Degradation and Development*, 28(3):799–810, 2017.

[20] Duong Cao Phan, Ta Hoang Trung, Van Thinh Truong, Taiga Sasagawa, Thuy Phuong Thi Vu, Dieu Tien Bui, Masato Hayashi, Takeo Tadono, and Kenlo Nishida Nasahara. First comprehensive quantification of annual land use/cover from 1990 to 2020 across mainland Vietnam. *Scientific Reports*, 11(1), dec 2021.

[21] Ziheng Sun, Laura Sandoval, Robert Crystal-Ornelas, S. Mostafa Mousavi, Jinbo Wang, Cindy Lin, Nicoleta Cristea, Daniel Tong, Wendy Hawley Carande, Xiaogang Ma, Yuhan Rao, James A. Bednar, Amanda Tan, Jianwu Wang, Sanjay Purushotham, Thomas E. Gill, Julien Chastang, Daniel Howard, Benjamin Holt, Chandana Gangodagamage, Peisheng Zhao, Pablo Rivas, Zachary Chester, Javier

Orduz, and Aji John. A review of Earth Artificial Intelligence. *Computers & Geosciences*, 159:105034, 2022.

[22] Swapan Talukdar, Pankaj Singha, Susanta Mahato, Shahfahad, Swades Pal, Yuei An Liou, and Atiqur Rahman. Land-use land-cover classification by machine learning classifiers for satellite observations-A review. *Remote Sensing*, 12(7), 2020.

[23] Van Thinh Truong, Cao Duong Phan, Kenlo Nishida Nasahara, and Takeo Tadono. How does land use/land cover map's accuracy depend on number of classification classes? *Scientific Online Letters on the Atmosphere*, 15:28–31, 2019.

[24] A. J. Pitman, F. B. Avila, G. Abramowitz, Y. P. Wang, S. J. Phipps, and N. De Noblet-Ducoudré. Importance of background climate in determining impact of land-cover change on regional climate. *Nature Climate Change*, 1(9):472–475, 2011.

[25] M. Bálint, S. Domisch, C. H.M. Engelhardt, P. Haase, S. Lehrian, J. Sauer, K. Theissinger, S. U. Pauls, and C. Nowak. Cryptic biodiversity loss linked to global climate change. *Nature Climate Change*, 1(6):313–318, 2011.

[26] Phuong Thao Thi Ngo, Tien Dat Pham, Viet Ha Nhu, Thu Trang Le, Dang An Tran, Duong Cao Phan, Pham Viet Hoa, José Lázaro Amaro-Mellado, and Dieu Tien Bui. A novel hybrid quantum-PSO and credal decision tree ensemble for tropical cyclone induced flash flood susceptibility mapping with geospatial data. *Journal of Hydrology*, 596, 2021.

[27] Viet Ha Nhu, Phuong Thao Thi Ngo, Tien Dat Pham, Jie Dou, Xuan Song, Nhat Duc Hoang, Dang An Tran, Duong Phan Cao, Ibrahim Berkan Aydilek, Mahdis Amiri, Romulus Costache, Pham Viet Hoa, and Dieu Tien Bui. A new hybrid firefly-pso optimized random subspace tree intelligence for torrential rainfall-induced flash flood susceptible mapping. *Remote Sensing*, 12(17), 2020.

[28] IPCC. Climate Change and Land: an IPCC special report. In *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*, pages 1–864. United Nations, 2019.

[29] USGS. [IGBP] USGS EROS Archive - Land Cover Products - Global Land Cover Characterization (GLCC), 2018.

[30] Helmut Herrmann and Herbert Bucksch. Global Land Cover Facility, 2014.

[31] A Hartley, J-F. Pekel, M Ledwith, J-L. Champeaux, E De Badts, and S A Bartalev. Global land cover 2000 - Europe, 2002.

[32] ESA. GlobCover - Global Land Cover Service, 2011.

[33] ESA. ESA Climate Change Initiative, 2017.

[34] Mark A. Friedl and D. Sulla-Menashe. MCD12Q2 MODIS/Terra+ Aqua 589 Land Cover Dynamics Yearly L3 Global 500m SIN Grid V006. NASA 590 EOSDIS LP DAAC, 2019.

[35] Toshiyuki Kobayashi, Ryutaro Tateishi, Bayan Alsaaideh, Ram C. Sharma, Takuma Wakaizumi, Daichi Miyamoto, Xiulian Bai, Bui D. Long, Gegentana Gegentana, Aikebaier Maitiniyazi, Destika Cahyana, Alifu Haireti, Yohei Morifuji, Gulijianati Abake, Rendy Pratama, Naijia Zhang, Zilaitigu Alifu, Tomohiro Shirahata, Lan Mi, Kotaro Iizuka, Aimaiti Yusupujiang, Fedri R. Rinawan, Richa Bhattarai, and Dong X. Phong. Production of Global Land Cover Data – GLCNMO2013. *Journal of Geography and Geology*, 9(3):1, 2017.

[36] Jamal Jokar Arsanjani, Amin Tayyebi, and Eric Vaz. GlobeLand30 as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries. *Habitat International*, 55:25–31, 2016.

[37] Xiao Zhang, Liangyun Liu, Xidong Chen, Yuan Gao, Shuai Xie, and Jun Mi. GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth System Science Data Discussions*, pages 1–31, 2020.

[38] Marcel Buchhorn, Bruno Smets, Luc Bertels, Myroslava Lesiv, Nandin-Erdene Tsendbazar, D. Masiliunas, L. Linlin, Martin Herold, and S. Fritz. Copernicus Global Land Service: Land Cover 100m: Collection 3: epoch 2015: Globe (Version V3.0.1). *Zenodo*, 2020.

[39] Krishna Karra, Caitlin Kontgis, Zoe Statman-Weil, Joseph C. Mazzariello, Mark Mathis, and Steven P. Brumby. Global land use / land cover with Sentinel 2 and deep learning. In *IGARSS 2021-2021 IEEE International Geoscience and Remote Sensing Symposium*, pages 4704–4707. IEEE, 2021.

[40] Eric F. Lambin and Daniele Ehrlich. Land-cover changes in Sub-Saharan Africa (1982-1991): Application of a change index based on remotely sensed surface temperature and vegetation indices at a continental scale. *Remote Sensing of Environment*, 61(2):181–200, 1997.

[41] L. Durieux, J. Kropáček, G. D. de Grandi, and F. Achard. Object-oriented and textural image classification of the Siberia GBFM radar mosaic combined with MERIS imagery for continental scale land cover mapping. *International Journal of Remote Sensing*, 28(18):4175–4182, 2007.

[42] Christelle Vancutsem, Eduardo Marinho, François Kayitakire, Linda See, and Steffen Fritz. Harmonizing and combining existing land cover/land use datasets for cropland area monitoring at the African continental scale. *Remote Sensing*, 5(1):19–41, 2013.

[43] Jovanka Nikolic, Shiyuan Zhong, Lisi Pei, Xindi Bian, Warren E. Heilman, and Joseph J. Charney. Sensitivity of low-level jets to land-use and land- cover change over the continental U.S. *Atmosphere*, 10(4), 2019.

[44] Marco Calderón-Loor, Michalis Hadjikakou, and Brett A. Bryan. High-resolution wall-to-wall land-cover mapping and land change assessment for Australia from 1985 to 2015. *Remote Sensing of Environment*, 252, 2021.

[45] Zander S. Venter and Markus A.K. Sydenham. Continental-scale land cover mapping at 10 m resolution over europe (Elc10). *Remote Sensing*, 13(12), 2021.

[46] Don C.I. Okpala. State of national land survey and large-scale mapping. *Land Use Policy*, 13(4):317–323, 1996.

[47] S. García-Gigorro, F. González-Alonso, S. Merino-de Miguel, A. Roldán-Zamarrón, and J. M. Cuevas. MERIS-FR potential for land use-land coer mapping in Spain. *International Journal of Remote Sensing*, 28(6):1405–1412, 2007.

[48] J. G.P.W. Clevers, M. E. Schaepman, C. A. Mücher, A. J.W. de Wit, R. Zurita-Milla, and H. M. Bartholomeus. Using MERIS on Envisat for land cover mapping in the Netherlands. *International Journal of Remote Sensing*, 28(3-4):637–652, 2007.

[49] J. Dash, A. Mathur, G. M. Foody, P. J. Curran, J. W. Chipman, and T. M. Lillesand. Land cover classification using multi-temporal MERIS vegetation indices. *International Journal of Remote Sensing*, 28(6):1137–1159, 2007.

[50] H. Carrão, A. Araújo, P. Gonçalves, and M. Caetano. Multitemporal MERIS images for land-cover mapping at a national scale: A case study of Portugal. *International Journal of Remote Sensing*, 31(8):2063–2082, 2010.

[51] James Wickham, Stephen V. Stehman, Daniel G. Sorenson, Leila Gass, and Jon A. Dewitz. Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States. *Remote Sensing of Environment*, 257, 2021.

[52] Thanh Tung Hoang, Van Thinh Truong, Masato Hayashi, Takeo Tadono, and Kenlo Nishida Nasahara. New JAXA high-resolution land use/land cover map for Vietnam aiming for natural forest and plantation forest monitoring. *Remote Sensing*, 12(17), 2020.

[53] Sowmya D., P. Deepa, and Venugopal K. Remote Sensing Satellite Image Processing Techniques for Image Classification: A Comprehensive Survey. *International Journal of Computer Applications*, 161(11):24–37, 2017.

[54] J. A. Lindesay, M. O. Andreae, J. G. Goldammer, G. Harris, H. J. Annegarn, M. Garstang, R. J. Scholes, and B. W. Van Wilgen. International geosphere-biosphere programme/international global atmospheric chemistry SAFARI-92 field experiment: Background and overview. *Journal of Geophysical Research Atmospheres*, 101(19):23521–23530, 1996.

[55] V. V. Klemas, J. E. Dobson, R. L. Ferguson, and K. D. Haddad. A coastal land cover classification system for the NOAA Coastwatch Change Analysis Project. *Journal of Coastal Research*, 9(3):862–872, 1993.

[56] Darius Phiri and Justin Morgenroth. Developments in Landsat land cover classification methods: A review, 2017.

[57] Sakshi Dhingra and Dharminder Kumar. A review of remotely sensed satellite image classification. *International Journal of Electrical and Computer Engineering*, 9(3):1720–1731, 2019.

[58] Lazhar Khelifi and Max Mignotte. Deep Learning for Change Detection in Remote Sensing Images: Comprehensive Review and Meta-Analysis. *IEEE Access*, 8:126385–126400, 2020.

[59] Chiman Kwan, Bulent Ayhan, Bence Budavari, Yan Lu, Daniel Perez, Jiang Li, Sergio Bernabe, and Antonio Plaza. Deep learning for land cover classification using only a few bands. *Remote Sensing*, 12(12), 2020.

[60] Giles M. Foody. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1):185–201, 2002.

[61] G. M. Foody. Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *International Journal of Remote Sensing*, 26(6):1217–1228, 2005.

[62] Neil Flood. Seasonal composite landsat TM/ETM+ Images using the medoid (a multi-dimensional median). *Remote Sensing*, 5(12):6481–6500, 2013.

[63] Pontus Olofsson, Giles M. Foody, Martin Herold, Stephen V. Stehman, Curtis E. Woodcock, and Michael A. Wulder. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148:42–57, 2014.

[64] D. Lu, P. Mausel, E. Brondízio, and E. Moran. Change detection techniques. *International Journal of Remote Sensing*, 25(12):2365–2401, 2004.

[65] Jean Philippe Puyravaud. Standardizing the calculation of the annual rate of deforestation. *Forest Ecology and Management*, 177(1-3):593–596, 2003.

[66] Russell Congalton, Jianyu Gu, Kamini Yadav, Prasad Thenkabail, and Mutlu Ozdogan. Global Land Cover Mapping: A Review and Uncertainty Analysis. *Remote Sensing*, 6(12):12070–12093, dec 2014.

[67] Martine Rutten, Michiel Van Dijk, Wilbert Van Rooij, and Henk Hilderink. Land use dynamics, climate change, and food security in Vietnam: A global-to-local modeling approach. *World Development*, 59:29–46, 2014.

[68] T. Tadono, H. Ishida, F. Oda, S. Naito, K. Minakawa, and H. Iwamoto. Precise Global DEM Generation by ALOS PRISM. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-4:71–76, 2014.

[69] Tung Thanh Hoang. *Analysis of Natural Forest and Plantation Forest Changes in Vietnam Using Remote Sensing Data*. Thesis, University of Tsukuba, 2021.

[70] Sunsanee Arunyawat and Rajendra P. Shrestha. Assessing land use change and its impact on ecosystem services in northern Thailand. *Sustainability (Switzerland)*, 8(8), 2016.

[71] Lars Koschke, Christine Fürst, Susanne Frank, and Franz Makeschin. A multi-criteria approach for an integrated land-cover-based assessment of ecosystem services provision to support landscape planning. *Ecological Indicators*, 21:54–66, 2012.

[72] Terefe Tolessa, Feyera Senbeta, and Moges Kidane. The impact of land use/land cover change on ecosystem services in the central highlands of Ethiopia. *Ecosystem Services*, 23:47–54, 2017.

[73] Shannon M. Sterling, Agnès Ducharne, and Jan Polcher. The impact of global land-cover change on the terrestrial water cycle. *Nature Climate Change*, 3(4):385–390, 2013.

[74] Eugenia Kalnay and Ming Cai. Impact of urbanization and land-use change on climate. *Nature*, 423(6939):528–531, 2003.

[75] Sérgio Niquisse, Pedro Cabral, Ângela Rodrigues, and Gabriela Augusto. Ecosystem services and biodiversity trends in Mozambique as a consequence of land cover change. *International Journal of Biodiversity Science, Ecosystem Services and Management*, 13(1):297–311, 2017.

[76] Li Jiang, Xiangzheng Deng, and Karen C. Seto. The impact of urban expansion on agricultural land use intensity in China. *Land Use Policy*, 35:33–39, 2013.

[77] Jacob H. Lafontaine, Lauren E. Hay, Roland J. Viger, R. Steve Regan, and Steven L. Markstrom. Effects of Climate and Land Cover on Hydrology in the Southeastern U.S.: Potential Impacts on Watershed Planning. *Journal of the American Water Resources Association*, 51(5):1235–1261, 2015.

[78] Patrick Meyfroidt. Environmental Cognitions, Land Change and Social-Ecological Feedbacks: Local Case Studies of Forest Transition in Vietnam. *Human Ecology*, 41(3):367–392, 2013.

[79] Nicolas Titeux, Klaus Henle, Jean Baptiste Mihoub, Adrián Regos, Ilse R. Geijzendorffer, Wolfgang Cramer, Peter H. Verburg, and Lluís Brotons. Biodiversity scenarios neglect future land-use changes. *Global change biology*, 22(7):2505–2515, 2016.

[80] Eric F. Lambin and Patrick Meyfroidt. Land use transitions: Socio-ecological feedback versus socio-economic change. *Land Use Policy*, 27(2):108–118, 2010.

[81] Valerio Avitabile, Michael Schultz, Nadine Herold, Sytze de Bruin, Arun Kumar Pratihast, Cuong Pham Manh, Hien Vu Quang, and Martin Herold. Carbon emissions from land cover change in Central Vietnam. *Carbon Management*, 7(5-6):333–346, 2016.

[82] Leonardo Disperati and Salvatore Gonario Pasquale Virdis. Assessment of land-use and land-cover changes from 1965 to 2014 in Tam Giang-Cau Hai Lagoon, central Vietnam. *Applied Geography*, 58:48–64, 2015.

[83] P.T. Dung and S. Sharma. Chapter 2 - responding to climate change in the agriculture and rural development sector in vietnam. In Mai Van Thanh, Tran Duc Vien, Stephen J. Leisz, and Ganesh P. Shivakoti, editors, *Redefining Diversity & Dynamics of Natural Resources Management in Asia, Volume 2*, pages 13–25. Elsevier, 2017.

[84] Fanny Langerwisch, Tomas Vaclavik, Werner Von Bloh, Tobias Vetter, and Kirsten Thonicke. Combined effects of climate and land-use change on the provision of ecosystem services in rice agro-ecosystems. *Environmental Research Letters*, 13(1), 2018.

[85] Han Soo Lee, Andhang Rakhmat Trihamdani, Tetsu Kubota, Satoru Iizuka, and Tran Thi Thu Phuong. Impacts of land use changes from the Hanoi Master Plan 2030 on urban heat islands: Part 2. Influence of global warming. *Sustainable Cities and Society*, 31:95–108, 2017.

[86] Pengfei Fan, Changming Liu, Wenshou Luo, and Xuelong Jiang. Can a group elicit duets from its neighbours? A field study on the black-crested gibbon (Nomascus concolor jingdongensis) in Central Yunnan, China. *Folia Primatologica*, 78(3):186–195, 2007.

[87] R. A. Houghton, J. I. House, J. Pongratz, G. R. Van Der Werf, R. S. Defries, M. C. Hansen, C. Le Quéré, and N. Ramankutty. Carbon emissions from land use and land-cover change. *Biogeosciences*, 9(12):5125–5142, 2012.

[88] Nguyen Kim Loi. Assessing the Impacts of Land use / Land cover Changes and Practices on Water Discharge and Sedimentation using SWAT : Case study in Dong Nai watershed – Vietnam. *International Symposium on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences 2010*, 2010.

[89] D. Serpa, J. P. Nunes, J. Santos, E. Sampaio, R. Jacinto, S. Veiga, J. C. Lima, M. Moreira, J. Corte-Real, J. J. Keizer, and N. Abrantes. Impacts of climate and land use changes on the hydrological and erosion processes of two contrasting Mediterranean catchments. *Science of the Total Environment*, 538:64–77, 2015.

[90] Philippe Mayaux, Hugh Eva, Javier Gallego, Alan H. Strahler, Martin Herold, Shefali Agrawal, Sergey Naumov, Evaristo Eduardo De Miranda, Carlos M. Di Bella, Callan Ordoyne, Yuri Kopin, and P. S. Roy. Validation of the global land cover 2000 map. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7):1728–1737, 2006.

[91] Thomas C.G. Bosch, René Augustin, Klaus Gellner, Konstantin Khalturin, and Jan U. Lohmann. In vivo electroporation for genetic manipulations of whole Hydra polyps. Technical Report 4-5, National Center for Biotechnology Information Search database, 2002.

[92] J. Scepan and J. E. Estes. Thematic validation of global land cover data sets - Procedures and interpretation methods. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 3, pages 1119–1121, 2001.

[93] Mark A. Friedl, Damien Sulla-Menashe, Bin Tan, Annemarie Schneider, Navin Ramankutty, Adam Sibley, and Xiaoman Huang. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114(1):168–182, 2010.

[94] Peng Gong, Jie Wang, L. Yu, Yongchao Zhao, Yuanyuan Zhao, L. Liang, Zhenguo Niu, Xiaomeng Huang, Haohuan Fu, Shuang Liu, Congcong Li, Xueyan Li, Wei Fu, Caixia Liu, Yue Xu, Xiaoyi Wang, Q. Cheng, Luanyun Hu, Wenbo Yao, Han Zhang, Peng Zhu, Ziying Zhao, Haiying Zhang, Yaomin Zheng, Luyan Ji, Yawen Zhang, Han Chen, A. Yan, Jianhong Guo, Liang Yu, Lei Wang, Xiaojun Liu, Tingting Shi, Menghua Zhu, Yanlei Chen, Guangwen Yang, Ping Tang, Bing Xu, Chandra Giri, Nicholas Clinton, Zhiliang Zhu, Jin Chen, and Jun Chen. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 34(7):2607–2654, 2013.

[95] C. J. Stubenrauch, W. B. Rossow, S. Kinne, S. Ackerman, G. Cesana, H. Chepfer, L. Di Girolamo, B. Getzewich, A. Guignard, A. Heidinger, B. C. Maddux, W. P. Menzel, P. Minnis, C. Pearl, S. Platnick, C. Poulsen, J. Riedi, S. Sun-Mack, A. Walther, D. Winker, S. Zeng, and G. Zhao. Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX radiation panel. *Bulletin of the American Meteorological Society*, 94(7):1031–1049, 2013.

[96] Terry Arvidson, Samuel Goward, John Gasch, and Darrel Williams. Landsat-7 long-term acquisition plan: Development and validation, 2006.

[97] Cornelius Senf, Pedro J. Leitão, Dirk Pflugmacher, Sebastian van der Linden, and Patrick Hostert. Mapping land cover in complex Mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. *Remote Sensing of Environment*, 156:527–536, 2015.

[98] Michael A. Wulder, Jeffrey G. Masek, Warren B. Cohen, Thomas R. Loveland, and Curtis E. Woodcock. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, 122:2–10, 2012.

[99] Nathan Torbick, Diya Chowdhury, William Salas, and Jiaguo Qi. Monitoring rice agriculture across myanmar using time series Sentinel-1 assisted by Landsat-8 and PALSAR-2. *Remote Sensing*, 9(2), 2017.

[100] Jinwei Dong, Xiangming Xiao, Bangqian Chen, Nathan Torbick, Cui Jin, Geli Zhang, and Chandrashekhar Biradar. Mapping deciduous rubber plantations through integration of PALSAR and multi-temporal Landsat imagery. *Remote Sensing of Environment*, 134:392–402, 2013.

[101] Tao Zhou, Zhaofu Li, and Jianjun Pan. Multi-feature classification of multi-sensor satellite imagery based on dual-polarimetric sentinel-1A, landsat-8 OLI, and hyperion images for urban land-cover classification. *Sensors (Switzerland)*, 18(2), 2018.

[102] Tien Dat Pham, Kunihiko Yoshino, Nga Nhu Le, and Dieu Tien Bui. Estimating aboveground biomass of a mangrove plantation on the Northern coast of Vietnam using machine learning techniques with an integration of ALOS-2 PALSAR-2 and Sentinel-2A data. *International Journal of Remote Sensing*, 39(22):7761–7788, 2018.

[103] Johannes Reiche, Richard Lucas, Anthea L. Mitchell, Jan Verbesselt, Dirk H. Hoekman, Jörg Haarpaintner, Josef M. Kellndorfer, Ake Rosenqvist, Eric A. Lehmann, Curtis E. Woodcock, Frank Martin Seifert, and Martin Herold. Combining satellite data for better tropical forest monitoring, 2016.

[104] Jose Don T. De Alban, Grant M. Connette, Patrick Oswald, and Edward L. Webb. Combined Landsat and L-band SAR data improves land cover classification and change detection in dynamic tropical landscapes. *Remote Sensing*, 10(2), 2018.

[105] Intergovernmental Panel on Climate Change. *Climate Change 2014 Mitigation of Climate Change*. The Intergovernmental Panel on Climate Change (IPCC), 2014.

[106] Martin Rudbeck Jepsen and Gregor Levin. Semantically based reclassification of Danish land-use and land-cover information. *International Journal of Geographical Information Science*, 27(12):2375–2390, 2013.

[107] Peter H. Verburg, Neville Crossman, Erle C. Ellis, Andreas Heinimann, Patrick Hostert, Ole Mertz, Harini Nagendra, Thomas Sikor, Karl Heinz Erb, Nancy Golubiewski, Ricardo Grau, Morgan Grove, Souleymane Konaté, Patrick Meyfroidt, Dawn C. Parker, Rinku Roy Chowdhury, Hideaki Shibata, Allison Thomson, and Lin Zhen. Land system science and sustainable development of the earth system: A global land project perspective. *Anthropocene*, 12:29–41, 2015.

[108] Thanh Tung Hoang, Kenlo Nishida Nasahara, and Jin Katagi. Analysis of Land Cover Changes in Northern Vietnam Using High Resolution Remote Sensing Data. In *Advances and Applications in Geospatial Technology and Earth Resources*, pages 134–151. Springer, Cham., 2018.

[109] Shahriar S. Heydari and Giorgos Mountrakis. Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. *Remote Sensing of Environment*, 204:648–658, 2018.

[110] Edward J. Knight and Geir Kvaran. Landsat-8 operational land imager design, characterization and performance. *Remote Sensing*, 6(11):10286–10305, 2014.

[111] Zhe Zhu and Curtis E. Woodcock. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118:83–94, 2012.

[112] Zhe Zhu, Shixiong Wang, and Curtis E. Woodcock. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images. *Remote Sensing of Environment*, 159:269–277, 2015.

[113] Eija Parmes, Yrjö Rauste, Matthieu Molinier, Kaj Andersson, and Lauri Seitsonen. Automatic cloud and shadow detection in optical satellite imagery without using thermal bands-application to Suomi NPP VIIRS images over Fennoscandia. *Remote Sensing*, 9(8), 2017.

[114] Eric Vermote, Chris Justice, and Ivan Csiszar. Early evaluation of the VIIRS calibration, cloud mask and surface reflectance Earth data records. *Remote Sensing of Environment*, 148:134–145, 2014.

[115] Lihang Zhou, Murty Divakarla, and Xingpin Liu. An overview of the joint polar satellite system (JPSS) science data product calibration and validation. *Remote Sensing*, 8(2), 2016.

[116] M. Piper and T. Bahr. A rapid cloud mask algorithm for suomi npp VIIRS imagery EDRS. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, volume 40, pages 237–242, 2015.

[117] Abd Rahman As-syakur, I. Wayan Sandi Adnyana, I. Wayan Arthana, and I. Wayan Nuarsa. Enhanced built-UP and bareness index (EBBI) for mapping built-UP and bare land in an urban area. *Remote Sensing*, 4(10):2957–2970, 2012.

[118] Y. Zha, J. Gao, and S. Ni. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3):583–594, 2003.

[119] Makoto; Kawamura, Sanath; Jayamanna, and Yuji Tsujiko. Relation between social and environmental conditions in Colombo Sri Lanka and the urban index

estimated by satellite remote sensing data. *International Archives of Photogrammetry and Remote Sensing*, XXXI(B7):321–326, 1996.

[120] Hongmei Zhao and Xiaoling Chen. Use of normalized difference bareness index in quickly mapping bare areas from TM/ETM+. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 3, pages 1666–1668, 2005.

[121] Compton J. Tucker. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2):127–150, 1979.

[122] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1-2):195–213, 2002.

[123] A. R. Huete. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3):295–309, 1988.

[124] Bo Cai Gao. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3):257–266, 1996.

[125] Weili Kou, Xiangming Xiao, Jinwei Dong, Shu Gan, Deli Zhai, Geli Zhang, Yuanwei Qin, and Li Li. Mapping deciduous rubber plantation areas and stand ages with PALSAR and landsat images. *Remote Sensing*, 7(1):1048–1073, 2015.

[126] Jinwei Dong, Xiangming Xiao, Sage Sheldon, Chandrashekhar Biradar, and Guishui Xie. Mapping tropical forests and rubber plantations in complex landscapes by integrating PALSAR and MODIS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 74:20–33, 2012.

[127] J. S. Lee, I. Jurkevich, P. Dewaele, P. Wambacq, and A. Oosterlinck. Speckle filtering of synthetic aperture radar images: a review. *Remote Sensing Reviews*, 8(4):313–340, 1994.

[128] Masanobu Shimada, Takuya Itoh, Takeshi Motooka, Manabu Watanabe, Tomohiro Shiraishi, Rajesh Thapa, and Richard Lucas. New global forest/non-forest maps from ALOS PALSAR data (2007-2010). *Remote Sensing of Environment*, 155:13–31, 2014.

[129] Shutaro Hashimoto, Takeo Tadono, Masahiko Onosato, Masahiro Hori, and Takashi Moriyama. Probabilistic land cover classification approach toward knowledge-based satellite data interpretations. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1513–1516, 2012.

[130] Shutaro Hashimoto, Takeo Tadono, Masahiko Onosato, and Masahiro Hori. Land use and land cover inference in large areas using multi-temporal optical satellite images. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3333–3336, 2013.

[131] Jin Katagi, Nishida Kenlo Nasahara, Kenichiro Kobayashi, Masanori Dotsu, and Takeo Tadono. Reduction of Misclassification Caused by Mountain Shadow in a High Resolution Land Use and Land Cover Map Using Multi-temporal Optical Images. *Journal of The Remote Sensing Society of Japan*, 38(1):30–34, 2018.

[132] Mitsunori Ishihara and Takeo Tadono. Land cover changes induced by the great east Japan earthquake in 2011. *Scientific Reports*, 7, 2017.

[133] Stacy L. Ozesmi and Marvin E. Bauer. Satellite remote sensing of wetlands. *Wetlands Ecology and Management*, 10(5):381–402, 2002.

[134] H. G. Lewis and M. Brown. A generalized confusion matrix for assessing area estimates from remotely sensed data. *International Journal of Remote Sensing*, 22(16):3223–3235, 2001.

[135] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

[136] Romie Jhonnerie, Vincentius P. Siregar, Bisman Nababan, Lilik Budi Prasetyo, and Sam Wouthuyzen. Random Forest Classification for Mangrove Land Cover Mapping Using Landsat 5 TM and Alos Palsar Imageries. *Procedia Environmental Sciences*, 24:215–221, 2015.

[137] Arief Wijaya and Richard Gloaguen. Fusion of ALOS Palsar and Landsat ETM data for land cover classification and biomass modeling using non-linear methods. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 3, 2009.

[138] Gaia Vaglio Laurin, Veraldo Liesenberg, Qi Chen, Leila Guerriero, Fabio Del Frate, Antonio Bartolini, David Coomes, Beccy Wilebore, Jeremy Lindsell, and Riccardo Valentini. Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa. *International Journal of Applied Earth Observation and Geoinformation*, 21(1):7–16, 2012.

[139] Bui Thi Dung, Henry Madsen, and Dang Tat The. Distribution of freshwater snails in family-based VAC ponds and associated waterbodies with special reference to intermediate hosts of fish-borne zoonotic trematodes in Nam Dinh Province, Vietnam. *Acta Tropica*, 116(1):15–23, 2010.

[140] Son Thi Thanh Dang and Anders Dalsgaard. Escherichia coli contamination of fish raised in integrated pig-fish aquaculture systems in Vietnam. *Journal of Food Protection*, 75(7):1317–1319, 2012.

[141] Asli Ozdarici-Ok, Ali Ozgun Ok, and Konrad Schindler. Mapping of agricultural crops from single high-resolution multispectral images-data-driven smoothing vs. parcel-based smoothing. *Remote Sensing*, 7(5):5611–5638, 2015.

[142] Jonas Schmedtmann and Manuel L. Campagnolo. Reliable crop identification with satellite imagery in the context of Common Agriculture Policy subsidy control. *Remote Sensing*, 7(7):9325–9346, 2015.

[143] Yasser Maghsoudi, Michael J. Collins, and Donald Leckie. Speckle reduction for the forest mapping analysis of multi-temporal Radarsat-1 images. *International Journal of Remote Sensing*, 33(5):1349–1359, 2012.

[144] Nguyen Quang Phuc, A. C.M.van Westen, and Annelies Zoomers. Agricultural land for urban development: The process of land conversion in Central Vietnam. *Habitat International*, 41:1–7, 2014.

[145] Alan de Brauw. Seasonal migration and agricultural production in Vietnam. *Journal of Development Studies*, 46(1):114–139, 2010.

[146] Duong Dang Khoi and Yuji Murayama. Forecasting areas vulnerable to forest conversion in the tam Dao National Park region, Vietnam. *Remote Sensing*, 2(5):1249–1272, 2010.

[147] Jukka Miettinen, Chenghua Shi, and Soo Chin Liew. Deforestation rates in insular Southeast Asia between 2000 and 2010. *Global Change Biology*, 17(7):2261–2270, 2011.

[148] EK Sadanandan Nambiar, Christopher E. Harwood, and Nguyen Duc Kien. Acacia plantations in Vietnam: research and knowledge application to secure a sustainable future, 2015.

[149] Truong Dao Minh, Masayuki Yanagisawa, and Yasuyuki Kono. Forest transition in Vietnam: A case study of Northern mountain region. *Forest Policy and Economics*, 76:72–80, 2017.

[150] Patrick Meyfroidt and Eric F. Lambin. Forest transition in Vietnam and its environmental impacts. *Global Change Biology*, 14(6):1319–1336, 2008.

[151] Kenneth Macdicken, Greg Reams, and Joberto De Freitas. Special Issue: Global Forest Resources Assessment 2015. *Forest Ecology and Management*, page 244, 2015.

[152] John A. Church and Jonathan M. Gregory. Sea level change. In J. Kirk Cochran, Henry J. Bokuniewicz, and Patricia L. Yager, editors, *Encyclopedia of Ocean Sciences (Third Edition)*, pages 493–499. Academic Press, Oxford, third edition edition, 2019.

[153] Zahra Rahneshan, Fatemeh Nasibi, and Ali Ahmadi Moghadam. Effects of salinity stress on some growth, physiological, biochemical parameters and nutrients in two pistachio (Pistacia vera L.) rootstocks. *Journal of Plant Interactions*, 13(1):73–82, 2018.

[154] Amira M.S. Abdul Qados. Effect of salt stress on plant growth and metabolism of bean plant Vicia faba (L.). *Journal of the Saudi Society of Agricultural Sciences*, 10(1):7–15, 2011.

[155] Collin Homer, Jon Dewitz, Limin Yang, Suming Jin, Patrick Danielson, George Xian, John Coulston, Nathaniel Herold, James Wickham, and Kevin Megown. Completion of the 2011 national land cover database for the conterminous United States – Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, 81(5):345–354, 2015.

[156] Curtis E. Woodcock, Richard Allen, Martha Anderson, Alan Belward, Robert Bindschadler, Warren Cohen, Feng Gao, Samuel N. Goward, Dennis Helder, Eileen Helmer, Rama Nemani, Lazaros Oreopoulos, John Schott, Prasad S. Thenkabail, Eric F. Vermote, James Vogelmann, Michael A. Wulder, and Randolph Wynne. Free access to landsat imagery. *Science*, 320(5879):1011, 2008.

[157] S. J. Baillarin, A. Meygret, C. Dechoz, B. Petrucci, S. Lacherade, T. Tremas, C. Isola, P. Martimort, and F. Spoto. Sentinel-2 level 1 products and image processing performances. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 7003–7006, 2012.

[158] Zbyněk Malenovský, Helmut Rott, Josef Cihlar, Michael E. Schaepman, Glenda García-Santos, Richard Fernandes, and Michael Berger. Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sensing of Environment*, 120:91–101, 2012.

[159] Ramon Torres, Paul Snoeij, Dirk Geudtner, David Bibby, Malcolm Davidson, Evert Attema, Pierre Potin, Bj Örn Rommen, Nicolas Floury, Mike Brown, Ignacio Navas Traver, Patrick Deghaye, Berthyl Duesmann, Betlem Rosich, Nuno Miranda, Claudio Bruno, Michelangelo L'Abbate, Renato Croci, Andrea Pietropaolo, Markus Huchler, and Friedhelm Rostan. GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120:9–24, 2012.

[160] Jixian Zhang. Multi-source remote sensing data fusion: Status and trends. *International Journal of Image and Data Fusion*, 1(1):5–24, 2010.

[161] Luigi Boschetti, David P. Roy, Christopher O. Justice, and Michael L. Humber. MODIS-Landsat fusion for large area 30m burned area mapping. *Remote Sensing of Environment*, 161:27–42, 2015.

[162] Lev V. Utkin, Andrei V. Konstantinov, Viacheslav S. Chukanov, Mikhail V. Kots, Mikhail A. Ryabinin, and Anna A. Meldo. A weighted random survival forest. *Knowledge-Based Systems*, 177:136–144, 2019.

[163] Yunfeng Hu, Yu Dong, and Batunacun. An automatic approach for land-change detection and land updates based on integrated NDVI timing analysis and the

CVAPS method with GEE support. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:347–359, 2018.

[164] Zhe Zhu and Curtis E. Woodcock. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment*, 144:152–171, 2014.

[165] Mateo Gašparović, Mladen Zrinjski, and Marina Gudelj. Automatic cost-effective method for land cover classification (ALCC). *Computers, Environment and Urban Systems*, 76:1–10, 2019.

[166] Pieter Kempeneers, Fernando Sedano, Lucia Seebach, Peter Strobl, and Jesús San-Miguel-Ayanz. Data fusion of different spatial resolution remote sensing images applied to forest-type mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 49(12 PART 2):4977–4986, 2011.

[167] ESA. GOCE Mission Requirements Document. *Earth*, 2000.

[168] Magdalena Main-Knorn, Bringfried Pflug, Jerome Louis, Vincent Debaecker, Uwe Müller-Wilm, and Ferran Gascon. Sen2Cor for Sentinel-2. In *Image and Signal Processing for Remote Sensing XXIII*, page 3, 2017.

[169] Peng Gong, Han Liu, Meinan Zhang, Congcong Li, Jie Wang, Huabing Huang, Nicholas Clinton, Luyan Ji, Wenyu Li, Yuqi Bai, Bin Chen, Bing Xu, Zhiliang Zhu, Cui Yuan, Hoi Ping Suen, Jing Guo, Nan Xu, Weijia Li, Yuanyuan Zhao, Jun Yang, Chaoqing Yu, Xi Wang, Haohuan Fu, Le Yu, Iryna Dronova, Fengming Hui, Xiao Cheng, Xueli Shi, Fengjin Xiao, Qiufeng Liu, and Lianchun Song. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Science Bulletin*, 64(6):370–373, 2019.

[170] Collin Homer, Chengquan Huang, Limin Yang, Bruce Wylie, and Michael Coan. Development of a 2001 National Land-Cover Database for the United States. *Photogrammetric Engineering and Remote Sensing*, 70(7):829–840, 2004.

[171] Joyce A. Fry, George Xian, Suming Jin, Jon A. Dewitz, Collin G. Homer, Limin Yang, Christopher A. Barnes, Nathaniel D. Herold, and James D. Wickham. Completion of the 2006 national land cover database for the conterminous united states. *Photogrammetric Engineering and Remote Sensing*, 77(9):858–864, 2011.

[172] Limin Yang, Suming Jin, Patrick Danielson, Collin Homer, Leila Gass, Stacie M. Bender, Adam Case, Catherine Costello, Jon Dewitz, Joyce Fry, Michelle Funk, Brian Granneman, Greg C. Liknes, Matthew Rigge, and George Xian. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:108–123, 2018.

[173] Jun Chen, Jin Chen, Anping Liao, Xin Cao, Lijun Chen, Xuehong Chen, Chaoying He, Gang Han, Shu Peng, Miao Lu, Weiwei Zhang, Xiaohua Tong, and Jon Mills. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:7–27, 2015.

[174] Matthew C. Hansen and Thomas R. Loveland. A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, 122:66–74, 2012.

[175] John A. Matthews. *Encyclopedia of Environmental Change*. SAGE Publications, Ltd., 2014.

[176] Neha Joshi, Matthias Baumann, Andrea Ehammer, Rasmus Fensholt, Kenneth Grogan, Patrick Hostert, Martin Rudbeck Jepsen, Tobias Kuemmerle, Patrick Meyfroidt, Edward T.A. Mitchard, Johannes Reiche, Casey M. Ryan, and Björn Waske. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, 8(1), 2016.

[177] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S. Lew. Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48, 2016.

[178] Bo Huang, Bei Zhao, and Yimeng Song. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sensing of Environment*, 214:73–86, 2018.

[179] Liheng Zhong, Lina Hu, and Hang Zhou. Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, 221:430–443, 2019.

[180] Linjing Zhang, Zhenfeng Shao, Jianchen Liu, and Qimin Cheng. Deep learning based retrieval of forest aboveground biomass from combined LiDAR and landsat 8 data. *Remote Sensing*, 11(12), 2019.

[181] Ce Zhang, Isabel Sargent, Xin Pan, Huapeng Li, Andy Gardiner, Jonathon Hare, and Peter M. Atkinson. Joint Deep Learning for land cover and land use classification. *Remote Sensing of Environment*, 221:173–187, 2019.

[182] Patrick Schäfer, Dirk Pflugmacher, Patrick Hostert, and Ulf Leser. Classifying land cover from satellite images using time series analytics. In *CEUR Workshop Proceedings*, volume 2083, pages 10–15, 2018.

[183] Mariana Belgiu and Lucian Drăgu. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, 2016.

[184] P. S.J. Minderhoud, L. Coumou, L. E. Erban, H. Middelkoop, E. Stouthamer, and E. A. Addink. The relation between land use and subsidence in the Vietnamese Mekong delta. *Science of the Total Environment*, 634:715–726, 2018.

[185] Quan H. Nguyen, Dung D. Tran, Khoi K. Dang, Dorien Korbee, Luan D.M.H. Pham, Lan T. Vu, Tang T. Luu, Loc H. Ho, Phat T. Nguyen, Trang T. Trang, Dung T.K. Nguyen, Andrew Wyatt, Maaike van Aalst, Thong A. Tran, and William B. Sea. Land-use dynamics in the Mekong delta: From national policy to livelihood sustainability. *Sustainable Development*, 28(3):448–467, 2020.

[186] F. J. Kriegler, W.A. Malila, R. F. Nalepka, and W. Richardson. Preprocessing transformations and their effects on multispectral recognition. In *Proceedings of the 6th international symposium on remote sensing of environment*, pages 97–131, 1969.

[187] S. K. McFeeters. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7):1425–1432, 1996.

[188] Zhangyan Jiang, Alfredo R. Huete, Kamel Didan, and Tomoaki Miura. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112(10):3833–3845, 2008.

[189] Nikita Roy Mukherjee and Christopher Samuel. Assessment of the temporal variations of surface water bodies in and around Chennai using landsat imagery. *Indian Journal of Science and Technology*, 9(18), 2016.

[190] A. P. Van Deventer, A. D. Ward, P. M. Gowda, and J. G. Lyon. Using thematic mapper data to identify contrasting soil plains and tillage practices. *Photogrammetric Engineering and Remote Sensing*, 63(1):87–93, 1997.

[191] Cong Wang, Jin Chen, Jin Wu, Yanhong Tang, Peijun Shi, T. Andrew Black, and Kai Zhu. A snow-free vegetation index for improved monitoring of vegetation spring green-up date in deciduous ecosystems. *Remote Sensing of Environment*, 196:1–12, 2017.

[192] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[193] A Liaw and M Wiener. Classification and Regression by randomForest. *R News*, 2(3):18–22, 2002.

[194] Vladimir Svetnik, Andy Liaw, Christopher Tong, J. Christopher Culberson, Robert P. Sheridan, and Bradley P. Feuston. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences*, 43(6):1947–1958, 2003.

[195] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1):93–104, 2012.

[196] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15:3133–3181, 2014.

[197] Aaron E. Maxwell, Timothy A. Warner, and Fang Fang. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9):2784–2817, 2018.

[198] Thomas Hilker, Michael A. Wulder, Nicholas C. Coops, Julia Linke, Greg McDermid, Jeffrey G. Masek, Feng Gao, and Joanne C. White. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment*, 113(8):1613–1627, 2009.

[199] Stephen V. Stehman. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62(1):77–89, 1997.

[200] Håkan Berg, Agnes Ekman Söderholm, Anna Sara Söderström, and Nguyen Thanh Tam. Recognizing wetland ecosystem services for sustainable rice farming in the Mekong Delta, Vietnam. *Sustainability Science*, 12(1):137–154, 2017.

[201] Tyler A. Keys and Durelle T. Scott. Monitoring volumetric fluctuations in tropical lakes and reservoirs using satellite remote sensing. *Lake and Reservoir Management*, 34(2):154–166, 2018.

[202] R. Paul Lawson and Paquita Zuidema. Aircraft microphysical and surface-based radar observations of summertime arctic clouds. *Journal of the Atmospheric Sciences*, 66(12):3505–3529, 2009.

[203] Wei Li, Courtney Schumacher, and Sally A. Mcfarlane. Radiative heating of the ISCCP upper level cloud regimes and its impact on the large-scale tropical circulation. *Journal of Geophysical Research Atmospheres*, 118(2):592–604, 2013.

[204] Robert Jackson, Jeffrey French, and Joseph Finlon. Chapter 4 - microphysical properties of convectively forced mixed-phase clouds. In Constantin Andronache, editor, *Mixed-Phase Clouds*, pages 69–96. Elsevier, 2018.

[205] Luyan Ji, Peng Gong, Xiurui Geng, and Yongchao Zhao. Improving the accuracy of the water surface cover type in the 30 m FROM-GLC product. *Remote Sensing*, 7(10):13507–13527, 2015.

[206] Shu'an Liu, Xing Li, Dan Chen, Yuanqiang Duan, Hanyu Ji, Liangpeng Zhang, Qi Chai, and Xiaodong Hu. Understanding Land use/Land cover dynamics and impacts of human activities in the Mekong Delta over the last 40 years. *Global Ecology and Conservation*, 22, 2020.

[207] Cristina Gómez, Joanne C. White, and Michael A. Wulder. Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116:55–72, 2016.

[208] Virginia H. Dale. The relationship between land-use change and climate change. *Ecological Applications*, 7(3):753–769, 1997.

[209] Aiguo Dai. Increasing drought under global warming in observations and models. *Nature Climate Change*, 3(1):52–58, 2013.

[210] Eric F. Lambin, B. L. Turner, Helmut J. Geist, Samuel B. Agbola, Arild Angelsen, John W. Bruce, Oliver T. Coomes, Rodolfo Dirzo, Günther Fischer, Carl Folke, P. S. George, Katherine Homewood, Jacques Imbernon, Rik Leemans, Xiubin Li, Emilio F. Moran, Michael Mortimore, P. S. Ramakrishnan, John F. Richards, Helle Skånes, Will Steffen, Glenn D. Stone, Uno Svedin, Tom A. Veldkamp, Coleen Vogel, and Jianchu Xu. The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change*, 11(4):261–269, 2001.

[211] Xiao Peng Song, Matthew C. Hansen, Stephen V. Stehman, Peter V. Potapov, Alexandra Tyukavina, Eric F. Vermote, and John R. Townshend. Global land change from 1982 to 2016. *Nature*, 560(7720):639–643, 2018.

[212] Nancy B. Grimm, Stanley H. Faeth, Nancy E. Golubiewski, Charles L. Redman, Jianguo Wu, Xuemei Bai, and John M. Briggs. Global change and the ecology of cities. *Science*, 319(5864):756–760, 2008.

[213] Yang Hu and Yunfeng Hu. Detecting forest disturbance and recovery in primorsky krai, russia, using annual landsat time series and multi–source land cover products. *Remote Sensing*, 12(1), 2020.

[214] R. Nemani. Nasa Earth Exchange: Next Generation Earth Science Collaborative. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-8/W20:17–17, 2012.

[215] Theodomir Mugiraneza, Andrea Nascetti, and Yifang Ban. Continuous monitoring of urban land cover change trajectories with landsat time series and landtrendr-google earth engine cloud computing. *Remote Sensing*, 12(18), 2020.

[216] Yuhao Jin, Xiaoping Liu, Jing Yao, Xiaoxiang Zhang, and Han Zhang. Mapping the annual dynamics of cultivated land in typical area of the Middle-lower Yangtze plain using long time-series of Landsat images based on Google Earth Engine. *International Journal of Remote Sensing*, 41(4):1625–1644, 2020.

[217] Jukka Miettinen, Chenghua Shi, and Soo Chin Liew. 2015 Land cover map of Southeast Asia at 250 m spatial resolution. *Remote Sensing Letters*, 7(7):701–710, 2016.

[218] Arsalan Ghorbanian, Mohammad Kakooei, Meisam Amani, Sahel Mahdavi, Ali Mohammadzadeh, and Mahdi Hasanlou. Improved land cover map of Iran using Sentinel imagery within Google Earth Engine and a novel automatic workflow for land cover classification using migrated training samples. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167:276–288, 2020.

[219] Huabing Huang, Jie Wang, Caixia Liu, Lu Liang, Congcong Li, and Peng Gong. The migration of training samples towards dynamic global land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161:27–36, 2020.

[220] Hasi Bagan and Yoshiki Yamagata. Land-cover change analysis in 50 global cities by using a combination of Landsat data and analysis of grid cells. *Environmental Research Letters*, 9(6), 2014.

[221] Julien Radoux, Céline Lamarche, Eric Van Bogaert, Sophie Bontemps, Carsten Brockmann, and Pierre Defourny. Automated training sample extraction for global land cover mapping. *Remote Sensing*, 6(5):3965–3987, 2014.

[222] Vu Kim Chi, Anton Van Rompaey, Gerard Govers, Veerle Vanacker, Birgit Schmook, and Nguyen Hieu. Land Transitions in Northwest Vietnam: An Integrated Analysis of Biophysical and Socio-Cultural Factors. *Human Ecology*, 41(1):37–50, 2013.

[223] Grace B. Villamor, Delia C. Catacutan, Van Anh T. Truong, and Luyen Doan Thi. Tree-cover transition in Northern Vietnam from a gender-specific land-use preferences perspective. *Land Use Policy*, 61:53–62, 2017.

[224] Thuy Dang Truong and Luat Huu Do. Mangrove forests and aquaculture in the Mekong river delta. *Land Use Policy*, 73:20–28, 2018.

[225] Xiaoming Xu, Atul K. Jain, and Katherine V. Calvin. Quantifying the biophysical and socioeconomic drivers of changes in forest and agricultural land in South and Southeast Asia. *Global Change Biology*, 25(6):2137–2151, jun 2019.

[226] Thanh Ngo-Duc, Chanh Kieu, Marcus Thatcher, Dzung Nguyen-Le, and Tan Phan-Van. Climate projections for Vietnam based on regional climate models. *Climate Research*, 60(3):199–213, 2014.

[227] David Saah, Karis Tenneson, Ate Poortinga, Quyen Nguyen, Farrukh Chishtie, Khun San Aung, Kel N. Markert, Nicholas Clinton, Eric R. Anderson, Peter Cutter, Joshua Goldstein, Ian W. Housman, Biplov Bhandari, Peter V. Potapov, Mir Matin, Kabir Uddin, Hai N. Pham, Nishanta Khanal, Sajana Maharjan, Walter L. Ellenberg, Birendra Bajracharya, Radhika Bhargava, Paul Maus, Matthew Patterson, Africa Ixmucane Flores-Anderson, Jeffrey Silverman, Chansopheaktra Sovann, Phuong M. Do, Giang V. Nguyen, Soukanh Bounthabandit, Raja Ram Aryal, Su Mon Myat, Kei Sato, Erik Lindquist, Marija Kono, Jeremy Broadhead,

Peeranan Towashiraporn, and David Ganz. Primitives as building blocks for constructing land cover maps. *International Journal of Applied Earth Observation and Geoinformation*, 85:101979, 2020.

[228] Robert G. Keys. Cubic Convolution Interpolation for Digital Image Processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6):1153–1160, 1981.

[229] Tom G. Farr, Paul A. Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, David Seal, Scott Shaffer, Joanne Shimada, Jeffrey Umland, Marian Werner, Michael Oskin, Douglas Burbank, and Douglas E. Alsdorf. The shuttle radar topography mission. *Reviews of Geophysics*, 45(2), 2007.

[230] Scott A. Soenen, Derek R. Peddle, and Craig A. Coburn. SCS+C: A modified sun-canopy-sensor topographic correction in forested terrain. *IEEE Transactions on Geoscience and Remote Sensing*, 43(9):2148–2159, 2005.

[231] Nazmi Saleous and Jonathan Kutler. LEDAPS Calibration , Reflectance , Atmospheric Correction Preprocessing Code , Version 2 Summary : Data Citation : Model Product Description :. *Oak Ridge National Laboratory Distributed Active Archive Center*, 2012.

[232] Eric Vermote, Chris Justice, Martin Claverie, and Belen Franch. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185:46–56, 2016.

[233] Steve Foga, Pat L. Scaramuzza, Song Guo, Zhe Zhu, Ronald D. Dilley, Tim Beckmann, Gail L. Schmidt, John L. Dwyer, M. Joseph Hughes, and Brady Laue. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment*, 194:379–390, 2017.

[234] Jerome Louis, Bringfried Pflug, Magdalena Main-Knorn, Vincent Debaecker, Uwe Mueller-Wilm, Rosario Quirino Iannone, Enrico Giuseppe Cadau, Valentina Boccia, and Ferran Gascon. Sentinel-2 Global Surface Reflectance Level-2a Product Generated with Sen2Cor. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 8522–8525, 2019.

[235] D. P. Roy, H. K. Zhang, J. Ju, J. L. Gomez-Dans, P. E. Lewis, C. B. Schaaf, Q. Sun, J. Li, H. Huang, and V. Kovalskyy. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, 176:255–271, 2016.

[236] David P. Roy, Jian Li, Hankui K. Zhang, Lin Yan, Haiyan Huang, and Zhongbin Li. Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, 199:25–38, 2017.

[237] Jong Sen Lee, Thomas L. Ainsworth, Yanting Wang, and Kun Shan Chen. Polarimetric SAR speckle filtering and the extended sigma filter. *IEEE Transactions on Geoscience and Remote Sensing*, 53(3):1150–1160, 2015.

[238] D. P. Roy, V. Kovalskyy, H. K. Zhang, E. F. Vermote, L. Yan, S. S. Kumar, and A. Egorov. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment*, 185:57–70, 2016.

[239] Martin Claverie, Junchang Ju, Jeffrey G. Masek, Jennifer L. Dungan, Eric F. Vermote, Jean Claude Roger, Sergii V. Skakun, and Christopher Justice. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219:145–161, 2018.

[240] Yoram J. Kaufman and Didier Tanré. Atmospherically Resistant Vegetation Index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2):261–270, 1992.

[241] Charles R. Perry and Lyle F. Lautenschlager. Functional equivalence of spectral vegetation indices. *Remote Sensing of Environment*, 14(1-3):169–182, 1984.

[242] Hui Qing Liu and Alfredo Huete. Feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Transactions on Geoscience and Remote Sensing*, 33(2):457–465, 1995.

[243] Anatoly A. Gitelson, Andrés Viña, Verónica Ciganda, Donald C. Rundquist, and Timothy J. Arkebauer. Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, 32(8):1–4, 2005.

[244] Alvin B. Baloloy, Ariel C. Blanco, Raymund Rhommel C. Raymund Rhommel, and Kazuo Nadaoka. Development and application of a new mangrove vegetation index (MVI) for rapid and accurate mangrove mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:95–117, 2020.

[245] M. J.López García and V. Caselles. Mapping burns and natural reforestation using thematic mapper data. *Geocarto International*, 6(1):31–37, 1991.

[246] J. P. Lacaux, Y. M. Tourre, C. Vignolles, J. A. Ndione, and M. Lafaye. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. *Remote Sensing of Environment*, 106(1):66–74, 2007.

[247] J. Penuelas, F. Baret, and I. Filella. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, 31(2):221–230, 1995.

[248] Li Shen and Changchun Li. Water body extraction from Landsat ETM+ imagery using adaboost algorithm. In *2010 18th International Conference on Geoinformatics, Geoinformatics 2010*, 2010.

[249] Chris Wright and Alisa Gallant. Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data. *Remote Sensing of Environment*, 107(4):582–605, 2007.

[250] C. F. Hutchinson. Techniques for combining Landsat and ancillary data for digital classification improvement. *Photogrammetric Engineering & Remote Sensing*, 48(1):123–130, 1982.

[251] Congcong Li, Jie Wang, Lei Wang, Luanyun Hu, and Peng Gong. Comparison of classification algorithms and training sample sizes in urban land classification with landsat thematic mapper imagery. *Remote Sensing*, 6(2):964–983, jan 2014.

[252] John A. Richards. *Remote sensing digital image analysis: An introduction*, volume 9783642300622. Springer, Berlin, Heidelberg, 2013.

[253] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F.H. Goetz. The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2-3):145–163, 1993.

[254] Yaqian He, Eungul Lee, and Timothy A. Warner. A time series of annual land use and land cover maps of China from 1982 to 2013 generated using AVHRR GIMMS NDVI3g data. *Remote Sensing of Environment*, 199:201–217, 2017.

[255] Russell G. Congalton. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1):35–46, 1991.

[256] Mario Schmidt. The Sankey diagram in energy and material flow management - Part II: Methodology and current applications. *Journal of Industrial Ecology*, 12(2):173–185, 2008.

[257] Shalei Song, Wei Gong, Bo Zhu, and Xin Huang. Wavelength selection and spectral discrimination for paddy rice, with laboratory measurements of hyperspectral leaf reflectance. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(5):672–682, 2011.

[258] Zoltan Szantoi, Scot E. Smith, Giovanni Strona, Lian Pin Koh, and Serge A. Wich. Mapping orangutan habitat and agricultural areas using Landsat OLI imagery augmented with unmanned aircraft system aerial photography. *International Journal of Remote Sensing*, 38(8-10):2231–2245, 2017.

[259] R. A. Hill and A. G. Thomson. Mapping woodland species composition and structure using airborne spectral and LiDAR data. *International Journal of Remote Sensing*, 26(17):3763–3779, 2005.

[260] Caitlin Kontgis, Annemarie Schneider, Jefferson Fox, Sumeet Saksena, James H. Spencer, and Miguel Castrence. Monitoring peri-urbanization in the greater Ho Chi Minh City metropolitan area. *Applied Geography*, 53:377–388, 2014.

[261] Tuyen V. Ha, Mike Tuohy, Matthew Irwin, and Pham V. Tuan. Monitoring and mapping rural urbanization and land use changes using Landsat data in the northeast subtropical region of Vietnam. *Egyptian Journal of Remote Sensing and Space Science*, 23(1):11–19, 2020.

[262] Christopher Bren D'Amour, Femke Reitsma, Giovanni Baiocchi, Stephan Barthel, Burak Güneralp, Karl Heinz Erb, Helmut Haberl, Felix Creutzig, and Karen C. Seto. Future urban land expansion and implications for global croplands. *Proceedings of the National Academy of Sciences of the United States of America*, 114(34):8939–8944, 2017.

[263] Tran Quoc Nhan, Le Thi Van Ly, and Le Van Tan. How Much Do Rice Farmers Earn from Their Crops? Evidence from a Rice-Exporting Country. *Journal of Agricultural Studies*, 8(1):302, 2020.

[264] Tran Thi Phung Ha, Han van Dijk, and Leontine Visser. Impacts of changes in mangrove forest management practices on forest accessibility and livelihood: A case study in mangrove-shrimp farming system in Ca Mau Province, Mekong Delta, Vietnam. *Land Use Policy*, 36:89–101, 2014.

[265] Thuy Ngan Le, Arnold K. Bregt, Gerardo E. van Halsema, Petra J.G.J. Hellegers, and Lam Dao Nguyen. Interplay between land-use dynamics and changes in hydrological regime in the Vietnamese Mekong Delta. *Land Use Policy*, 73:269–280, 2018.

[266] Quy Van Khuc, Bao Quang Tran, Patrick Meyfroidt, and Mark W. Paschke. Drivers of deforestation and forest degradation in Vietnam: An exploratory analysis at the national level. *Forest Policy and Economics*, 90:128–141, 2018.

[267] Patrick Meyfroidt, Tan Phuong Vu, and Viet Anh Hoang. Trajectories of deforestation, coffee expansion and displacement of shifting cultivation in the Central Highlands of Vietnam. *Global Environmental Change*, 23(5):1187–1198, 2013.

[268] Kirsten L. Findell, Alexis Berg, Pierre Gentine, John P. Krasting, Benjamin R. Lintner, Sergey Malyshev, Joseph A. Santanello, and Elena Shevliakova. The impact of anthropogenic land use and land cover change on regional climate extremes. *Nature Communications*, 8(1), 2017.

[269] Philipp Semenchuk, Christoph Plutzar, Thomas Kastner, Sarah Matej, Giorgio Bidoglio, Karl Heinz Erb, Franz Essl, Helmut Haberl, Johannes Wessely, Fridolin Krausmann, and Stefan Dullinger. Relative effects of land conversion and land-use intensity on terrestrial vertebrate diversity. *Nature Communications*, 13(1):615, 2022.

[270] Min Gon Chung, Kenneth A. Frank, Yadu Pokhrel, Thomas Dietz, and Jianguo Liu. Natural infrastructure in sustaining global urban freshwater ecosystem services. *Nature Sustainability*, 4(12):1068–1075, 2021.

[271] Ran Goldblatt, Klaus Deininger, and Gordon Hanson. Utilizing publicly available satellite data for urban research: Mapping built-up land cover and land use in Ho Chi Minh City, Vietnam. *Development Engineering*, 3:83–99, 2018.

[272] Nguyen Cung Que Truong, Hong Quan Nguyen, and Akihiko Kondoh. Land use and land cover changes and their effect on the flow regime in the upstream Dong Nai River Basin, Vietnam. *Water (Switzerland)*, 10(9), 2018.

[273] Patrick Laux, Phuong Ngoc Bich Nguyen, Johannes Cullmann, and Harald Kunstmann. Impacts of Land-Use/Land-Cover Change and Climate Change on the Regional Climate in the Central Vietnam. In *Water Resources Development and Management*, pages 143–151. Springer Nature, 2017.

[274] Phan Cao Duong, Ta Hoang Trung, Kenlo Nishida Nasahara, and Takeo Tadono. JAXA high-resolution land use/land cover map for Central Vietnam in 2007 and 2017. *Remote Sensing*, 10(9), 2018.

[275] J. K. Ord and Arthur Getis. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, 27(4):286–306, 1995.

[276] Zhihai Ma, Benjamin Zuckerberg, William F. Porter, and Lianjun Zhang. Use of localized descriptive statistics for exploring the spatial pattern changes of bird species richness at multiple scales. *Applied Geography*, 32(2):185–194, 2012.

[277] F. Y. Hsieh, Philip W. Lavori, Harvey J. Cohen, and John R. Feussner. An overview of variance inflation factors for sample-size calculation. *Evaluation and the Health Professions*, 26(3):239–257, 2003.

[278] Ziheng Sun, Laura Sandoval, Robert Crystal-Ornelas, S. Mostafa Mousavi, Jinbo Wang, Cindy Lin, Nicoleta Cristea, Daniel Tong, Wendy Hawley Carande, Xiaogang Ma, Yuhan Rao, James A. Bednar, Amanda Tan, Jianwu Wang, Sanjay Purushotham, Thomas E. Gill, Julien Chastang, Daniel Howard, Benjamin Holt, Chandana Gangodagamage, Peisheng Zhao, Pablo Rivas, Zachary Chester, Javier Orduz, and Aji John. A review of earth artificial intelligence. *Computers & Geosciences*, 159:105034, 2022.

[279] Patrick Leinenkugel, Michel L. Wolters, Natascha Oppelt, and Claudia Kuenzer. Tree cover and forest cover dynamics in the Mekong Basin from 2001 to 2011. *Remote Sensing of Environment*, 158:376–392, 2015.

[280] Isaline Jadin, Veerle Vanacker, and Huong Thi Thu Hoang. Drivers of forest cover dynamics in smallholder farming systems: The case of northwestern vietnam. *Ambio*, 42(3):344–356, 2013.

[281] Nobuo Imai, Takuya Furukawa, Riyou Tsujino, Shumpei Kitamura, and Takakazu Yumoto. Factors affecting forest area change in southeast Asia during 1980-2010. *PLoS ONE*, 13(5), 2018.

[282] Gabrielle Kissinger. Policy responses to direct and underlying drivers of deforestation: Examining rubber and coffee in the central highlands of Vietnam. *Forests*, 11(11), 2020.

[283] Jawaid Ashraf, Rajiv Pandey, and Wil de Jong. Assessment of bio-physical, social and economic drivers for forest transition in Asia-Pacific region. *Forest Policy and Economics*, 76:35–44, 2017.

[284] Thi Thuy Hanh Nguyen. Drivers of forest change in Hoa Binh, Vietnam in the context of integration and globalization. *Singapore Journal of Tropical Geography*, 40(3):452–475, 2019.

[285] Daniele La Rosa and Viviana Pappalardo. Planning for spatial equity - A performance based approach for sustainable urban drainage systems. *Sustainable Cities and Society*, 53, 2020.

[286] N. T. Hai, B. Dell, V. T. Phuong, and R. J. Harper. Towards a more robust approach for the restoration of mangroves in Vietnam. *Annals of Forest Science*, 77(1), 2020.

[287] Thinh An Nguyen, Dung Anh Vu, Phai Van Vu, Thanh Ngoc Nguyen, Tam Minh Pham, Hang Thi Thuy Nguyen, Hai Trinh Le, Thanh Viet Nguyen, Lich Khac Hoang, Thanh Duc Vu, Tung Song Nguyen, Tuyen Thi Luong, Ngoc Phuong Trinh, and Luc Hens. Human ecological effects of tropical storms in the coastal area of Ky Anh (Ha Tinh, Vietnam). *Environment, Development and Sustainability*, 19(2):745–767, 2017.

[288] Daniel Muller and Manfred Zeller. Land use dynamics in the central highlands of Vietnam: a spatial model combining village survey data with satellite imagery interpretation. *Agricultural Economics*, 27(3):333–354, 2002.

[289] Hermann A. Jürgen Pohlan and Marc J.J. Janssens. Growth and Production of Coffee. *Soils, Plant growth and crop production. Volume III*, 2010.

[290] Zhe Li and Jefferson M. Fox. Mapping rubber tree growth in mainland Southeast Asia using time-series MODIS 250 m NDVI and statistical data. *Applied Geography*, 32(2):420–432, 2012.

[291] Jean Christophe Castella, Pham Hung Manh, Suan Pheng Kam, Lorena Villano, and Nathalie Rachel Tronche. Analysis of village accessibility and its impact on land use dynamics in a mountainous province of northern Vietnam. *Applied Geography*, 25(4):308–326, 2005.

[292] Thomas Sikor. The allocation of forestry land in Vietnam: Did it cause the expansion of forests in the northwest? *Forest Policy and Economics*, 2(1):1–11, 2001.

[293] John C. O'Connor, Stefan C. Dekker, Arie Staal, Obbe A. Tuinenburg, Karin T. Rebel, and Maria J. Santos. Forests buffer against variations in precipitation. *Global Change Biology*, 27(19):4686–4696, 2021.

[294] Mara Baudena, Obbe A. Tuinenburg, Pendula A. Ferdinand, and Arie Staal. Effects of land-use change in the Amazon on precipitation are likely underestimated. *Global Change Biology*, 27(21):5580–5587, 2021.

[295] Ruben D. Molina, Juan Fernando Salazar, J. Alejandro Martínez, Juan Camilo Villegas, and Paola A. Arias. Forest-Induced Exponential Growth of Precipitation Along Climatological Wind Streamlines Over the Amazon. *Journal of Geophysical Research: Atmospheres*, 124(5):2589–2599, 2019.

[296] Do Hoai Nam, Phan Cao Duong, Duong Hai Thuan, Dang Thanh Mai, and Nguyen Quoc Dung. Assessment of near-term runoffresponse at a river basin scale in central Vietnam using direct CMIP5 high-resolution model outputs. *Water (Switzerland)*, 10(4), 2018.

[297] Do Hoai Nam, Tran Dinh Hoa, Phan Cao Duong, Duong Hai Thuan, and Dang Thanh Mai. Assessment of flood extremes using downscaled CMIP5 high-resolution ensemble projections of near-term climate for the upper Thu Bon catchment in Vietnam. *Water (Switzerland)*, 11(4), 2019.

[298] Juliana Siqueira-Gay, Laura J. Sonter, and Luis E. Sánchez. Exploring potential impacts of mining on forest loss and fragmentation within a biodiverse region of Brazil's northeastern Amazon. *Resources Policy*, 67, 2020.

[299] Ho Nguyen, Ta Hoang Trung, Duong Cao Phan, Thong Anh Tran, Nguyen Thi Hai Ly, Kenlo Nishida Nasahara, Alexander V. Prishchepov, and Norbert Hölzel. Transformation of rural landscapes in the vietnamese mekong delta from 1990 to 2019: A spatio-temporal analysis. *Geocarto International*, 0(ja):1–23, 2022.

[300] Thiha, Edward L. Webb, and Kiyoshi Honda. Biophysical and policy drivers of landscape change in a central Vietnamese district. *Environmental Conservation*, 34(2):164–172, 2007.

[301] Stephen E. Fick and Robert J. Hijmans. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12):4302–4315, 2017.

[302] Tom G. Farr and Mike Kobrick. Shuttle radar topography mission produces a wealth of data. *Eos*, 81(48):583–585, 2000.

[303] Yamei Wang, Zhongwu Li, Zhenghong Tang, and Guangming Zeng. A GIS-Based Spatial Multi-Criteria Approach for Flood Risk Assessment in the Dongting Lake Region, Hunan, Central China. *Water Resources Management*, 25(13):3465–3484, 2011.

[304] Debrati Guha-Sapir Margareta Wahlstrom. The Human Costs of Weather Related Disasters. Technical report, The Centre for Research on the Epidemiology of Disasters and the United Nations International Strategy for Disaster Reduction, Geneva, Switzerland, 2015.

[305] Lorenzo Alfieri, Berny Bisselink, Francesco Dottori, Gustavo Naumann, Ad de Roo, Peter Salamon, Klaus Wyser, and Luc Feyen. Global projections of river flood risk in a warmer world. *Earth's Future*, 5(2):171–182, 2017.

[306] Mohammed Sarfaraz Gani Adnan, Abu Yousuf Md Abdullah, Ashraf Dewan, and Jim W. Hall. The effects of changing land use and flood hazard on poverty in coastal Bangladesh. *Land Use Policy*, 99, 2020.

[307] Paramita Roy, Subodh Chandra Pal, Rabin Chakrabortty, Indrajit Chowdhuri, Sadhan Malik, and Biswajit Das. Threats of climate and land use change on future flood susceptibility. *Journal of Cleaner Production*, 272, 2020.

[308] Mahfuzur Rahman, Chen Ningsheng, Golam Iftekhar Mahmud, Md Monirul Islam, Hamid Reza Pourghasemi, Hilal Ahmad, Jules Maurice Habumugisha, Rana Muhammad Ali Washakh, Mehtab Alam, Enlong Liu, Zheng Han, Huayong Ni, Tian Shufeng, and Ashraf Dewan. Flooding and its relationship with land cover change, population growth, and road density. *Geoscience Frontiers*, 12(6), 2021.

[309] Shahab S. Band, Saeid Janizadeh, Subodh Chandra Pal, Asish Saha, Rabin Chakrabortty, Assefa M. Melesse, and Amirhosein Mosavi. Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. *Remote Sensing*, 12(21):1–23, 2020.

[310] Yi Wang, Zhice Fang, Haoyuan Hong, and Ling Peng. Flood susceptibility mapping using convolutional neural network frameworks. *Journal of Hydrology*, 582, 2020.

[311] Hai Hoa Nguyen, Nghia Huu Nghia, Hien Thi Thu Nguyen, An Thanh Le, Lan Thi Ngoc Tran, Linh Vo Khanh Duong, Simone Bohm, and Michael J. Furniss. Classification methods for mapping mangrove extents and drivers of change in Thanh Hoa province, Vietnam during 2005-2018. *Forest and Society*, 4(1):225–242, 2020.

[312] Quan H. Nguyen, Dung D. Tran, Khoi K. Dang, Dorien Korbee, Luan D.M.H. Pham, Lan T. Vu, Tang T. Luu, Loc H. Ho, Phat T. Nguyen, Trang T. Trang, Dung T.K. Nguyen, Andrew Wyatt, Maaike van Aalst, Thong A. Tran, and William B. Sea. Land-use dynamics in the Mekong delta: From national policy to livelihood sustainability. *Sustainable Development*, 28(3):448–467, 2020.

[313] Thanh Van Nguyen, Jie Hua Lv, Thi Thanh Huyen Vu, and Bin Zhang. Determinants of non-timber forest product planting, development, and trading: Case study in central Vietnam. *Forests*, 11(1), 2020.

[314] Thuong V. Tran, Duy X. Tran, Soe W. Myint, Cho ying Huang, Hoa V. Pham, Tung H. Luu, and Tien M.T. Vo. Examining spatiotemporal salinity dynamics in the Mekong River Delta using Landsat time series imagery and a spatial regression approach. *Science of the Total Environment*, 687:1087–1097, 2019.

[315] T. Nguyen, S. Lawler, and W. Paul. Socioeconomic and indigeneity determinants of the consumption of non-timber forest products in Vietnam's Bu Gia Map National Park. *International Journal of Sustainable Development and World Ecology*, 26(7):646–656, 2019.

[316] Minh Tu Nguyen, Fabrice G. Renaud, and Zita Sebesvari. Drivers of change and adaptation pathways of agricultural systems facing increased salinity intrusion in coastal areas of the Mekong and Red River deltas in Vietnam. *Environmental Science and Policy*, 92:331–348, 2019.

[317] Philippe Rufin, Christian Levers, Matthias Baumann, Jonas Jägermeyr, Tobias Krueger, Tobias Kuemmerle, and Patrick Hostert. Global-scale patterns and determinants of cropping frequency in irrigation dam command areas. *Global Environmental Change*, 50:110–122, 2018.

[318] Leon T. Hauser, Giang Nguyen Vu, Binh An Nguyen, Emma Dade, Hieu Minh Nguyen, Trang Thi Quynh Nguyen, Toan Quang Le, Long Huu Vu, Ai Thi Huyen Tong, and Hoa Viet Pham. Uncovering the spatio-temporal dynamics of land cover change and fragmentation of mangroves in the Ca Mau peninsula, Vietnam using multi-temporal SPOT satellite imagery (2004–2013). *Applied Geography*, 86:197–207, 2017.

[319] Roland Cochard, Dung Tri Ngo, Patrick O. Waeber, and Christian A. Kull. Extent and causes of forest cover changes in Vietnam's provinces 1993-2013: A review and analysis of official data. *Environmental Reviews*, 25(2):199–217, 2017.

[320] Hoang Huu Nguyen, Paul Dargusch, Patrick Moss, and Ammar Abdul Aziz. Land-use change and socio-ecological drivers of wetland conversion in Ha Tien Plain, Mekong Delta, Vietnam. *Land Use Policy*, 64:101–113, 2017.

[321] Trung Thanh Nguyen, Loc Duc Nguyen, Rattiya Suddeephong Lippe, and Ulrike Grote. Determinants of Farmers' Land Use Decision-Making: Comparative Evidence From Thailand and Vietnam. *World Development*, 89:199–213, 2017.

[322] Alexander Chapman and Stephen Darby. Evaluating sustainable adaptation strategies for vulnerable mega-deltas using system dynamics modelling: Rice agriculture in the Mekong Delta's An Giang Province, Vietnam. *Science of the Total Environment*, 559:326–338, 2016.

[323] Christian M.S. Knudsen and Ole Mertz. Improved land tenure not the driver of economic development in a Vietnamese community. *Geografisk Tidsskrift - Danish Journal of Geography*, 116(1):82–84, 2016.

[324] Sarah Turner and Thi Thanh Hiên Pham. "Nothing is like it was before": The dynamics between land-use and land-cover, and livelihood strategies in the northern Vietnam borderlands. *Land*, 4(4):1030–1059, 2015.

[325] Thomas K. Rudel. Have tropical deforestation's changing dynamics created conservation opportunities? A historical analysis. *Environmental Conservation*, 42(2):108–118, 2015.

[326] Hai Hoa Nguyen, Clive McAlpine, David Pullar, Stephen Joseph Leisz, and Gramotnev Galina. Drivers of Coastal Shoreline Change: Case Study of Hon Dat Coast, Kien Giang, Vietnam. *Environmental Management*, 55(5):1093–1108, 2015.

[327] Quyet Manh Vu, Quang Bao Le, Emmanuel Frossard, and Paul L.G. Vlek. Socio-economic and biophysical determinants of land degradation in Vietnam: An integrated causal analysis at the national level. *Land Use Policy*, 36:605–617, 2014.

[328] T. T.T. Thuy, N. T. Yen, N. T.A. Tuyet, L. L. Te, and D. de Waele. Population dynamics of Meloidogyne incognita on black pepper plants in two agro-ecological regions in Vietnam. *Archives of Phytopathology and Plant Protection*, 45(13):1527–1537, 2012.

[329] Floriane Clement, Didier Orange, Meredith Williams, Corinne Mulley, and Michael Epprecht. Drivers of afforestation in Northern Vietnam: Assessing local variations using geographically weighted regression. *Applied Geography*, 29(4):561–576, 2009.

[330] Thao Tran. Landscapes of mangrove forests and littoral dynamics in the South Viêt-Nam. *Journal of Coastal Conservation*, 13(2):65–75, 2009.

[331] Andreas Heinimann, Peter Messerli, Dietrich Schmidt-Vogt, and Urs Wiesmann. The dynamics of secondary forest landscapes in the lower Mekong Basin: A regional-scale analysis. *Mountain Research and Development*, 27(3):232–241, 2007.

[332] N.T. Son. Relation between Land-Use Change and Rural Income Determinants: A Case Study of Policy Implication in Tri Ton, Vietnam. *The Electronic Journal of Information Systems in Developing Countries*, 29(1):1–9, 2007.

[333] Jean Christophe Castella and Peter H. Verburg. Combination of process-oriented and pattern-oriented models of land-use change in a mountain area of Vietnam. *Ecological Modelling*, 202(3-4):410–420, 2007.

[334] Patrick Meyfroidt and Eric F. Lambin. Forest transition in Vietnam and displacement of deforestation abroad. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38):16139–16144, 2009.

[335] Hoang Huu Nguyen, Paul Dargusch, Patrick Moss, and Da Binh Tran. A review of the drivers of 200 years of wetland degradation in the Mekong Delta of Vietnam. *Regional Environmental Change*, 16(8):2303–2315, 2016.

[336] Petrina Rowcroft. Frontiers of change: The reasons behind land-use change in the Mekong Basin. *Ambio*, 37(3):213–218, 2008.

Appendix A

A

Table A.1: The fundamental characteristics of global land cover databases.

Database		**IGBP**	**UMD**	**GLC**
Producer		USGS, UNL, and JRC	UMD	JRC
Sensor		AVHRR	AVHRR	SPOT VEGETATION-1
Input Data	Primary Input	Monthly NDVI	41 metrics from bands and NDVI	4 bands and NDVI
	Collection Date	4/1992-3/1993	4/1992-3/1993	11/1999-12/2000
	Ancillary Data	DEM Atlases of ecoregion, soils, vegetation, and Land cover maps	Landsat MSS	Radar DMSP and Elevation Data (ETOPO5)
	Spatial Resolution	1000 m	1000 m	1000 m
Processing	Projection	Goode Homolosine Equal Area	Goode Homolosine Equal Area	Lat-Lon (WGS)
	Geometric Correction	Geo-registered to Goode Homolosine Equal Area	Geo-registered to Goode Homolosine Equal Area	Ortho-rectification with ETOPO5, resampled by bi-cubic convolution

Continued on next page

Table A.1 – continued from previous page

	Atmospheric Correction	Reduction of contamination and off-nadir viewing effects by NDVI	Correct Ozone, Rayleigh scattering and solar zenith angle	Cloud screening and Reduce Abrupt signal drops
Classification	No. Classes	17	14	22
	Training Site	N/A	37294	N/A
	Classification Scheme	IGBP	IGBP	LCCS
	Classification Method	Unsupervised clustering with post-classification refinement	Supervised Decision Tree	Unsupervised classification with ISODATA algorithm
	Processing Sequence	Continent-by-continent	Global	Region by region
Accuracy Assessment	Validation Method	Statistical validation	None	Statistical validation
	Sampling Method	Stratified random sampling by classes	Unknown	Two-stage stratified clustered sampling
	Reference Data	Landsat TM and Spot images	Unknown	Landsat TM and Spot images
	Overall Accuracy	66.90%	Unknown	68.60%
Sources (Availability)		1992	1992	2000
Continued				
Database		**GlobCover**	**CCI**	**MODIS LC**
Producer		ESA	ESA	USGS
Sensor		ENVISAT MERIS	AVHRR, SPOT-VGT MERIS, and PROBA-V	MODIS

Continued on next page

Table A.1 – continued from previous page

Input Data	Primary Input	13 bands and NDVI	Bands and NDVI	Spectro-temporal features and EVI
	Collection Date	1/2009-12/2009	1992-2018	2001-2019
	Ancillary Data	Altimeter Corrected Elevation (Getasse 30)	Unknown	DEM, LST, and Texture.
	Spatial Resolution	300 m	300 m	500/1000 m
Processing	Projection	Lat-Lon (WGS)	Lat-Lon (WGS)	Lat-Lon (WGS)
	Geometric Correction	Level 1B data corrected into Level 3 Mosaics using AMORGOS tool	Orthorectified images-so called MERIS FSG products from MERIS FR products	Sub-pixel geometric registration
	Atmospheric Correction	Cloud screening Rayleigh scattering & Aerosol correction	Relies on a neural network derived from MOMO model	A simple SACA algorithm
Classification	No. Classes	22	22	17
	Training Site	Unknown	3167	1860
	Classification Scheme	LCCS	LCCS	IGBP, UMD, LAI, BGC, PFT, and LCCS.
	Classification Method	Per-pixel supervised (urban and wetland) and unsupervised	Supervised and per-pixel unsupervised classification	Supervised Decision Tree
	Processing Sequence	Global	Region by region	Global

Continued on next page

Table A.1 – continued from previous page

Accuracy Assessment	Validation Method	Statistical validation	Statistical validation	Cross-validation analysis
	Sampling Method	Stratified random sampling by classes	Stratified random sampling by classes	Random sampling
	Reference Data	Reference dataset from multiple sources	Reference dataset from multiple sources	Training site database
	Overall Accuracy	67.50%	73.40%	74.80%
Sources (Availability)		2009	1992-2018	2001-2019
Continued				
Database		**GLCNMO**	**FROM-GLC**	**Globeland30**
Producer		Chiba University	National Geomatics Center of China	China
Sensor		MODIS, Landsat, and DMSP-OLS	Landsat TM, ETM+, and OLI	Landsat TM, ETM+ and HJ-1
Input Data	Primary Input	7-bands	Spectral bands	Spectral bands and spectral indexes
	Collection Date	2008	1981-2011	2000/2010
	Ancillary Data	Previous land cover data	MODIS EVI. DEM	DEM and ecological zones.
	Spatial Resolution	500 m	30 m	30 m
Processing	Projection	Lat-Lon (WGS)	Lat-Lon (WGS)	Lat-Lon (WGS)
	Geometric Correction	A root mean square error (RMSE)	A root mean square error (RMSE)	Satisfactorily geo-referenced and collinearity equations and Lagrange interpolation

Continued on next page

Table A.1 – continued from previous page

	Atmospheric Correction	Filled cloud-contaminated pixels by time-series linear interpolation	Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm	MODerate resolution atmospheric TRANsmission (MODTRAN)
Classification]	No. Classes	20	18	10
	Training Site	904	36,630	150,000
	Classification Scheme	LCCS	LCCS and IGBP	LCCS and IGBP
	Classification Method	Supervised Decision Tree	Support Vector Machine	Pixel-Object-Knowledge-based (POK-based) method
	Processing Sequence	Class-by-class	Global	Tile by tile (5o x 5o)
Accuracy Assessment	Validation Method	A confusion matrix	A confusion matrix	A confusion matrix
	Sampling Method	Stratified random sampling	Unaligned sampling strategy	Stratified random sampling by classes
	Reference Data	Reference dataset	Reference dataset from multiple sources	Reference dataset from multiple sources
	Overall Accuracy	77.90%	64.90%	83.50%
Sources (Availability)		2003/2008	2015/2017	2000/2010/2020
Continued				
Database		**GLC_FCS30**	**CGLS-LC100**	**ESRI-LC10**
Producer		Chinese academy of sciences	Copernicus Global Land Service	ESRI

Continued on next page

Table A.1 – continued from previous page

Sensor		Landsat OLI	PROBA-V and Sentinel-2	Sentinel-2
Input Data	Primary Input	Six optical bands	183 metrics	Spectral bands
	Collection Date	2014-2016; 2019-2020	2015-2019	2020
	Ancillary Data	DEM terrain elevation data	Height, slope, aspect, and purity	N/A
	Spatial Resolution	30 m	100 m	10 m
Processing	Projection	Lat-Lon (WGS)	Lat-Lon (WGS)	Universal Transverse Mercator (UTM)
	Geometric Correction	Unknown	Ortho-rectification and spatial registration on a global reference system	Sen2Cor for Sentinel-2
	Atmospheric Correction	Landsat Surface Reflectance Code (LaSRC) and CFMask.	Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Surface Reflectance Algorithm	Sen2Cor for Sentinel-2
Classification	No. Classes	10	23	10
	Training Site	44,043	21,700	23,693,961
	Classification Scheme	LCCS	LCCS	LCCS
	Classification Method	Random Forest Algorithm	Supervised classification and regression	Deep learning

Continued on next page

Table A.1 – continued from previous page

	Processing Sequence	Tile by tile	Region by region	N/A
Accuracy Assessment	Validation Method	A confusion matrix	A confusion matrix	N/A
	Sampling Method	Stratified random sampling by classes	Stratified random sampling by classes	N/A
	Reference Data	Reference dataset from multiple sources	In-situ data	N/A
	Overall Accuracy	82.50%	80.60%	85.96%
Sources (Availability)		2015/2020	2015-2019	2020

Table A.2: Common software was used recently for land use/land cover analyses.

Software	**Source**	**Key functions**	**Advantage**
ArcGIS	ESRI (Environmental Systems Research Institute)	Geospatial reference, pan-sharpening, mosaicking, spatio-temporal processing, supervised and unsupervised classification, post-classification, accuracy validation, and change detection and analysis.	This tool can efficiently process with multispectral and hyper-spectral remote sensing images.
ENVI	Harris Geospatial Solutions	Radiometric correction, topographic correction, image sharpening, anomaly detection, geometric correction, feature extraction, mosaicking, supervised and unsupervised classification, post-classification, accuracy validation, and spatio-temporal change assessment.	This tool can effectively handle with multispectral and hyper-spectral remote sensing datasets.

Continued on next page

Table A.2 – continued from previous page

Software	Source	Key functions	Advantage
QGIS	QGIS Developers Team	Preprocessing, post-processing, supervised and unsupervised classification, projection, and calibrating the terrain features.	This tool can be easily to add in newly effective plugins.
ERDAS Imagine	Hexagon Geospatial	Pan-sharpening, geometric calibration, radiometric processing, sub-pixel, supervised and unsupervised classification, and change analysis.	This software can practically process with multispectral and hyper-spectral image data.
IDRISI	Clark Labs	Preprocessing, clustering, modelling dependent and independent variables, classification, and LULC change analysis.	IDRISI allows to deal effectively with different remote sensing sources, especially hyper-spectral imagery data.
GEE	Google	Pre-processing, data fusion and harmonization, classification, clustering and LU/LC change detection.	GEE can effectively process with time-series images and large-scale analyses.
Python	Python Software	Feature extraction, mosaicking, supervised and unsupervised classification, post-classification, accuracy validation, and spatio-temporal change assessment.	Python provides with various geospatial libraries, which can handle effectively different tasks.
Matlab	Math Works	Importing and exporting the geographic data, and vector data representation. It can also deal with projections, coordinate transformations, terrain processing, and web mapping.	Matlab is an open-source, which is flexibly to handle with new processes.
Rstudio	RStudio	Reading, writing and plotting spatial data, time-series analysis, and supervised and unsupervised classification.	Rstudio can deal efficiently with time-series analysis.

Table A.3: A list of location notices of different land use/land cover types (LULCT) in Google Map and GPS photos were taken from our field surveys in 2020.

Land types	Access: url	Land types	Access: url
Rice paddy	here	Coffee	here
Coffee	here	Pipeapple	here
Constructions	here	Melaleuca	here
Aquaculture	here	Wet grass	here
Mangrove	here	Urban forest	here
Coconut	here	Burned forest	here
Grassland	here	Natural forest	here
Cropland	here	Shrub-land	here
Lotus	here	Cashew	here
Cane	here	Dry forest	here
Potato	here	Evergreen neddleleaf forest	here
Orchard	here	Pepper	here
Bamboo	here	House	here
Dragon	here	Acacia	here
Longan	here	Sand dune	here
Mango	here	Cassava	here
Flower	here	Rubber	here
Chilly	here	Corn	here
Banana	here	Field survey photos	here

Table A.4: The description of land use/cover classification system.

Primary dominant land cover	Code	Secondary dominant land cover	Description
Residential land	R1	High developed areas	Land covered by buildings and other man-made structures > 50% of constructed cover and < 20% vegetation.
	R2	Low developed areas	Land covered by buildings and other man-made structures < 50% of constructed cover and < 20% vegetation.
Rice paddies	RP	Rice paddies	Rice fields have a area of greater than 30 x 30 m

Continued on next page

Table A.4 – continued from previous page

Primary dominant land cover	Code	Secondary dominant land cover	Description
Cropland	WC	Woody crops	Perennial woody crops such coffee, tea, woody orchards, and so on.
	IC	In-house crops	Crops are planted in agricultural plastic houses, which have an area go larger than 30 m by 30 m.
	OC	Other croplands	Lands covered with temporary crops followed by harvest and a bare soil period (e.g., single and multiple cropping systems). It is noted that perennial woody crops will be classified as the woody crops or the appropriate forest or shrub land cover type.
Grassland	GL	Grassland/ Herbaceous vegetation	> 80 % of herbaceous land; not intensive management but can be utilized for grazing. Tree and shrub cover is less than 10%
Barren land	BL	Barren land	Lands with exposed soil, sand, rocks, or snow and never have more than 10% vegetated cover during any time of the year.
Scrub land	SL	Scrub/Shrub	Areas are dominated by shrubs less than 5% of 5 meter tall. Shrub canopy is typically greater than 10% of total vegetation. This category includes tree shrubs, young trees in an early successional stage, or trees stunted from environmental conditions.
Forest land	DBF	Deciduous broadleaf forest	Lands dominated by woody vegetation with a percent cover > 60% and height exceeding 5 m. Consisting of broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.
	EBF	Evergreen broadleaf forest	Lands dominated by broadleaf woody vegetation with a percent cover > 60% and height exceeding 5 m. Almost all trees remain green year round. Canopy is never without green foliage.
	ENF	Evergreen needleleaf forest	Lands dominated by needle-leaf woody vegetation with a percent cover > 60% and height exceeding 5 m. Almost all trees remain green all year. Canopy is never without green foliage.

Continued on next page

Table A.4 – continued from previous page

Primary dominant land cover	Code	Secondary dominant land cover	Description
	PL	Plantation land	Lands dominated by plantation trees with a percent cover > 60% and height exceeding 5 m.
	BA	Bamboo area	Lands dominated by bamboo with a percent cover > 60%.
Wetland	IW	Inland wetland	Include tidal and non-tidal wetlands dominated by vegetation. Note that mangrove is classified as the mangrove forests.
	MF	Mangrove	Lands dominated by mangrove trees with a percent cover > 60% and height exceeding 5 m.
Open water	OW	Open water	Oceans, seas, lakes, reservoirs, and rivers which can be either fresh or saltwater bodies.
Aquaculture	AC	Aquaculture	Contains areas dominated by aquatic organism and plant farming, including artificial ponds and temporarily flooded areas.

Table A.5: Bands, indices, and ancillary data were used to create annual land use/cover maps of Vietnam from 1990 to 2020.

Data source	Order	Covariates	Resolution (m)	
			1990-2014	2015-2020
Landsat TM, ETM+ and OLI, and Sentinel MSI	1	blue	30	10
	2	green	30	10
	3	red	30	10
	4	nir	30	10
	5	swir1	30	10
	6	swir2	30	10
	7	thermal	30	10
	8	p20_green	30	10
	9	p20_nir	30	10
	10	p20_blue	30	10
	11	p20_red	30	10
	12	p20_swir1	30	10
	13	p20_swir2	30	10
	14	p80_green	30	10

Continued on next page

Table A.5 – continued from previous page

Data source	Order	Covariates	Resolution (m)	
			1990-2014	2015-2020
	15	p80_nir	30	10
	16	p80_blue	30	10
	17	p80_red	30	10
	18	p80_swir1	30	10
	19	p80_swir2	30	10
	20	blue_stdDev	30	10
	21	red_stdDev	30	10
	22	green_stdDev	30	10
	23	nir_stdDev	30	10
	24	swir1_stdDev	30	10
	25	swir2_stdDev	30	10
	26	ND_green_swir1_stdDev	30	10
	27	ND_nir_red_stdDev	30	10
	28	ND_nir_swir2_stdDev	30	10
	29	blue/green	30	10
	30	red/blue	30	10
	31	red/green	30	10
	32	red/nir	30	10
	33	nir/(redswir1)	30	10
	34	ARVI	30	10
	35	DVI	30	10
	36	EBBI	30	10
	37	EVI	30	10
	38	GCI	30	10
	39	MVI	30	10
	40	NBR	30	10
	41	NDBaI	30	10
	42	NDBI	30	10
	43	NDPI	30	10
	44	NDTI	30	10
	45	MDVI	30	10
	46	NDWI	30	10
	47	SAVI	30	10
	48	SIPI	30	10
	49	UI	30	10
	50	WRI	30	10
Sentinel SAR GRD	51	VV		10
	52	VH		10
	53	ND_VV_VH	-	10

Continued on next page

Table A.5 – continued from previous page

Data source	Order	Covariates	Resolution (m)	
			1990-2014	2015-2020
AW2D30	54	elevation	30	10
	55	slope	30	10
	56	aspect	30	10
OpenStreetMap	57	distRoad	30	10
	58	distBuilding	30	10
OpenDevelopment Mekong	59	distCoast	30	10
	60	distRiver	30	10
	61	soilType	30	10

Table A.6: Synthesis of the case studies on changes in forests in Vietnam.

Study#	LULCC	Key results (Drivers)	Details	Methods	Reference
1	Deforestation	Rubber and coffee production	Location: Central Highlands Study period: 2005 - 2015	Multi-level-factor analysis framework.	[282]
2	Land use/land cover change	Economic development; and Land policy	Location: Mekong Delta Study period: 1979 - 2015	Descriptive statistics	[206]
3	Reforestation (Mangrove gain)	National programs of mangrove rehabilitation	Location: Thanh Hoa province Study period: 2005 - 2018	Descriptive statistics	[311]
4	Agriculture expansion	National policy	Location: Mekong Delta Study period: 1990 - 2020	Institutional analyses	[312]

Continued on next page

Table A.6 – continued from previous page

Study#	LULCC	Key results (Drivers)	Details	Methods	Reference
5	Reforestation	Income; Market influence; Close proximity to forest; Technical training; Number of labours; and Understanding of forestry economic policies.	Location: Central Vietnam Study period: 2010 - 2015	Combining logistic and tobit methods	[313]
6	Agricultural shifting (Rice)	Salinity intrusion	Location: Mekong Delta Study period: 1989 - 2018	Landsat imagery classification and a spatial regression	[314]
7	Deforestation (Non-timber forest consumption)	Indigeneity; Education; Labour; Fuel wood; Family size; and The area of land.	Location: Bu Gia Map National Park Study period: 2015	Interview analysis	[315]
8	Deforestation and reforestation (forest regrowth, and plantation expansion).	Accessibility; Livelihood; Key policies; and Local economic development	Location: Hoa Binh province Study period: 2005 - 2017	Multiple Logistic Regression	[284]
9	Agriculture expansion	Policy intervention; Market influence; and Salinity intrusion.	Location: Mekong Delta Study period: 2015 – 2016	The lens of complex adaptive system theory	[316]

Continued on next page

Table A.6 – continued from previous page

Study#	LULCC	Key results (Drivers)	Details	Methods	Reference
10	Agriculture expansion	Dam construction	Location: Vietnam-wide Study period: 2011 – 2012	Boosted Regression Tree	[317]
11	Deforestation	Initial forest cover; Per capital income; Agriculture production; Population growth; and Poverty.	Location: Vietnam-wide Study period: 2000 - 2010	A structural regression model and a regression tree model.	[266]
12	Agricultural increase (Intensification of agriculture)	Hydrological regime changes	Location: Mekong Delta Study period: 2001 – 2012	Time-series imagery interpretation and descriptive analysis.	[265]
13	Deforestation (Mangrove)	Expansion of aquaculture production and urbanization	Location: Ca Mau province Study period: 2004 – 2013	SPOT satellite imagery interpretation.	[318]
14	Forest change (Plantation expansion, deforestation and afforestation)	Deforestation due to cash crop expansion, and migration and population growth; and Reforestation due to policies intervention and effects of forest management.	Location: Vietnam-wide Study period: 1993 – 2013	Review and analysis.	[319]

Continued on next page

Table A.6 – continued from previous page

Study#	LULCC	Key results (Drivers)	Details	Methods	Reference
15	Deforestation (Wetland conversion)	Food supply; Improving income; and Unclear land tenure.	Location: Kien Giang province Study period: 2004 - 2014	A system thinking approach	[320]
16	Agriculture increase (intensification)	Agricultural system increase; and Technical and economic factors	Location: Hai Duong province Study period: 1980 - 2010	A systematic analysis of agricultural systems	[19]
17	Reforestation	Agricultural reduction; and Urban population.	Location: Vietnam-wide Study period: 1962 – 2011	Discriminant analysis	[283]
18	Agriculture change	Weather; Accessibility; and Livelihood	Location: Vietnam-wide Study period: 2007 – 2013	A sustainable livelihood framework	[321]
19	Agriculture increase (Rice intensification)	Irrigation construction and operation (dyke and sluice).	Location: An Giang province Study period: 2005 - 2015	A system dynamics method	[322]
20	Crop transition	Policy; and Land tenure.	Location: Vietnam-wide Study period: Since 1986	Review and descriptive method	[323]
21	Reforestation	Plantation expansion	Location: Lai Chau, Lao Cai, and Ha Giang provinces Study period: 2000 – 2009	Remote sensing image and fieldwork interpretation.	[324]

Continued on next page

Table A.6 – continued from previous page

Study#	LULCC	Key results (Drivers)	Details	Methods	Reference
22	Deforestation	Easy access ability; and Lack of conservation policies	Location: Vietnam-wide Study period: since 1990s	Review and analysis	[325]
23	Deforestation (Mangrove loss)	Aquaculture and agricultural expansion.	Location: Kien Giang province Study period: 1995 – 2009	Generalized linear mixed-effects models.	[326]
24	Agricultural decline (agriculture conversion into aquaculture)	Expansion of aquaculture and urbanization;	Location: Ca Mau province Study period: 1973 – 2011	SPOT imagery interpretation.	[15]
25	Deforestation	Forest clearance; and Agricultural expansion	Location: Mekong Delta Study period: 2001 – 2011	MODIS imagery interpretation	[279]
26	Deforestation Agricultural decline (rice decrease)	Climate change	Location: Vietnam-wide Study period: since 1990s	Global-to-local modelling	[67]
27	Deforestation	Agricultural expansion; and Population growth	Location: Vietnam-wide Study period: 2000 – 2010	Multi-linear regression and binary logistic regression.	[327]
28	Deforestation and reforestation.	Expansion of agriculture and plantation.	Location: Lao Cai province Study period: since 1990s	Review and statistical analysis	[280]

Continued on next page

Table A.6 – continued from previous page

Study#	LULCC	Key results (Drivers)	Details	Methods	Reference
29	Agriculture increase	Rainfall and temperature differences; and Population density	Location: Quang Tri and Dak Lak provinces Study period: 2010 – 2013	Sample collection and statistical analysis.	[328]
30	Reforestation (Afforestation)	Wood-processing industry demand; Distance to highways; and Population density.	Location: Northern Vietnam Study period: 1993 – 2000	Geographically weighted regression	[329]
31	Deforestation (Mangrove)	Defoliated; and Boom	Location: Ca Mau province Study period: 1961 – 1975	Remote sensing imagery interpretation	[330]
32	Reforestation	Plantation expansion; Policy introduction; and Decline cultivation of crops	Location: Vietnam-wide Study period: since 1990s	Spatial lag regression model	[150]
33	Deforestation	Expansion of aquaculture and crops	Location: Mekong Delta Study period: 1993 – 1997	Review and history data analysis.	[331]
34	Forest change	Policy	Location: Dac Nong province Study period: 1975 – 2004	Linear regression model	[300]
35	Deforestation and agricultural expansion.	Change in benefit of crops; and Income.	Location: An Giang province Study period: 2000 – 2005	Descriptive method based on historical data	[332]

Continued on next page

Table A.6 – continued from previous page

Study#	LULCC	Key results (Drivers)	Details	Methods	Reference
36	Land use change	Accessibility to market	Location: Bac Can province Study period: 2000	GIS distance analysis	[291]
37	Land cover (deforestation and agriculture expansion)	Expansion of agriculture; Increasing accessibility; Increase irrigated areas; Policy introduction; Rainfall; Topography; and Soil types	Location: Central Highlands Study period: 1975 – 2000	Multinomial logit model	[288]
38	Agriculture conversion to forest	Increasing accessibility of agricultural output market; and Application of new technology for agriculture	Location: Northwest Study period: 1954 – 2000	Review and historical data analysis	[292]
39	Agriculture change	Altitude/elevation; and Distance to villages	Location: Cho Don district, Ba Kan province. Study period: 1990 - 2001	Rule-based CLUE-s model	[333]
40	Reforestation	Wood imports;and Policies restricting forest exploitation.	Location: Vietnam-wide Study period: 1987 - 2006	Literature review and analysis.	[334]

Continued on next page

Table A.6 – continued from previous page

Study#	LULCC	Key results (Drivers)	Details	Methods	Reference
41	Deforestation	Floods; Shifting cultivation of agriculture; Forest transitions; Policies; Slope, logging and fire.	Location: four villages in Cho Don district, Bac Kan province. Study period: 1975 - 2007	Statistical analysis of feedbacks from field surveys, including interviews, group discussions, and secondary sources.	[78]
42	Reforestation	Policies; and Plantation expansion.	Location: three villages, including Dong Cao, Dong Dau, and Que Vai, in Tien Xuan commune, Luong Son district, Hoa Binh province. Study period: 1960s - 2005	The Institutional Analysis and Development framework based on focus group, key informant, and household interviews.	[267]
43	Deforestation (Wetland loss)	Economic development policies; Population growth and urbanization, resettlement or infrastructure development; Agriculture expansion; and Forest exploitation.	Location: Mekong delta Study period: 1816 - 2015	Case study analysis framework.	[335]

Continued on next page

Table A.6 – continued from previous page

Study#	LULCC	Key results (Drivers)	Details	Methods	Reference
44	Land transitions	Culture; Market economy; Accessibility and infrastructure; Distance to roads and settlement; and Elevation, slope, and soil type.	Location: Suoi Muoi river in Son La province. Study period: 1973 - 2008	Multiple logistic regression; and Multiple corresponding analysis on household and key informant interviews.	[222]
45	Reforestation	Elevation; and Labour availability.	Location: Muong Do commune in Phu Yen district, Son La province. Study period: 2000 – 2013.	Regression model analysis from interviews	[223]
46	Agriculture expansion	Population growth; Poverty; and Infrastructure development.	Location: Mekong delta; Study period: 1990 - 2005		[336]
47	Deforestation	Agriculture expansion (food production); Population increase; Low food production; and Market of export and import wood products.	Location: regional study covering Vietnam. Study period: 1980 – 2010.	Multiple correlation analysis from field survey data.	[281]

Table A.7: Description and characteristics of the biophysical and socio-economic drivers.

Class	Name	Main drivers	Unit	Tem.Res	Spa.Res.	Code and variable metric		Source
C1	Climate	Precipitation	mm/y	Yearly (1990–2020)	1 km	B1	Mean, RoC, and SD of annual values	Source 1
						B2		
						B3		
		Temperature	oC			B4		
						B5		
						B6		
		Water vapour	kPa			B7		
						B8		
						B9		
C2	Terrain and soil	Elevation	m	Constant	0.09 km	B10	Mean of the AOI	Source 2
		Slope	rad			B11		
		Aspect	rad			B12		
		Topographic Position	index			B13		
		Terrain Ruggedness	index			B14		
		Soil types	class		1 km	B15		Source 3
C3	Demographic data	Population density	pp/km^2	Yearly (2000–2019)	1 km	S16	Mean, RoC, and SD of annual values	Source 4
						S17		
						S18		
		Urban change	%	Yearly (1990–2020)	100 m	S19		
						S20		
						S21		

Continued on next page

Table A.7 – continued from previous page

Class	Name	Main drivers	Unit	Tem.Res	Spa.Res.	Code and variable metric		Source
		Built-settlement growth		Yearly (2000–2020)		S22		
						S23		
						S24		
		Net-migration rate		Yearly (2005-2019)		S25	SD of annual values	Source 5
		Percentage of graduates		Yearly (2001-2017)	Commune	S26		
C4	Accessibility	To main rivers	m	Constant	100 m	B27	Mean of the AOI	Source 6
		To main transportation				S28		
		To the sea				B29		
		To drainage				S30		Source 3
		To embankment of water bodies				S31		
		To irrigation systems				S32		
		To canal				S33		
C5	Agriculture and forestry	Paddy yield	kg/km^2	Yearly (1995-2019)	Commune	S34	SD of annual values	Source 5
		Paddy production				S35		
		Cereal production				S36		
		Fuel wood (logging and charcoal)				S37		

Continued on next page

Table A.7 – continued from previous page

Class	Name	Main drivers	Unit	Tem.Res	Spa.Res.	Code and variable metric		Source
		Agriculture expansion	km^2	Yearly (1990-2019)		S38		
		Shifting agriculture				S39		
		Aquaculture				S40		
C6	Livestock	Poultry	$head/km^2$	Yearly (1990-2019)	Commune	S41	SD of annual values	Source 5
		Cattle				S42		
		Pig				S43		
		Industrial production	index	Yearly (2012-2019)		S44		
C7	Economy	Income	VND/head	Yearly (2006-2018)	Commune	S45	SD of annual values	Source 5
		Income dependency on agriculture				S46		
		Poverty rate	%			S47		
		Market accessibility	index	Constant		S48	Mean of the AOI	

Note

Source 1: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas.
Source 2: The USGS EROS Archive – Digital Elevation – Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global;
Source 3: The Ministry of Agriculture and Rural Development of the Socialist Republic of Vietnam.
Source 4: Gridded Population of the World, Version 3 (GPWv3): Land and Geographic Unit Area Grids; and NASA Socioeconomic Data and Applications Center (SEDAC).
Source 5: The General Statistics Office (GSO) of Viet Nam; https://www.gso.gov.vn/en/about-gso/.
Source 6:The Open Development Vietnam (ODV); https://vietnam.opendevelopmentmekong.net/

Continued on next page

Table A.7 – continued from previous page

Class	Name	Main drivers	Unit	Tem.Res	Spa.Res.	Code and variable metric	Source
Further explanation							
RoC: Rate of change; SD: Standard deviation the of mean;							
The area of interest (AOI) is a boundary area of a community;							
Soil was class based on chemical, depth, drainage, fertility and texture;							

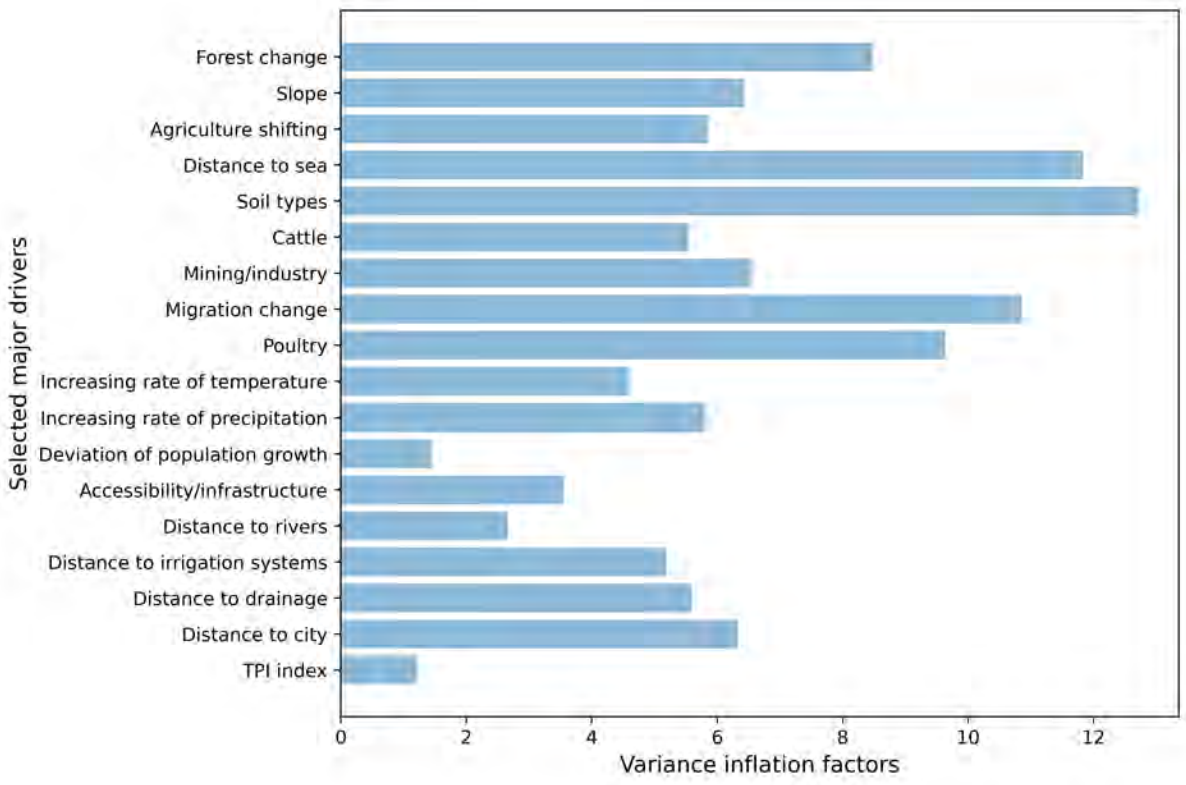


Figure A.1: The measurement of variance inflation factors (VIF) of driver proxies. This figure presents selected driver proxies. The measurement of the VIF is conducted by using Python Version 3.9.10 in Ubuntu Version 20.04.4.

Table A.8: Results from the multiple least square regression model: analyses of the forest gain. The model is conducted by using the main statsmodels API (statsmodels.api) with support of Python Version 3.9.10 in Ubuntu Version 20.04.4.

Log-Likelihood:	177.42		R-squared:	0.901		
AIC:	-318.8		Adj. R-squared:	0.895		
BIC:	-252.7		F-statistic:	146.8		
Df Model:	17		Prob (F-statistic):	2.05E-126		
Drivers	**Coefficients**	**Std err**	**t**	**P> | t |**	**[0.025**	**0.975]**
Constant	0.2268	0.008	28.494	0	0.211	0.242
TPI index	0.1691	0.004	44.294	0	0.162	0.177
Distance to city	0.0392	0.005	7.141	0	0.028	0.05
Distance to drainage	0.0805	0.006	13.394	0	0.069	0.092
Distance to irrigation systems	-0.0618	0.006	-9.697	0	-0.074	-0.049
Distance to rivers	-0.039	0.007	-5.266	0	-0.054	-0.024
Accessibility/infrastructure	-0.0117	0.008	-1.5	0.135	-0.027	0.004
Deviation of population growth	0.0439	0.008	5.387	0	0.028	0.06
Increasing rate of precipitation	-0.0275	0.008	-3.24	0.001	-0.044	-0.011
Increasing rate of temperature	-0.0014	0.01	-0.136	0.892	-0.021	0.018
Poultry	0.0382	0.011	3.632	0	0.018	0.059
Migration change	-0.0397	0.011	-3.556	0	-0.062	-0.018
Mining/industry	0.0189	0.013	1.473	0.142	-0.006	0.044
Cattle	-0.0423	0.014	-3.122	0.002	-0.069	-0.016
Soil types	0.0465	0.014	3.211	0.001	0.018	0.075
Distance to sea	-0.0666	0.017	-3.89	0	-0.1	-0.033
Agriculture shifting	0.1777	0.021	8.431	0	0.136	0.219
Slope	0.0532	0.024	2.179	0.03	0.005	0.101

Table A.9: Results from the multiple least square regression model: analyses of the forest loss. The model is conducted by using the main statsmodels API (statsmodels.api) with support of Python Version 3.9.10 in Ubuntu Version 20.04.4.

Log-Likelihood:	114.89		R-squared:	0.854		
AIC:	-193.8		Adj. R-squared:	0.845		
BIC:	-126.9		F-statistic:	98.09		
Df Model:	17		Prob (F-statistic):	3.12E-108		
Drivers	**Coefficients**	**Std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Constant	0.2508	0.01	25.568	0	0.232	0.27
TPI index	0.1837	0.005	36.44	0	0.174	0.194
Distance to city	-0.0101	0.007	-1.528	0.128	-0.023	0.003
Distance to drainage	0.0848	0.008	11.303	0	0.07	0.1
Distance to irrigation systems	-0.0684	0.008	-8.401	0	-0.084	-0.052
Distance to rivers	-0.0508	0.009	-5.776	0	-0.068	-0.034
Accessibility/infrastructure	0.0253	0.009	2.714	0.007	0.007	0.044
Deviation of population growth	-0.0367	0.01	-3.773	0	-0.056	-0.018
Increasing rate of precipitation	0.0532	0.011	4.909	0	0.032	0.075
Increasing rate of temperature	0.0069	0.011	0.632	0.528	-0.015	0.028
Poultry	-0.0005	0.012	-0.036	0.971	-0.025	0.024
Migration change	0.0061	0.013	0.454	0.65	-0.02	0.032
Mining/industry	-0.0562	0.015	-3.757	0	-0.086	-0.027
Cattle	0.0299	0.016	1.84	0.067	-0.002	0.062
Soil types	0.0341	0.017	2.023	0.044	0.001	0.067
Distance to sea	-0.0044	0.019	-0.228	0.82	-0.043	0.034
Agriculture shifting	0.1443	0.026	5.596	0	0.094	0.195
Slope	0.066	0.026	2.543	0.012	0.015	0.117