

**DETAILED LAND COVER MAPPING IN A SEASONALLY DRY TROPICAL FOREST
LANDSCAPE USING MULTIPLE SENSOR TYPES**

by

Adrian Dwiputra

B.Sc., Institut Teknologi Bandung, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Geography)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

September 2021

© Adrian Dwiputra, 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

DETAILED LAND COVER MAPPING IN A SEASONALLY DRY TROPICAL FOREST
LANDSCAPE USING MULTIPLE SENSOR TYPES

submitted by Adrian Dwiputra in partial fulfillment of the requirements for

the degree of Master of Science

in Geography

Examining Committee:

Dr. Naomi Schwartz, Assistant Professor, Department of Geography, UBC

Supervisor

Dr. Claire Kremen, Professor, Department of Zoology, UBC

Supervisory Committee Member

Dr. Brian Klinkenberg, Professor, Department of Geography, UBC

Supervisory Committee Member

Abstract

Detailed mapping of land cover is essential for supporting science-based sustainable landscape management. Despite the importance of land cover mapping in monitoring landscapes dynamics, land cover data are not always available. Even when the land cover data are available, they often lack detailed discrimination between forest types and plantations. This issue was found in a seasonally dry tropical forest landscape in Siem Reap and Preah Vihear, Cambodia. In this thesis, I explored the potential of (1) the fusion of optical and radar data in developing detailed land cover maps and revealing the driver of landscape change, and (2) vertical vegetation structure acquired by the Global Ecosystem Dynamics Investigation (GEDI) mission—a new mission that harnesses a space-borne waveform lidar sensor installed on the International Space Station—to improve the vegetation mapping in the studied landscape. The fusion between radar (Sentinel-1) and optical (Sentinel-2) satellite data slightly improved the land cover classification accuracy (1.6% overall accuracy increase) relative to Sentinel-2-only classification. Between 2015 and 2019, I detected a 247,781.04 Ha dry deciduous forest loss; most were due to logging (147,314 Ha). Land designations, such as the protected areas and the economic land concessions (ELCs), significantly determine land cover change. The classification of vegetation types using GEDI data had 81.9% overall accuracy despite the limited spatial coverage of GEDI data. The GEDI-only classification results could identify the seasonally inundated forests with better accuracy than the land cover map derived from the fusion of optical-radar data. These results demonstrate the potential of structural information acquired by Sentinel-1 and GEDI to improve our ability to identify vegetation types in complex, heterogeneous landscapes.

Lay Summary

Monitoring land cover is essential in the face of global environmental change. However, data gap is often an issue that hinders land cover monitoring from being put in place. In the first part, I used the combination of two different satellites: Sentinel-2 (optical) and Sentinel-1 (radar), to create land cover maps for 2015 and 2019. Map's accuracy slightly increased following the addition of Sentinel-1 data into land cover mapping. The analysis of 2015 and 2019 land cover maps showed that logging in the dry deciduous forest was prevalent. In the second part, I explored the potential of vegetation structure data collected by the new Global Ecosystem Dynamics Investigation (GEDI) mission in improving the identification of different vegetation types in the studied landscape. These results demonstrate the potential of structural information acquired by Sentinel-1 and GEDI to improve our ability to identify vegetation types in complex, heterogeneous landscapes.

Preface

The results presented here are the author's original works conducted during the two-year Master's study period in 2019-2021. The first research chapter, i.e. "Chapter 2. Drivers of Land Cover Change in Seasonally Dry Tropical Forest Landscape-A Case Study from Siem Reap and Preah Vihear", is an original, unpublished work designed and conducted by the author under the supervision of Dr. Naomi Schwartz. Professor Claire Kremen and Professor Brian Klinkenberg provided feedback that guided the author to refine the manuscript. The second chapter, "Chapter 3. GEDI Waveform Metrics in Vegetation Mapping – A Case Study from A Complex Tropical Forest Landscape", originated from the author's work under the supervision of Dr. Nicholas Coops and Dr. Naomi Schwartz. Some of the results of the second chapter had been presented as:

Dwiputra, Adrian, Nicholas C. Coops, and Naomi Schwartz. 2020. "Utilising Space-borne 3D Vegetation Metrics to Improve Land Cover Mapping: A Case Study from A Complex Tropical Landscape." *AGU Fall Meeting 2020 Abstracts*

Table of Contents

Abstract	iii
Lay Summary	iv
Preface.....	v
Table of Contents	vi
List of Tables	ix
List of Figures	x
Acknowledgements	xi
Dedication	xii
Chapter 1. Introduction	1
Chapter 2. Drivers of Land Cover Change in Seasonally Dry Tropical Forest Landscape-A Case Study from Siem Reap and Preah Vihear	5
2.1. Introduction	5
2.2. Objectives	7
2.3. Study Area.....	7
2.4. Data and Methods.....	7
2.4.1. Data	7
2.4.2. Land Cover Mapping	11
2.4.3. Land Cover Change Analysis	14
2.5. Results	16

2.5.1.	Land Cover Mapping	16
2.5.2.	Land Cover Change Analysis	18
2.6.	Discussion	22
2.6.1.	Comparison with other land cover mapping efforts	22
2.6.2.	Predictor variables' contribution.....	24
2.6.3.	Land cover change analysis	24
2.6.4.	Uncertainties and limitations	26
2.6.5.	Management relevance	27
2.7.	Conclusion.....	28
Chapter 3. GEDI Waveform Metrics in Vegetation Mapping – A Case Study from A Complex Tropical Forest Landscape.....		29
3.1.	Introduction	29
3.2.	Data and Method	30
3.2.1.	Study Area	30
3.2.2.	GEDI Dataset	31
3.2.3.	Waveform Metrics	31
3.2.4.	Vegetation Classification	33
3.2.5.	Comparison with Land Cover Data	34
3.3.	Results	35
3.3.1.	Vegetation classification using GEDI metrics	35

3.3.2. Comparison between GEDI-only classification and optical-radar classification ...	36
3.4. Discussion	37
3.5. Conclusion.....	39
Chapter 4. Conclusion.....	40
Bibliography	42
Appendix.....	52
A.1. A list of Sentinel scenes for compositing in land cover classification.....	52
A.2. Formulas of the spectral indices used in land cover classification	55

List of Tables

Table 2.1 List of variables included in two different variable input sets for land cover classification..	11
Table 2.2 Descriptions of land cover classes of interest and the associated number of reference points.....	12
Table 2.3 Explanatory variables included in the land cover change analysis.....	14
Table 2.4 Land cover classification accuracy obtained from different input variable sets.....	16
Table 2.5 Confusion matrix that contains class-specific accuracy and mapped area confidence interval	17
Table 3.1 Description of the GEDI-derived metrics extracted from GEDI Level 1B and Level 2A collections.	33
Table 3.2. Accuracy matrix of the GEDI-only classification	35

List of Figures

Figure 2.1 “True-color” Sentinel-2 satellite image of the study area from 2019. Siem Reap is the province on the left, while Preah Vihear is on the right.	8
Figure 2.2 Steps implemented in land cover mapping for 2019 and 2015 with nodes and edges that illustrate the processing flow	13
Figure 2.3 Relative importance of variables in the optical-radar fusion approach.....	17
Figure 2.4 Relative area plot showed land cover classes that underwent area increase and decrease in 2015-2019.	18
Figure 2.5 Land cover map of Siem Reap and Preah Vihear in 2019 and some of the detected changes in 2015-2019	19
Figure 2.6 The largest changes in 2015-2019 showed the key role of cropland (paddy, annual crops) expansions in the landscape.	21
Figure 2.7 The grass-dominated dry deciduous forest is prone to tree removal that induces conversion of DDF into grassland.	21
Figure 2.8 Response curves of explanatory variables included in the land cover change analysis.	22
Figure 3.1. Example of GEDI waveform data and the associated vegetation classes..	32
Figure 3.2. Classification result showing the distribution of the vegetation classes in the study area.	35
Figure 3.3. Variable importance as indicated by the Mean Decrease Accuracy and Mean Decrease Gini.	36
Figure 3.4. Comparison of the classification result and 2019 land cover map.....	37

Acknowledgements

The thesis and the rest of my Master's study will not be as successful without the generous help and support from colleagues, mentors, friends, and family. I would like to express my sincere gratitude to

- Naomi Schwartz, my amazing supervisor, without whom the whole study will not be possible in the first place. Thank you for patiently guiding me throughout the learning process all this time.
- Claire Kremen for being an accessible role model in conservation science, a field I am deeply passionate about throughout my academic journey. I am honoured to have you on my supervisory committee. Your feedback and kindness have been an inspiration that made finishing my study in pandemic much bearable.
- Brian Klinkenberg for the valuable feedback and discussions that had made navigating the complex and sometimes poorly documented field of Geographic Information Science (GIS) and remote sensing less daunting. I should mention that the online resources you had put together helped clarify technical aspects related to remote sensing and GIS.
- Nicholas Coops for the fantastic remote sensing course that inspired many important parts of this remote sensing study. It was an honour to collaborate with you in exploring the novel space-borne lidar technology.
- Chloe Lam for the valuable help in the tedious desktop-based reference data collection.
- Tony Yon for being a great surveying companion throughout my adventures in Cambodia
- my friends and family for making the learning experience enjoyable and exciting. I want to deliver special thanks to Kyle Wlodarczyk for being a dear friend who makes the pandemic much less isolating. Elise, Sarah, and Tony, thank you so much for being awesome labmates! Your cool works really inspired and motivated me.

for Yanto Wibawa and Tin Widyani Satriawan, who had not had the chance to meet each other

Chapter 1. Introduction

Tackling the global issues of food security, climate change, and biodiversity crisis requires more sustainable management of landscapes. The growing demand for agricultural products due to the global population growth has been causing an expansion of agricultural land since 1700 despite the increased productivity per unit land associated with more intensified management practices (Ramankutty et al., 2018). At the same time, as a consequence of the growing land demand among other drivers, the conversion of carbon-rich vegetation, such as natural forests, has been widespread across the globe (Hansen et al., 2013; Potapov et al., 2019), with identified hotspots in the tropical forest landscapes (Gibbs et al., 2010). Protecting forests in tropical landscapes can reduce greenhouse gases (GHG) emissions from avoided deforestation and ecosystem degradation and significantly contribute to achieving the 1.5°C above the pre-industrial global temperature goal (Roe et al., 2019). Furthermore, conserving carbon-rich terrestrial ecosystems provides co-benefits for biodiversity since habitats are protected from further loss and degradation. Despite the potential benefits of sustainable tropical forest landscape management, a recent global assessment identified tropical deciduous forest biomes as the largest hotspot of net tree canopy loss at approximately 179,000 km² between 1982 to 2016 (Song et al., 2018).

Effective conservation and sustainable management of tropical forest landscapes require a monitoring system to provide timely and accurate evidence that allows managers to make informed decisions (Portillo-Quintero et al., 2021). Accurate, regularly updated land cover maps depict landscapes' current and historical state; this information is the primary inputs to time-series analysis to unveil the drivers of the landscape changes (Leinenkugel et al., 2015; Tellman et al., 2020). A better understanding of the drivers of change can enable decision-makers to design holistic pathways to control the alarming rate of ecosystem degradation and deforestation in tropical forest landscapes.

Despite the importance of land cover mapping in monitoring tropical landscapes, land cover data are not always available, especially in developing economies. A recent study identified land managers' need for land cover data in the Lower Mekong region (Saah et al., 2019). A regional land cover monitoring system has been developed to address the identified gap by harnessing modern remote sensing technologies (Saah et al., 2020). Unfortunately, such a monitoring system

provides land cover data at a coarse scale (pixel size = 1 km²) and cannot identify different natural forest types.

In this thesis, I explored the potential of new remote sensing technologies, specifically (1) the fusion of radar and optical satellite data and (2) the space-borne lidar data, to accurately map land cover in a seasonally dry (deciduous) tropical forest landscape in Siem Reap and Preah Vihear, Cambodia. The general properties of the remote sensing data used in this research will be briefly described in the following subsections.

a. Optical Remote Sensing

Optical remote sensing systems are the first established remote sensing techniques used in land cover mapping (Anderson *et al.* 1976; Allan 1980). Optical sensors work very similarly to cameras to capture the lights reflected by objects of interest. It belongs to a passive remote sensing category because it relies solely on the sun as the primary light source. The widely-used Landsat and MODIS missions with a medium or coarse spatial resolution, with pixel size 30 and 250 meters, respectively, belong to this group (Chhetri and Thai 2019). This group also includes high-resolution satellites (spatial resolution ≤ 10 m), such as QuickBird, GeoEye, SPOT, and WorldView. With the long mission history, the earliest efforts of land cover monitoring from space were conducted with optical sensors (Allan 1980).

Optical sensors provide a means to observe the earth's surface from space systematically. With the oldest data archive dated back in 1982, Landsat TM allows for long multi-temporal analysis (Chhetri & Thai, 2019). Data collected by a newer satellite, Sentinel-2, can accurately map detailed land cover classes in a complex tropical landscape where diverse vegetation types are close to each other (Nomura & Mitchard, 2018). Multispectral bands used in Landsat TM and Sentinel-2 data can reveal patterns that would be invisible when only visible red, green, blue (RGB) bands are used.

Previous studies have identified several limitations of optical remote sensing. In optical systems, like in many other passive remote sensing systems, the major light source for optical remote sensing systems is the sunlight. Accordingly, only limited features are visible in scenes acquired in the nighttime. In places with frequent and dense cloud cover, the optical sensors captured clouds

instead of the reflectance of the land surface beneath them (Chhetri and Thai 2019; Nomura and Mitchard 2018; Leinenkugel *et al.* 2013). This limitation is due to the inherent properties of shorter wavelengths that are harder to ‘penetrate’ through the clouds. With such a limitation, it has been challenging to make dense temporal series land cover change analyses because of the limited availability of cloud-free scenes. A past effort in the tropical region of Southeast Asia required 11 years of mosaics to generate a single vegetation map on the regional level (Leinenkugel *et al.*, 2013). Even though the newer optical earth observation satellites are still unable to “see” through clouds, the higher observation frequency of the new satellites allows images with less cloud cover to be “stitched” together to form a cloud-free composite.

b. Radar Remote Sensing

As an active system, radar remote sensing involves transmissions of microwaves onto the earth’s surface and record the echoes that bounce back from the earth’s surface (Sabins, 1997). Unlike passive optical sensors that rely on the light source availability, radar data acquisitions are independent of the sunlight. The microwaves used in such a system have longer wavelengths (1 mm-30 cm) than the spectrum captured by optical systems. Consequently, radar pulses can penetrate through smoke, cloud, dust, snow, and rain. Its sensitivity to surface roughness, dielectric properties of objects, i.e. the sensitivity of materials to be polarized in response to an electric field, allows for patterns invisible to the optical sensor to be revealed.

Radar systems also have some drawbacks in monitoring land cover change. The sensitivity of microwaves to the objects’ water content sometimes causes the pulse to penetrate the object of interest instead of being bounced back. When this happens, it leads to the failure of the data acquisition in capturing the desired object. Therefore, radar sensors are prone to variations due to seasonal dryness/wetness of the area of interest and were found to perform better in wetter periods (Mendes *et al.*, 2019). In terms of temporal coverage, radar systems are inferior to optical systems, which generally have older mission history. Nevertheless, the Sentinel-1 mission has been collecting global data since 2014 and thus, allows for monitoring of recent dynamics of landscapes.

Despite the different limitations of radar and optical systems, an increasing number of studies show the potential of fusion between optical and radar data to improve land cover mapping (Joshi *et al.*, 2016; Mendes *et al.*, 2019).

c. Lidar Remote Sensing

One of the most cutting-edge advancements in remote sensing is lidar (light detection and ranging) imaging. Lidar can accurately map vegetation structures using light-transmitting devices paired with light detector sensors on airborne vehicles or satellites (Dubayah et al., 2020). Until recently, lidar missions are often airborne, and thus, have limited coverage area (Singh et al., 2019). Nevertheless, previous studies have shown that lidar data are helpful in identifying different forest types, either by itself (Marselis et al., 2018) or in combination with optical images (Fagan et al., 2018).

A new NASA remote sensing mission, GEDI (Global Ecosystem Dynamics Investigation), has recently been launched to provide lidar data to better understand vegetation dynamics worldwide (Potapov *et al.*, 2019; Marselis *et al.* 2018). In the two-year mission that started in 2019, the GEDI instrument has been collecting complex full-waveform data from the international space station. The data contain information that captures the complexity of vegetation structures. Since January 2020, the datasets have been made publicly accessible.

The studies presented in this thesis were designed to provide a case study that assesses the potential of newer remote sensing data in addressing the data gap in the biodiverse landscape of Siem Reap and Preah Vihear province, Cambodia. Despite the technical application of the three remote sensing systems described above as the thesis' focus, I included some discussions that draw some contextual information from literature about land governance in the study area to provide relevant background.

Chapter 2. Drivers of Land Cover Change in Seasonally Dry Tropical Forest Landscape-A Case Study from Siem Reap and Preah Vihear

2.1. Introduction

Rapid land cover change in tropical landscapes, especially in biodiversity hotspots, gives rise to environmental concerns (Chaplin-Kramer et al., 2015; Fitzherbert et al., 2008; Sánchez-Cuervo et al., 2020; Wohlfart et al., 2014). These biodiversity hotspots are home to diverse species of flora and fauna, many of which are threatened with local or total extinction due to habitat loss and fragmentation following rapid expansions of agricultural land. The expansion of agricultural land in the tropics has been occurring for decades (Gibbs et al., 2010; Ramankutty et al., 2018). At the same time, global tree cover loss was concentrated on tropical deciduous forest biomes (Song et al., 2018). Even under an optimistic IPCC's future climate scenario B2 that assumes a controlled increase of greenhouse gas emissions in the future, land cover change can completely eradicate suitable habitats for endangered bird species from Southeast Asia's protected areas (Singh, 2020). Without a significant change in the current trend, we can expect more greenhouse gas emitted into the atmosphere and more species to extinct in the near future.

Land cover monitoring is an essential tool to support the evidence-based sustainable management of dynamic tropical landscapes. Land cover monitoring systems have been developed to detect land cover changes at the global and regional scales (Hansen et al., 2013; Potapov et al., 2019). With such systems, policymakers can access empirical information that is regularly updated and, therefore, can make better decisions on landscape management issues. Land cover change monitoring can provide details about the underlying drivers of the observed change based on the typology and pattern of the change (Bey et al., 2020; Meyfroidt et al., 2014). Knowledge of land cover change drivers helps policymakers in devising change towards sustainable landscape management.

Modern land cover monitoring systems have been built on the advancement of remote sensing technologies. Remote sensing has been used in earth observation for decades; in the early days of remote sensing, scientists analyzed photos taken by optical cameras mounted on planes and satellites (Allan, 1980; Sabins, 1997). As the sensor technologies advance, various sensors, such

as radar and sophisticated optical sensors with multi-spectral, and hyperspectral bands, are onboard many earth observation satellites (Sabins, 1997). The vast array of sensors allows us to harness different perspectives that unveil the patterns of landscape dynamics worldwide in ever-improving detail.

Monitoring the landscape dynamics in the region using solely optical sensors faces many challenges. One of the most common issues is the persistent cloud cover that shrouds the area of interest (Leinenkugel et al., 2013). Even though cloud cover is less frequent in the drier parts of the tropics, smoke from fires often covers parts of the image. A new land cover monitoring system developed mainly using optical images failed to identify different natural forest types in the lower Mekong region (Saah et al., 2020). Since optical sensors are more sensitive to spectral than structural differences of different land cover classes, combining an optical sensor with other sensors that better delimit structural differences and are less sensitive to cloud cover will improve our land cover monitoring accuracy.

A fusion of optical and radar sensors can improve our ability to monitor tropical landscape dynamics over time (De Alban et al., 2018; Poortinga et al., 2019; Shimada et al., 2014). Radar remote sensing uses the microwave that allows observations through cloud and smoke cover (Sabins, 1997). In addition, radar sensor increases our ability to tell apart land cover classes with similar spectral properties but distinct structural properties (Joshi et al., 2016). Land cover mapping that harnesses a fusion of radar and optical sensors can improve our ability to tell apart different vegetation types that appear similar but bear distinct ecological and economic significance. In turn, these more accurate and detailed land cover maps can reveal change patterns that landscape managers otherwise overlook.

This study took place in a seasonally dry tropical forest landscape in northern Cambodia. Home to critically endangered species like the giant ibis (*Thaumatibis gigantea*), which is threatened by habitat conversion associated with agricultural land expansions (Loveridge & Srun, 2015), the landscape is dominated by mosaics of natural habitats and cultivated lands. The presence of smallholder farms and the large company-managed estates, which coexist with the protected areas, would provide an interesting case study to see the effect of land designations and other potential drivers, such as previous deforestation and accessibility, on the land cover dynamics.

2.2. Objectives

1. Evaluate the accuracy of land cover mapping with a fusion of Sentinel-2 (optical) and Sentinel-1 (radar) data
2. Identify the drivers of recent land cover change in the seasonally dry tropical forest landscape in Siem Reap and Preah Vihear province

2.3. Study Area

The studied landscape is a 24,087 km² seasonally dry tropical landscape located in Siem Reap and Preah Vihear province in the northern part of Cambodia. In a preliminary survey, I found that diverse land cover classes—such as evergreen forest, dry deciduous forest, as well as commodity plantations like mango, cassava, cashew, rice field—are present as mosaics close to each other and, thus, increase the challenge to accurately mapping them. Few studies had documented LULC in the area, all of them had a regional focus, i.e., the Lower Mekong Basin (Leinenkugel et al., 2015; Potapov et al., 2019; Saah et al., 2019). Home to several protected areas, the study area also hosts economic land concessions (ELC) that cover a total area of 2,384.399 km² (Open Development Cambodia, 2020). It was worth noting that some ELC areas overlapped with the protected area. The temporal scope of the study was the year 2015 and 2019; the former was the earliest year when Sentinel-2 data were available, while the latter was chosen due to the availability of reference data collected in a field ground-truthing.

2.4. Data and Methods

2.4.1. Data

2.4.1.1. Satellite data

I analyzed optical and radar data acquired by Sentinel-2 and Sentinel-1 satellites, respectively. These data were made publicly accessible by the European Space Agency. Details on the Sentinel-2 and Sentinel-1 data and the processing steps that were applied before the land cover classification are described in the following section. Except for the ones mentioned otherwise, all steps were run in Google Earth Engine (Gorelick et al., 2017). Originally varies in their spatial resolutions, all input layers were resampled into rasters with a cell size of 10-meters, the highest original spatial

resolution of some of the data.

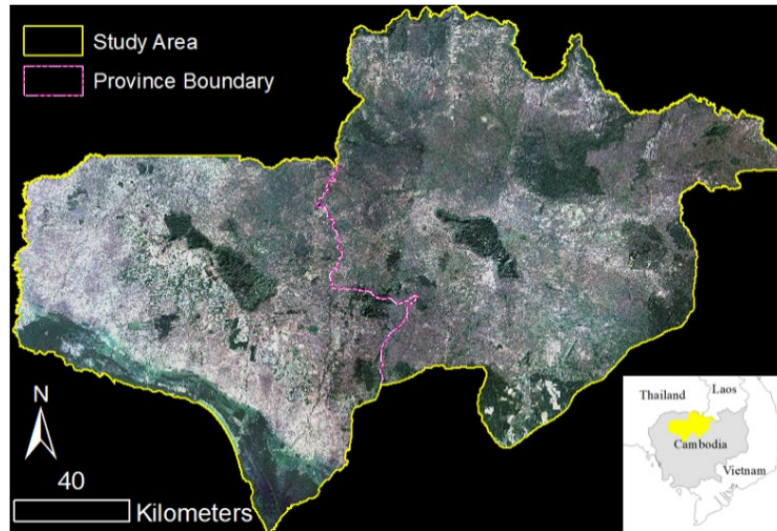


Figure 2.1 “True-color” Sentinel-2 satellite image of the study area from 2019. Siem Reap is the province on the left, while Preah Vihear is on the right.

a. Sentinel-2

I used Sentinel-2 data acquired between June 2015-March 2016 and June 2019-March 2020 as inputs for the land cover classification. For 2019, the data provider had implemented preprocessing steps comprising radiometric and terrain correction to the images published as Level-2A Sentinel-2 data. To obtain the same processing level for 2015 data, I ran the correction steps on Level 1C Sentinel-2 data using Sen2Cor software (Main-Knorn et al., 2017) version 2.8. I accepted all processing parameters as defaults to match the processing parameters used by the data providers. Sentinel-2 Level-2A contains surface reflectance measurements recorded at 11 bands, sensitive to different wavelengths of electromagnetic radiation.

I included 10 of the original Sentinel-2 bands and additional 13 layers that contain several spectral indices calculated using different sets of the Sentinel-2 band (Table 2.1). The original resolution of these layers was either 10 or 20 meters before I resampled them into 10 meters to maintain consistency across all inputs. While most layers were derived from the median composite of Sentinel-2 scenes acquired in December 2015 and 2019 (see Appendix A.1), two of them, the 10th and 90th percentile recorded NDVI, were extracted from an extensive set of Sentinel-2 scenes acquired between June 2015-March 2016 and June 2019-March 2020 for 2015 and 2019 land cover

maps, respectively. Median composite is a simple statistical approach to obtain the best available pixel value based on the median of the recorded values at each pixel within the period of interest. The application of median compositing removes pixels with extreme values associated with observation error or interference by unwanted objects, such as smoke or cloud. To reduce the likelihood of getting abnormal values in the median composite, clouds were masked out using *s2cloudless* (Sentinel Hub, 2021), with the maximum cloud probability parameter set to 12%. Shadows were also removed based on the near-infrared and shortwave infrared measurements at a given pixel (Poortinga et al., 2019). Two layers were not derived from the median composite of scenes acquired in December: *_p10* and *NDVI_p90*. Instead, I calculated them from a collection of cloud-masked Sentinel-2 Level-2A scenes acquired in June - March; *_p10* denotes the 10th percentile, while *_p90* denotes the 90th percentile. The resulting 23 layers are included in the optical-only and optical-radar variable sets, which were used as the land cover classification input (Table 2.1).

b. Sentinel-1

I included three radar layers and the 23 optical layers in the optical-radar set of inputs for the land cover classification (Table 2.1). All of the radar layers were created using Sentinel-1 Ground Range Detected (GRD) data. Sentinel-1 GRD data was collected using a synthetic-aperture radar (SAR) instrument that utilized C-band microwaves ($\lambda = 5.548$ cm). The radar images contain backscatter values, a coefficient that indicates the strength of the radar signal reflected back by the surface on the earth's surface. The backscatter coefficient is sensitive to the reflecting object's moisture, surface texture, and orientation (Sabins, 1997). The unique properties of radar data are expected to complement the optical information provided by Sentinel-2 derived layers in the land cover classification.

I implemented two approaches in preparing the radar inputs: with terrain flattening and without terrain flattening. Terrain-flattened backscatter data (gamma-naught, γ^0) have lower variations across two or more adjacent paths, and hence, have better consistency than the terrain-corrected sigma-naught (σ^0) data over a wide area (Small et al., 2021). Accordingly, the variable *VV* was created using the mosaic of γ^0 Sentinel-1 scenes acquired in December 2019 for the 2019 land cover input and in December 2015 for the 2015 land cover (see Appendix A.1). I run preprocessing

steps to create the terrain-flattened data using the Sentinel-1 toolbox in SNAP software version 8 (ESA, 2020). Two other radar input layers (*VV_p10* and *VV_p90*) were created using a collection of Sentinel-1 GRD σ^0 scenes acquired in June - March 2015 - 2016 and 2019 - 2020 that the data provider preprocessed.

2.4.1.2. Reference data

Reference data provided the basic information that defines the land cover class in the mapping processes. The reference data indicate the land cover classes found in the studied landscape and some of their known locations in the particular year mapped. Two reference data sources were field visits to the study area in 2019 and a desktop survey conducted in 2020. A detailed list that summarizes the source of the reference data, i.e. image sources and the associated acquisition dates, is available upon a reasonable request addressed to the Author.

The reference data contained 4,815 points collected from the ground survey in 2019 and 7,947 from the desktop survey. Ground survey points were collected opportunistically along roads to respect private properties access restrictions. The presence of landmines in the landscape caused safety concerns that added to the challenge of the probabilistic sampling implementation during the field survey. The desktop survey points were distributed randomly in the study area to enhance the spatial balance of the reference data. Two surveyors assessed the randomly distributed points using Collect Earth software (Bey et al., 2016) and assigned appropriate labels that indicate the identified land cover classes (Table 2.2). I identified and omitted outliers in the desktop survey results based on high-resolution global tree height data in 2019 (Potapov et al., 2021).

I separated the reference data into two groups, each serving a different purpose in the land cover classification. The first group that contained 9,352 (73.28%) reference points was the training data input to the classification. The other group ($n = 3,410$; 26.72%) was used in the accuracy assessment I conducted following the map production.

Table 2.1 List of variables included in two different variable input sets for land cover classification. The formulas for the spectral indices are provided in Appendix A.2.

Layers	Source	Description	optical-only	optical-radar
blue_median	S-2	band 2 of Sentinel-2 Level-2A median composite	o	o
green_median	S-2	band 3 of Sentinel-2 Level-2A median composite	o	o
red_median	S-2	band 4 of Sentinel-2 Level-2A median composite	o	o
red1_median	S-2	band 5 of Sentinel-2 Level-2A median composite	o	o
red2_median	S-2	band 6 of Sentinel-2 Level-2A median composite	o	o
red3_median	S-2	band 7 of Sentinel-2 Level-2A median composite	o	o
nir_median	S-2	band 8 of Sentinel-2 Level-2A median composite	o	o
swir1_median	S-2	band 11 of Sentinel-2 Level-2A median composite	o	o
swir2_median	S-2	band 12 of Sentinel-2 Level-2A median composite	o	o
NDVI	S-2 calc.	Normalized Difference Vegetation Index	o	o
SAVI	S-2 calc.	Soil Adjusted Vegetation Index	o	o
NDWI	S-2 calc.	Normalized Difference Water Index 1	o	o
NDWIt	S-2 calc.	Normalized Difference Water Index 2	o	o
NBR	S-2 calc.	Normalized Burn Ratio	o	o
EVI	S-2 calc.	Enhanced Vegetation Index	o	o
GNDVI	S-2 calc.	Green Normalized Difference Vegetation Index	o	o
EVI2	S-2 calc.	Enhanced Vegetation Index 2	o	o
MSI	S-2 calc.	Moisture Stress Index	o	o
MCARI	S-2 calc.	Modified Chlorophyll Absorption in Reflectance Index	o	o
PSSR	S-2 calc.	Pigment Specific Simple Ratio	o	o
NDVItx	S-2 calc.	focal statistics of 5x5 cells kernel applied on NDVI layer	o	o
NDVI_p10	S-2 calc.	The 10th percentile of recorded NDVI values in June-March	o	o
NDVI_p90	S-2 calc.	The 90th percentile of recorded NDVI values in June-March	o	o
VV	S-1	terrain flattened radar backscatter in December	-	o
VV_p10	S-1 calc.	The 10th percentile of recorded radar backscatter values in June-March	-	o
VV_p90	S-1 calc.	The 90th percentile of recorded radar backscatter values in June-March	-	o
S-2 calc. : calculated based on Sentinel-2 bands			o : included	
S-1 calc. : calculated based on Sentinel-1 bands			- : omitted	

2.4.2. Land Cover Mapping

Land cover mapping was the first main component of the analysis. I developed land cover maps for 2015 and 2019 as inputs to the land cover change modelling in the second part. I ran a supervised classification with Random Forest (Breiman, 2001) to map the land cover classes of interest (Table 2.2). These classes were the dominant land cover classes in the landscape that bear economic and ecological significance, i.e., important commodity plantation and natural ecosystem cover. I conducted all land cover classification processes in Google Earth Engine because of its massive computing capacity (Gorelick et al., 2017). The flowchart in Figure 2.2 summarizes the

overall land cover mapping framework.

Table 2.2 Descriptions of land cover classes of interest and the associated number of reference points

Land cover class	Description	n Reference
Other tree covers	Tree dominated land cover (tree cover > 30%) that includes pulp and paper plantation, agroforestry, and others that do not belong to any of the other class defined	221
Paddy	Paddy fields, either irrigated or dry	3238
Barren land and built	Unvegetated lands, including concretes, building, and rocks	261
Evergreen forest	Non-deciduous forest dominated by evergreen tree species	2303
Anacardiaceae plantation	Mature (≥ 2 m) mango or cashew plantation	720
Annual or herbaceous crops	Cultivated lands dominated by annual crops other than rice, e.g., cassava, sugar cane; includes non mature tree	1274
Dry dipterocarp forest	Deciduous forest with open canopy cover ($\geq 25\%$) and grass-dominated understory	1578
Grassland	Grass or shrub dominated lands; including wetlands that meet the criteria	1262
Seasonally inundated forest	Forest dominated by trees that can withstand seasonal flooding	241
Water bodies	Water bodies: rivers, lakes, water reservoirs, ponds	772
Rubber plantation	Mature rubber tree (≥ 5 m) plantation	892

I produced two 2019 land cover maps with a pixel size of 20 m, using two approaches, each of which uses a different set of input variables. The first approach used 23 optical-only variables, while the second used 26 variables that included three additional variables derived from Sentinel-1 radar data. The two approaches used the same training data and yielded two Random Forest classification results. While the input satellite data comprised raster layers with a pixel size of 10 m, the land cover maps have a spatial resolution of 20 m for computation efficiency. I applied morphological filters that consisted of a majority filter followed by a dilation-erosion-dilation filter to remove “salt and pepper” noise that is common in pixel-based classifications (Kavitha, Srikrishna, and Satyanarayana, 2021), as well as in radar data. The accuracy of the results, both unfiltered and filtered, was evaluated using the independent validation data. I conducted accuracy assessments according to the good practice described by Olofsson *et al.*, 2014 that considers the land cover area proportion in the study area and the associated uncertainties in area estimates. The

assessment was implemented in R version 4.0.5 (R Core Team, 2021). I took into account the area-adjusted accuracy assessment results and the out-of-bag error estimate produced earlier in the Random Forest classifier training when comparing the two different approaches.

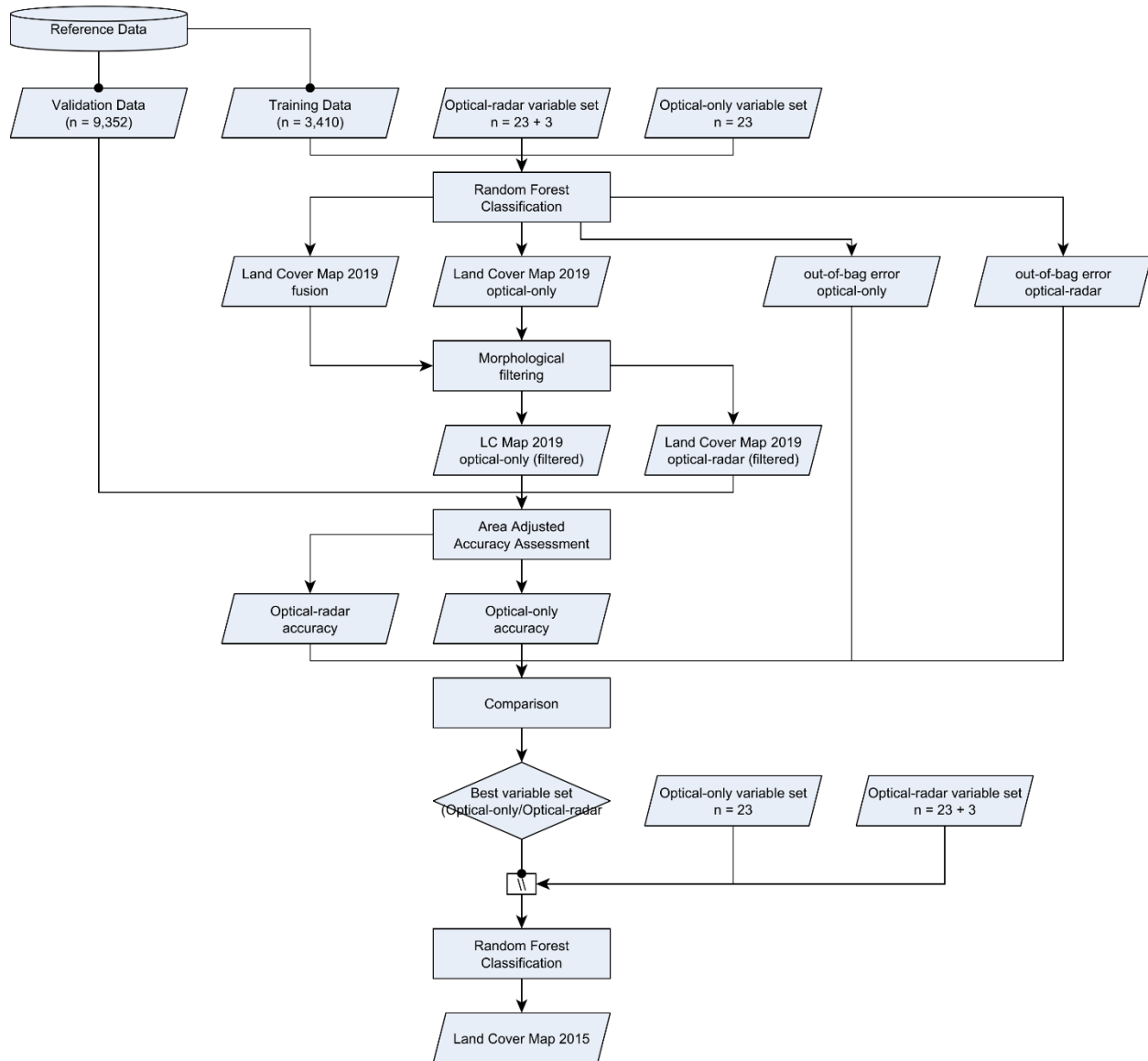


Figure 2.2 Steps implemented in land cover mapping for 2019 and 2015 with nodes and edges that illustrate the processing flow

Following the accuracy assessment, I replicated the approach with the highest accuracy to produce

the 2015 land cover map. I input the same variables as the set that resulted in the most accurate 2019 land cover map—according to the accuracy assessment step—into the previously trained Random Forest classifier; for the 2015 land cover map, predictor variables were derived from satellite data acquired in 2015 - 2016 instead. If the morphological filtering increases the 2019 land cover map accuracy, the same filters were to be applied to the 2015 land cover map after the Random Forest classification.

2.4.3. Land Cover Change Analysis

I identified land cover changes that occurred in 2015 - 2019 based on the land cover maps produced in the previous steps. The first step in land cover change analysis was to remove detected changes that seemed improbable. I overlaid the land cover rasters from 2015 and 2019 and reclassified the pixels in 2015 land cover that were involved in changes that were very unlikely to occur in 4 years, such as transformations into evergreen, seasonally inundated, or deciduous dipterocarp forest from intensively managed croplands. The corrections assumed that these improbable changes were due to misclassifications in 2015 land cover, while no actual change occurred in 2015 – 2019 in those locations. In total, 2,080,575 out of 61,441,010 pixels (3.38%) in 2015 land cover map were reclassified in this manner.

Table 2.3 Explanatory variables included in the land cover change analysis

Variable	Source	Category	Description
Accessibility to cities in 2015	Weiss <i>et al.</i> , 2018	Anthropogenic	Estimated travel time from each pixel to reach the nearest population center (in minutes) resampled from 1 km pixels
Proximity to the previous deforestation	Potapov <i>et al.</i> , 2019	Anthropogenic	Euclidean distance of each pixel to the nearest area that had undergone significant tree canopy loss (> 70%) resampled from 30 pixels
Distance to the protected area border	Open Development Cambodia, 2020	Anthropogenic	Euclidean distance of each pixel to the protected area border; multiplied by -1 for pixels inside the border
Distance to the economic land concession border	Open Development Cambodia, 2020	Anthropogenic	Euclidean distance of each pixel to the economic land concession border; multiplied by -1 for pixels inside the border
Soil type	Open Development Cambodia, 2020	Biophysical	Identified dominant soil class in each pixel

I included five variables to explain the observed change in the landscape: proximity to deforestation in the previous period 2010 - 2015, accessibility, soil types, distance to the protected area border, and distance to the economic land concession border. Distances to borders range from negative values for the area inside the border to positive values outside. These variables were extracted from many sources, as listed in Table 2.3. I used ArcMap 10.7 in the euclidean distance measurements.

I modelled the effects of the explanatory variables on the likelihood of identified changes in 2015 - 2019 using the weights of evidence approach in DINAMICA EGO software version 5 (Soares-Filho et al., 2004). The model would calculate the weight of evidence coefficients associated with a transition from class i into j for each range, or each class for categorical variables, in the explanatory variables using the following Bayesian formula (Bonham-Carter, 1994; Soares-Filho et al., 2004):

$$O\{D/B\} = O\{D\} \frac{P\{B/D\}}{P\{B/\bar{D}\}}$$

$$\log\{D/B\} = \log\{D\} + W^+,$$

where:

$O\{D\}$: the prior odds of event D

$O\{D/B\}$: the posterior (conditional) odds of event D given a spatial pattern B

$P\{B/D\}$: the posterior probability of the presence of spatial pattern B where event D occurred.

$P\{B/\bar{D}\}$: the posterior probability of the presence of spatial pattern B where event D *did not* occur.

W^+ : the weight of evidence of event D 's occurrence given a spatial pattern B

Higher, positive W^+ indicates the higher likelihood of land cover change from i into j to occur at a spatial pattern B , while negative values indicate the repelling effect of B . $W^+ = 0$ indicates neither

positive nor negative effect of the spatial pattern B on the change occurrence likelihood of concern.

2.5. Results

2.5.1. Land Cover Mapping

2.5.1.1. Accuracy

Land cover mapping with Sentinel-2 and Sentinel-1 sensors was slightly more accurate than the map produced using only Sentinel-2 data. The out-of-bag error calculated during the random forest classification also supported this finding (Sentinel-2 only = 17.39% > fusion = 15.06%). The application of post-classification morphological filters improved the accuracy of both approaches. Therefore, I decided to use land cover maps produced using the fusion approach with morphological filters hereon. The land cover map for 2015 was created using the fusion of Sentinel-2 and Sentinel-1 data with morphological filters.

Table 2.4 Land cover classification accuracy obtained from different input variable sets

Variable set	Overall Accuracy	
	Pre-filter	Post-filter
Sentinel-2 only (optical-only)	80.73	84.13
Sentinel-1 and 2 (optical-radar)	83.45	85.76

In terms of class-specific accuracy measures, the classifier had higher specificity relative to sensitivity. User's accuracy (UA) of mapped classes ranged between 77.5 - 100 %, while producer's accuracy (PA) lower range reached 41.27% for the barren land and built structure class (Table 2.4). This class and the grassland class (PA = 65.44%) had the lowest PA. Barren land was rare and difficult to identify because many were unvegetated only for a short time before being replanted. Grassland and paddy classes were often confused because both of them are dominated by grasses (family Poaceae) that appear very similar to each other and might share similar spectral and radar properties, primarily because not all paddy fields are continuously irrigated throughout the year.

2.5.1.2. Variable Importance

Variables derived from Sentinel-2 and Sentinel-1 contributed well to the land cover classification (Figure 2.3). The most important predictors reflected the seasonal variations; the 10th and 90th percentile of NDVI and radar backscatter were the variables with the highest purity gain. A proxy of texture derived from NDVI raster (*NDVI_tx*) was the third most important variable, followed by the 90th percentile of the radar backscatter (*VV_p90*). In general, the variable importance, along with the improvement of the classifier's performance following the introduction of radar variables, showed the positive effect of fusing radar with optical data in land cover mapping.

Table 2.5 Confusion matrix that contains class-specific accuracy and mapped area confidence interval

Reference class	OTC	P	B	EvF	AnP	AnC	DDF	G	SIF	W	RP	User's Acc. (%)	2019 Mapped Area	
Mapped class													km ²	km ² ±
Other tree covers (OTC)	50	0	0	0	0	0	0	0	0	0	0	100	171.32	91.92
Paddy (P)	4	916	11	0	4	43	0	49	1	0	0	89.11	5859.85	191.83
Barren land and built (B)	0	1	26	0	0	0	0	0	0	0	0	96.3	279.75	80.75
Evergreen forest (EvF)	1	0	6	661	4	0	19	1	5	0	1	94.7	2939.75	137.81
Anacardiaceae plantation (AnP)	1	11	4	0	171	8	1	12	0	0	2	81.43	1010.8	102.06
Annual or herbaceous crops (AnC)	2	18	13	0	21	299	0	32	1	0	0	77.46	2237.87	190.72
Dry deciduous forest (DDF)	2	2	0	25	4	4	432	28	7	0	0	85.71	6160.4	241.05
Grassland (G)	6	20	3	0	2	21	13	248	5	3	0	77.26	5030.7	298.25
Seasonally inundated forest (SIF)	0	0	0	0	2	0	1	8	53	0	0	82.81	518.36	108.62
Water bodies (W)	0	0	0	0	0	0	0	1	0	21	0	95.45	124.98	56.26
Rubber plantation (RP)	0	0	0	4	0	0	0	0	0	0	96	96	242.84	17.24
Producer's Accuracy (%)	76	95	41	96	82	80	93	65	74	88	97			

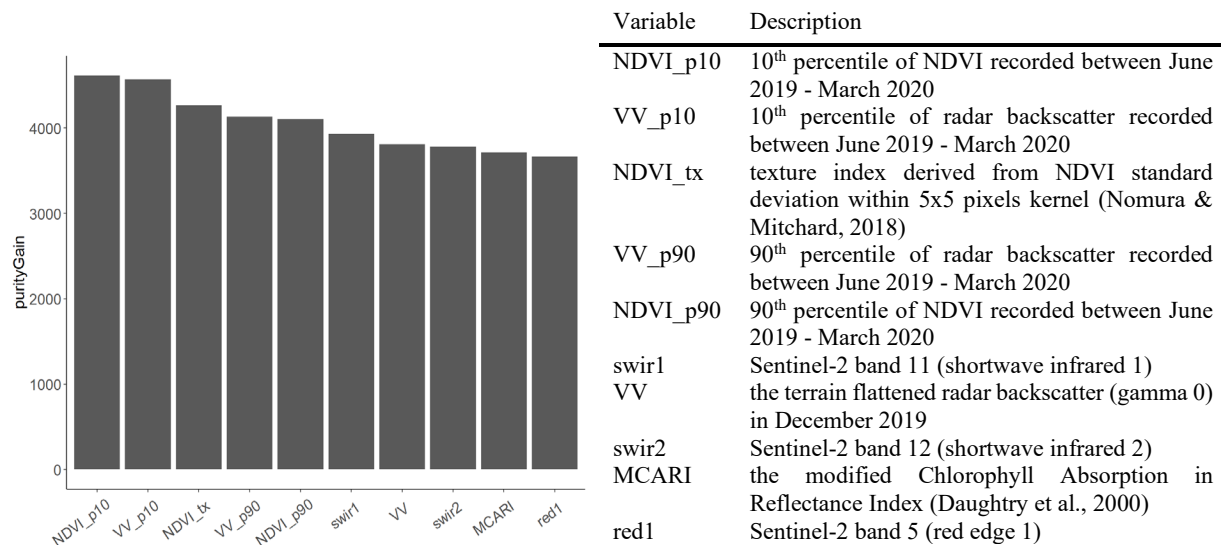


Figure 2.3 Relative importance of variables in the optical-radar fusion approach

2.5.2. Land Cover Change Analysis

Cultivated land area increased in 2015-2019, while natural or semi-natural land coverage decreased. Dry deciduous forest (DDF), the dominant class in the landscape, had undergone the largest contraction from 8,473.04 in 2015 to 6,838.05 square kilometres in 2019. I calculated this estimate based on a simple overlay of the land cover rasters under analysis, i.e., one for 2015 and the other for 2019. Due to the absence of validation data from 2015, such a simple approach does not incorporate the uncertainty of the area estimate, unlike the results presented in Table 2.5. Even though the magnitude may contain some error, the trend of change was apparent. The coverage of the other two natural or semi-natural cover classes, i.e., evergreen forest and seasonally inundated forest, also declined, but not as much as DDF.

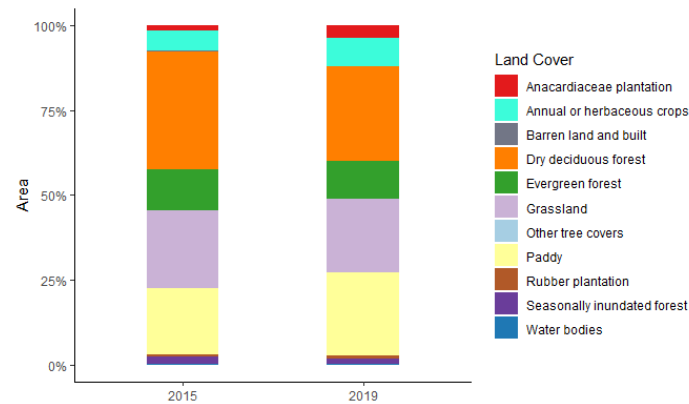


Figure 2.4 Relative area plot showed land cover classes that underwent area increase and decrease in 2015-2019.

On the other hand, expansions of cultivated land cover classes that comprise Anacardiaceae plantation, annual or herbaceous crops, paddy, and rubber plantation were apparent. In Figure 2.6, interlinkages between the contraction of natural or semi-natural covers and the increased area of cultivated land became clear. For example, DDF conversion into paddy and annual crops were among the top 10 observed changes with the largest area. However, the limited ability of the land cover classifier to tell apart paddy, grassland, and annual crops made the area estimates of these specific changes less accurate.

The change with the largest area was the conversion of DDF into grassland (Figure 2.6). Since grasses naturally dominate DDF floors (Figure 2.7), such a change can be considered a tree

2019 Land cover

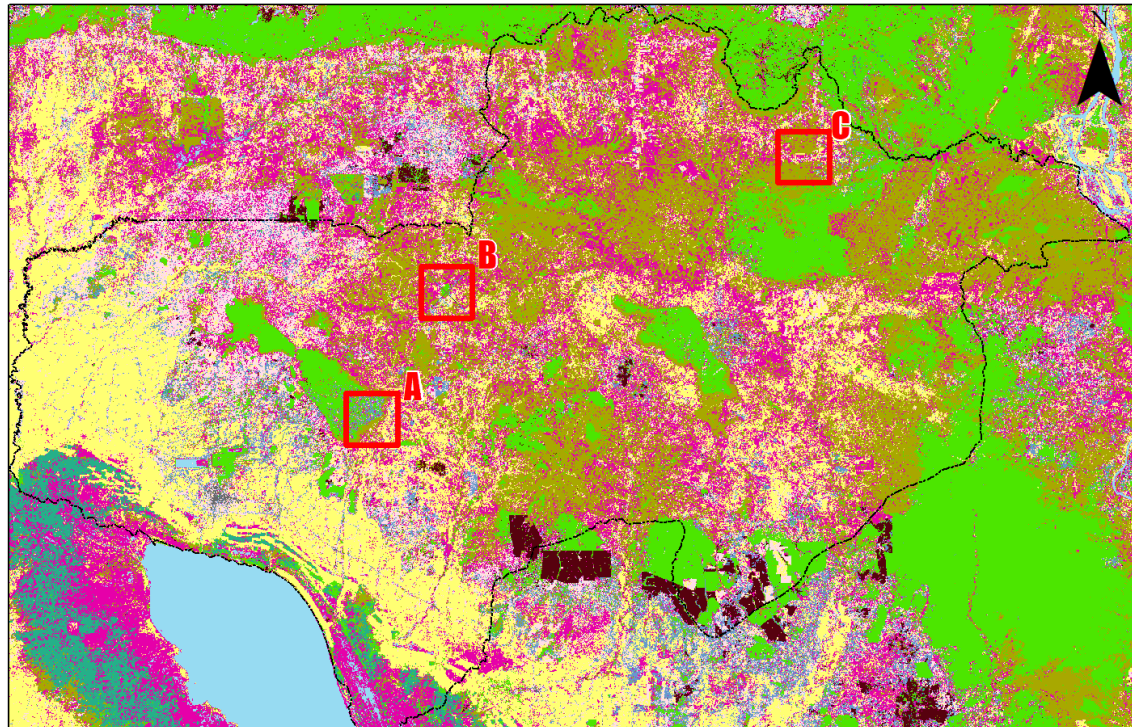
Land cover

- Anacardiaceae plantation
- Annual or herbaceous crops
- Barren land and built
- Dry deciduous forest
- Evergreen forest
- Grassland
- Other tree covers
- Paddy
- Rubber plantation
- Seasonally inundated forest
- Water bodies
- Study area

Scale

Main map (right) 30 Kilometers

Insets (bottom) 6.5 Kilometers



2015-2019 Land cover dynamics

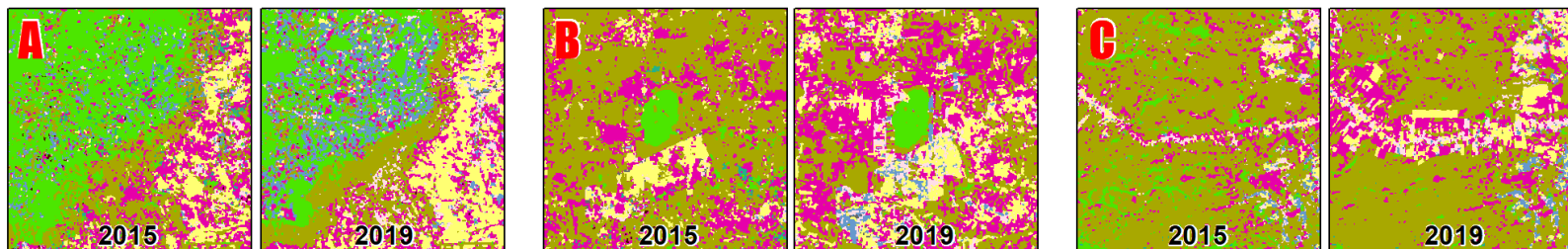


Figure 2.5 Land cover map of Siem Reap and Preah Vihear in 2019 and some of the detected changes in 2015-2019

removal rather than a conversion phenomenon. The recovery of grassland into DDF was also prevalent—indicated by the fact that such a change has the third-largest area. The prevalence of land cover changes that involved DDF, which is the class with dominant coverage, made this class worth further investigation to reveal the associated changes' pattern.

2.5.2.1. Change in focus: DDF dynamics

Three changes from/into DDF that had a magnitude above 45 thousand hectares (Figure 2.6) were among the top 6 changes with the most extensive occurrence. The first two of them were transitions between DDF and grassland, each with a different direction. The change with the lowest area among the three (45,346.24 hectares) was a conversion of DDF into paddy.

Conversion of DDF into grassland (*DDF defor.* in Figure 2.8) was associated with anthropogenic factors. The results indicated that tree removal from DDF was likely to occur at locations 1 to 2 hours away from population centers, like cities and towns, or forest patches within 500 meters radius from previously deforested areas in 2010-2015. I observed a stronger relationship between the likelihood of tree removal from DDF at the interior of economic land concessions. Protected areas seemed to prevent DDF deforestation, as indicated by the negative $W+$ coefficient value at negative distances (Figure 2.8). However, areas located inside the national park with a distance to the border of around 20 km appeared to have positive $W+$ coefficients, indicating a likelihood of tree removal from DDF. Most soil types had positive $W+$ coefficients, indicating that tree removal from DDF might occur across different soil types. The exceptions to this were acid lithosol, which has a shallow soil layer, and the unknown soil type located at cliffs near the Cambodian border in the north. Instead of the repelling effect of the unknown soil type, this may reflect the inaccessibility of areas close to the border, and hence, the low DDF tree removal risk.

The transition from grassland into DDF (*DDF recov.*) displayed different patterns from the ones observed in *DDF defor.* (Figure 2.8). The negative $W+$ values associated with the negative distance to the ELC border indicate that it was unlikely for grasslands inside ELCs to be (re)colonized by trees and become DDF. DDF tree colonization occurred in grasslands located at least 6 hours away from population centers or inside protected areas. The positive $W+$ coefficient associated with the unknown soil type emphasized the higher likelihood of DDF recovery in remote areas.

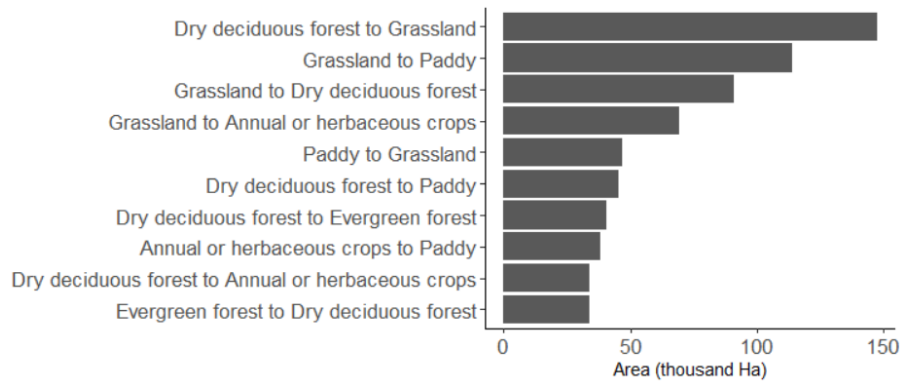


Figure 2.6 The largest changes in 2015-2019 showed the key role of cropland (paddy, annual crops) expansions in the landscape.



Figure 2.7 The grass-dominated dry deciduous forest is prone to tree removal that induces conversion of DDF into grassland.

Displacement of DDF associated with paddy expansion occurred at places suitable for paddy cultivation. I observed strong positive associations in the soil types' response curve, specifically at cultural hydromorphics and lacustrine alluvial soils; both are suitable for paddy. Protected areas prevent encroachment of paddy fields. DDF at locations near population centers (≤ 2 hours away) was prone to conversion into paddy fields. DDF inside economic land concessions was also prone to replacement by paddy field, which is not compliant with the ELCs' intended use.

2.6. Discussion

2.6.1. Comparison with other land cover mapping efforts

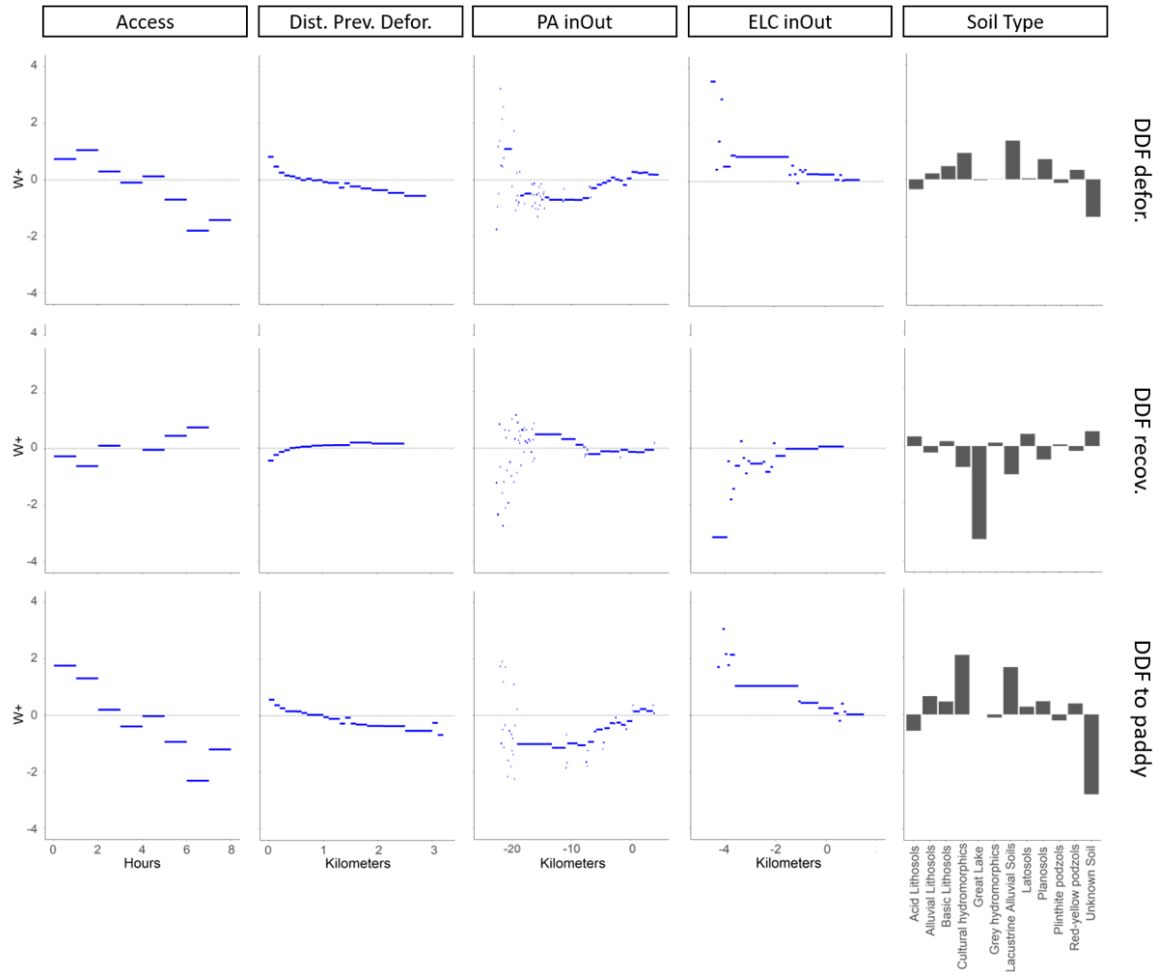


Figure 2.8 Response curves of explanatory variables included in the land cover change analysis. DDF defor. represents the “conversion” from DDF into grassland, DDF recov. represents the reverse of DDF defor.

The land cover map’s overall accuracy met the recommended minimum level of land-use/cover map accuracy (85%, Sabins, 1997) and is comparable with other efforts in similar contexts (Jia et al., 2019; Nomura & Mitchard, 2018; Poortinga et al., 2019). With more reference data, both the overall accuracy and class-specific accuracy should increase (Jia et al., 2019; Nomura & Mitchard, 2018). Aggregating land cover classes into more generic classes, such as “plantation”, “forest”, and “other” (as in Jia *et al.* 2019), improves accuracy at the expense of reduced details of the change typology. Based on the confusion matrix in Table 2.5, accuracy would increase following

a regrouping of the classes: grassland, paddy, and annual crops, which were often confused with each other. However, given the three classes' distinct ecological and socio-economic implications (Kong et al., 2021; Mahanty & Milne, 2016) and the moderately high accuracy associated with the detailed mapping, I retained the detailed depiction of land cover classes in the landscape. The omission of details in land cover mapping has been found to increase the risk of overlooking the most important driver(s) (Elz et al., 2015).

Adding Sentinel-1 (radar) data improved the classification accuracy by a small amount compared to the Sentinel-2 (optical) only classification. More sophisticated radar preprocessing steps can reduce the inconsistencies between scenes and thus, further improve the results (Small et al., 2021). A more pronounced increase in mapping accuracy was reported in a study that used Landsat (optical) and Advanced Land Observing Satellite-2 Phased Array L-band Synthetic Aperture Radar-2 (ALOS-2/PALSAR-2) (De Alban et al., 2018). Another study that used a combination of Landsat, Sentinel-2, and Sentinel-1 resulted in a similar overall accuracy level with my results (Poortinga et al., 2019). According to these findings, combining radar data with longer wavelengths better complements optical data than adding more optical data from different sensors. Unfortunately, there was no accessible ALOS-2 data coverage for the study area in 2019 during the analysis.

While the mapping accuracy was moderately high for the detailed land cover in 2019, it was not the case with the 2015 land cover map. The sparse temporal coverage of Sentinel 2 data in 2015, which relied on the observations made solely by the Sentinel 2A satellite, might have increased the inconsistency between the two mapped periods. This limitation was not a major issue for mapping the 2019 land cover map after the Sentinel 2B satellite launch on March 7th, 2017, which provides a more detailed temporal coverage. Given the limited data availability, the mapping errors for 2015 land cover maps introduced some errors in the land cover change analysis. However, I retained the detailed classification because the 2019 land cover map can depict the detailed classes relevant to the landscape's context at an acceptable accuracy level. A more accurate 2015 land cover map would bring significant improvement to the analysis. Unfortunately, such an improvement was not affordable in this study because of the limited capacity of the accessible, very high-resolution satellite images to facilitate accurate identification of detailed land cover in 2015.

2.6.2. Predictor variables' contribution

The land cover classification benefitted from predictor layers that depict contrasts between dry and wet seasons. All but one variable with the highest contribution were the 10th and 90th percentile of either the NDVI or VV backscatter recorded in June 2019-March 2020. In this case, the VV backscatter values are most sensitive to the seasonal change of the surface's moisture content while also possibly depicting some structural change associated with vegetation phenological dynamics. The importance of variables sensitive to phenological changes has been found in wider-scale studies in the area (Leinenkugel et al., 2013; Venkatappa et al., 2019). However, the variables used here were much simpler than those used in previous works; both cited works above derived phenological indices from much more sophisticated harmonic analyses. The temporal range used in deriving the simple variables also played an important role. Different ranges may not capture the critical differentiation period where vegetation types appear most distinct (Heupel, Spengler, and Itzerott, 2018).

Other variables with high importance values contain various structural and spectral information. A texture variable (*NDVI_tx*) derived from Sentinel-2 data was the third variable with the highest contribution to the classification process. This finding resonates with previous studies that found the significance of texture information (Glinskis & Gutiérrez-Vélez, 2019; Nomura & Mitchard, 2018). Two shortwave infrared bands from Sentinel-2 data were among the satellite bands with the highest contributions. These bands' significance in an optical-radar fusion was also found in a previous study (De Alban et al., 2018). A terrain flattened radar backscatter layer with VV polarization was the 7th variable with the highest purity gain. Previous studies showed that the cross-polarized VH backscatter band is more useful in identifying vegetations (Sabins, 1997). However, including a VH band in my classifier increased the out-of-bag error rate by 0.2%; therefore, I only included the VV backscatter layer.

2.6.3. Land cover change analysis

Expansion of commodity croplands was prevalent in the study area. The primary "source" of land for the predominant land expansion was grasslands. While some of the conversion from grassland to paddy can be attributed to classification errors, the widespread conversion from grassland to paddy or annual crop may indicate the recultivation of fallow lands being left for a period of time,

a common cycle in swidden agriculture that was common in the area (Wales, 2020). A fallow land tax applied by the Ministry of Agriculture in 2017 seemed to be one of the drivers of fallow land recultivation (Roeun, 2017). However, as indicated by the high level of confusion in identifying the non-tree dominated vegetation classes, i.e. grassland, paddy, and other annual crops, a more accurate mapping of the 2015 land cover is evidently required to confirm the hypothesis. Furthermore, the distinction between non-native grassland and paddy fields using appropriate sensors that can detect the slight difference in spectral signatures can further refine the change analysis.

While only a small portion of the commodity cropland expansion, i.e. the conversion into paddy fields and annual croplands, involved direct conversion from natural or semi-natural forests (Figure 2.6), an increase in land demand for crop production may be indirectly linked with the observed decline in natural forest cover (Ingalls et al., 2018). The dominant change from the dry deciduous forest into grassland could be followed by the conversion of grassland into paddy fields or other croplands. The high occurrence of DDF change into grassland in areas suitable for crop production, as well as the relatively high occurrences of grassland conversion into paddy or other annual crops (Figure 2.6), may support this hypothesis. However, a more detailed and extended monitoring period may provide stronger empirical proof of these indirect change trajectories that involve grassland as an intermediary state.

Amid the tree-crop “boom” phenomenon (Hurni et al., 2017), the associated land cover classes, such as rubber, cashews (as represented by the Anacardiaceae plantation class), as well as the pulp and paper plantations (other tree cover class), increased slightly (Figure 2.4). In contrast, annual crop and paddy underwent a substantial increase in area (Figure 2.4). The slow expansion of rubber plantations, which are mostly found inside ELCs, seemed to be related to a moratorium on new economic land concessions (ELCs) in 2012 by the national government (Neef et al., 2013). Another factor that explained the slow rubber expansion seemed to be the global rubber price that had peaked in 2011-2012 (Grogan et al., 2019). Compared to tree crops, annual crops, such as cassava and maize, provide more contribution to food security and quicker economic returns while demanding less investment capital; therefore, they are preferred by smallholder farmers (Kong et al., 2021; Mahanty & Milne, 2016).

The observed DDF change pattern displayed how accessibility and land allocation influenced some of the observed changes. The accessibility of a location affected the risk of tree removal or conversion into paddy from DDF. This finding makes sense since accessible lands are easier and economically convenient to exploit. In addition, the presence of hidden landmines in the landscape made travelling to and cultivating remote lands unsafe (Lin et al., 2021). Economic land concessions facilitated DDF deforestation and conversion. However, a surprising finding was the strong tendency of DDF inside ELCs to be converted into paddy fields instead of industrial crops, such as rubber or pulp plantation, which are more aligned with the ELC purpose. Despite being officially designated for industrial commodities plantation, 60.8 % of the ELCs in Siem Reap and Preah Vihear failed to operate (Magliocca et al., 2020). The local people's resistance was among the identified primary causes of these failures (Magliocca et al., 2020).

Protected areas played essential roles in controlling DDF deforestation and conversion as well as promoting DDF restoration. The risk of DDF conversion into paddy fields was low inside protected areas. I found a similar negative effect of protected areas on the risk of DDF deforestation, i.e. the conversion from DDF into grassland not yet followed with an evident land-use change. However, DDF in the deep interior—around 20 km from the border—of protected areas was still prone to illegal logging. Controlling illegal logging in the area was complex due to the limited coverage of regularly patrolled areas and other factors, such as the weak political will to protect nature, as seen in the release of a large portion of protected areas to the economic concessions (Ingalls et al., 2018; De Lopa, 2001; Milne, 2015).

2.6.4. Uncertainties and limitations

Given the limited availability of very high-resolution satellite images from 2015, I could not assess the accuracy of the 2015 land cover map. Sentinel-2 Level 2A data, which had been atmospherically and terrain corrected, minimized the 2019 and 2015 data variation. Temporal normalization, such as pseudo-invariant-feature normalization (Schott et al., 1988; Traganos et al., 2018), can further reduce variations associated with different sensing environments. Ideally, to test for the accuracy of the 2015 land cover map, one could use reference data collected back in 2015, for example, as in Saah *et al.*, 2020.

There were some issues related to satellite image interpretation during the desktop survey and land

cover map classification. Some land cover classes, such as grassland and different crop types, were not adequately recognizable from the very high-resolution satellite images. The confusion could introduce over/underestimation to the land cover change interpretation. This limitation could be addressed by collaboration with local partners, for example, through crowd-sourcing campaigns, which can improve the accuracy of reference data and the land cover maps. In the remote sensing context, the development of hyperspectral satellites, such as the Italian Space Agency's PRISMA, can increase the classifier's ability to identify different vegetation types. This research demonstrates that land cover change analysis will benefit from detailed land cover mapping with reliable accuracy.

The land cover change modelling results seemed to be affected by the errors in land cover classification. For instance, the similarity in patterns of the response curves of DDF changes into grassland (*DDF defor.*) with the DDF changes into paddy (*DDF to paddy*) might reflect some confusion in identifying the similar non-tree dominated vegetation classes, i.e., grassland and paddy, apart. While the misclassification-induced uncertainty was clearly present in the modelling results, the direction of the effect was unknown. One of the most common measures to address the change modelling uncertainty involves removing impossible land cover trajectories and fine-tuning the rates of changes and weights of evidence (W^+) values through series of trials-and-errors until acceptable accuracy is acquired (Elz et al., 2015). Another approach to minimize the uncertainty in land cover change modelling due to classification error using a novel algorithm named Compound Maximum a Posteriori (CMAP) had been developed recently (Reis et al., 2020). However, inadequate references and supporting data prevented this study from incorporating the identified approaches to improve the modelling outcomes. In a more specific study focusing on certain types of more general land cover change typology, e.g. DDF deforestation, we can group different changes into a more generic class to reduce the uncertainties in the model associated with the land cover classification error. With more generic classes, systematic assessment of change mapping accuracy can be more easily conducted. In this research, I retained the highest attainable details in the classification to avoid overlooking the major change driver(s) among the many possible drivers in the study area (Elz et al., 2015).

2.6.5. Management relevance

This work contributes to land cover change monitoring by providing spatial data that depicts the

recent state of the seasonally dry tropical forest landscape. The prevalence of dry deciduous forest deforestation signals the critical need for more sustainable management of the seasonally dry tropical landscape. Without a stronger commitment from the landscape managers, the natural habitat cover in the landscape, especially the dry deciduous forest, will continue to plummet under the pressure of the agricultural land expansion. The failure to control deforestation will exacerbate the extinction threats to species with concerning conservation status that rely on forest cover in the landscape.

The land cover maps for 2015 and 2019 provide valuable input for scenario development in sustainable landscape planning. The four-year period used in this analysis matches the duration of validity of development plans in the area. After some improvements to address the 2015 mapping inconsistencies, a business-as-usual baseline can be developed using the assumption that the land cover change pattern and rate observed in 2015-2019 will carry on in the future. Such a spatially explicit prediction can be developed using Markov chain-Cellular automata analysis in modelling software, e.g. DINAMICA EGO (Soares-Filho et al., 2004). Multiple scenarios can be developed by tweaking the rate or the weight coefficients of relevant changes. Based on multiple scenario analyses, comparisons can be made to estimate possible outcomes of different scenarios and the associated trade-offs, which allow land managers to make informed decisions that define the future of the landscape. Such a complex process will require close collaboration between different stakeholders but is plausible and essential to achieving sustainable landscape management and greener economic growth (Mulia et al., 2019).

2.7. Conclusion

1. The fusion of Sentinel-1 and Sentinel-2 data produced the highest accuracy in land cover mapping.
2. Timber extraction from the natural or semi-natural forest and expansion of cultivated lands, especially annual crops, were the major drivers of land cover change in the landscape.

Chapter 3. GEDI Waveform Metrics in Vegetation Mapping – A Case Study from A Complex Tropical Forest Landscape

3.1. Introduction

The spatial distribution of vegetation types is essential information in understanding the land sector's role in global climate, terrestrial carbon cycle, and biodiversity (Bajželj and Richards 2014, Malhi *et al.* 2008, Powers and Jetz 2019). The most common approach to derive this information is by classifying satellite images acquired using optical sensors. However, vegetation types that are spectrally similar but structurally different in complex landscapes are often confused (Saah *et al.*, 2020). These vegetation types often comprise different species of tree-dominated vegetation with unique ecological and/or economic significance. Distinguishing different natural forest types and tree plantations may be challenging with spectral information alone (Fagan *et al.* 2018, De Alban *et al.* 2018, Singh *et al.* 2019).

Limited accessibility, among other factors, has been a challenge in collecting references to map vegetation types in data-limited areas, especially in the developing world. Therefore, reference data often rely on high-resolution satellite images. In addition to being limited in availability, these optical images, with a spatial resolution of < 5 meters, often failed to capture the presence of tree canopies because they are often acquired in the dry season with less cloud cover when deciduous trees shed their leaves. These limitations complicate the mapping of vegetation types in these areas.

Vegetation structure information has been used to improve the identification of vegetation types in complex and dynamic landscapes (De Alban *et al.* 2018, Reiche *et al.* 2018, Fagan *et al.* 2018). In a study in southern Myanmar, adding L-band radar data with optical data increased the classification accuracy of land cover mapping up to 2% relative to the optical-only classification result (De Alban *et al.* 2018). More recently, the combination of radar data with optical data yielded satisfactory classification accuracies (Poortinga *et al.*, 2019). It should be noted, however, that radar data are sensitive to an array of factors other than vegetation structure per se, including moisture content, biomass and orientation.

With a higher information specificity, lidar (light detection and ranging) sensors directly interact with vegetation structure making it very useful in vegetation mapping, either alone (Marselis *et al.*

2018) or in combination with other remotely sensed data (Singh *et al.* 2019). However, most of the lidar data had been collected by airborne laser scanning (ALS) (Dubayah *et al.* 2020); hence it is difficult to obtain structural information at a large scale in a standardized manner. In late 2018, NASA launched the Global Ecosystem Dynamics Investigation (GEDI) mission to collect vegetation structure data at a global scale using full-waveform lidar.

GEDI is a novel experimental space-borne waveform lidar mission onboard the International Space Station (ISS) launched on 5 December 2018 by NASA. GEDI instrument comprises three lasers: one coverage laser that shoots four weaker beams and two full power lasers, each of which shoots a pair of stronger intensity beams. The full power lasers are expected to have a better penetration ability to sense through vegetation canopies than the coverage laser (GEDI Science Team 2020a). One of the mission's objectives is to detect the vegetation structural change across the temperate and tropical regions of the globe (Dubayah *et al.* 2020).

This study will explore the utilization of GEDI waveform data in identifying different vegetation types in a complex tropical landscape in Cambodia. We aimed to (1) evaluate the potential of variables derived from Level 1B and Level 2A GEDI data in vegetation type classification based on the classification accuracy and variable importance, and (2) compare the GEDI-only classification results with an existing land cover map of the region. The findings will provide an empirical assessment of GEDI data robustness for vegetation mapping in Cambodia and other complex landscapes where similar challenges in the vegetation mapping linger.

3.2. Data and Method

3.2.1. Study Area

This study covered two provinces in Cambodia: Siem Reap and Preah Vihear. The two provinces cover 10,299 and 13,788 square kilometres, respectively. Cambodia had undergone the largest amount of deforestation between 2011 and 2017 among other countries in the Lower Mekong (Potapov *et al.* 2019). Despite being categorized as humid (Title and Bemmels 2018, Bastin *et al.* 2017), the area is strongly affected by seasonal monsoon dynamics (Thoeun 2015). The monsoonal climate and a long history of anthropogenic land modification had formed a complex landscape covered by a mosaic of natural forest, both deciduous and evergreen, savannah, cropland, and tree

plantations (Figure 3.1) (Kiyono *et al.* 2018, Singh *et al.* 2019, Grogan *et al.* 2019).

3.2.2. GEDI Dataset

We downloaded 19 collections of GEDI data Level 1B and Level 2A version 001 acquired in April – early August 2019 using the GEDI Finder service (<https://lpdaacsvc.cr.usgs.gov/services/gedifinder>). Level 1B data contains the “raw” geolocated full-waveform data. In contrast, level 2A contains the elevation and height metrics extracted from the full-waveform data, such as the relative heights of the wave components, the elevation of the presumed ground and canopy, and energy intensity metrics. Potapov *et al.* (2021) used GEDI Level 2A data combined with time-series Landsat data to produce a wall-to-wall global forest canopy height map. Another study in Gabon found that Plant Area Index (PAI) extracted from GEDI Level 2B data was pivotal in identifying different successional vegetation types (Marselis *et al.* 2018).

After applying a quality filter based on a quality flag index in Level 2A data, we obtained 79,498 granules; each contains measurements of all eight dithered 25 m footprints, separated approximately 600 meters across-track and 60 m along-track, generated from four beams used in a “push-broom” scanning (Hofton *et al.*, 2019).

3.2.3. Waveform Metrics

We extracted 120 metrics based on the Level 1B and Level 2A granules (Table 3.1), 101 of which were relative height (rh) metrics directly extracted from Level 2A data products. Metrics were grouped into categories, namely topographic (1 metric), intensity (6 metrics), height (104 metrics; the rh metrics belong to this category), shape (8 metrics), and beam type (1 metric). We summarized the 101 rh metrics using a simple linear regression between the rh values and the square root of the associated percentile ids (0-100). The linear model produced three metrics: *coef.slope*, *coef.int*, and *rsdVar* represent the slope, intercept, and the model’s residual variance, respectively. We replaced all rh metrics, except for rh 0, 10, 99, and 100, with the three modelled variables resulting in 23 variables for analysis.

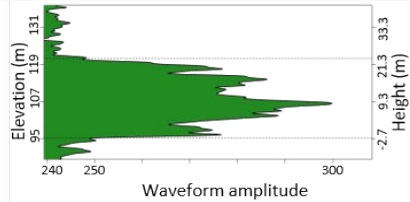

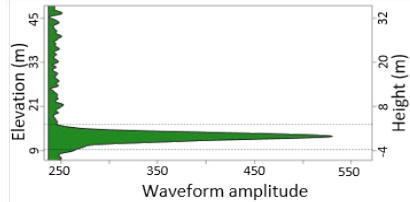

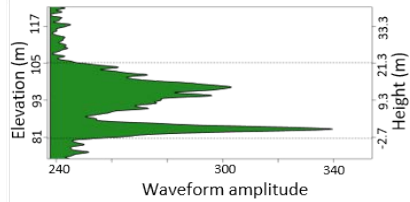

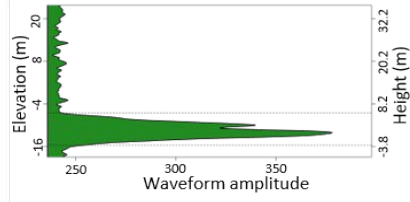

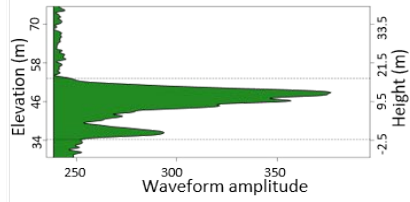

Vegetation types	Abbreviation	Description	Waveform sample	Photo sample
Evergreen forest	Evergreen	Forest dominated by evergreen tree species with tree cover $\geq 40\%$		
Cropland or grassland	CropGrss	Cropland or grass or shrub; may include seedlings or saplings.		
Deciduous forest	DecidF	Deciduous forest with open canopy cover ($\geq 25\%$) and grass-dominated understory		
Seasonally inundated forest	FloodFor	Forest dominated by trees that can withstand seasonal flooding		
Rubber plantation	Rubber	Mature rubber tree (Hevea sp.) plantation		

Figure 3.1. Example of GEDI waveform data and the associated vegetation classes. Dashed lines in the waveform samples indicate the top and bottom edge of the waveform, i.e. the relative height 100 and 0.

Table 3.1 Description of the GEDI-derived metrics extracted from GEDI Level 1B and Level 2A collections.

Code	Description
Metrics extracted from L2A	
elev_lowestmode	The elevation of the last detected mode relative to a reference ellipsoid
num_detectedmodes	The count of the waveform peaks detected after smoothing and noise cleaning
rh_x where $x = [0,100]$	The relative height (rh) of the point where $x\%$ of the laser intensity was below this point (Dubayah et al. 2010). rh_{100} is associated with the ground height of the highest canopy, while rh_0 is considered the end of the waveform.
coef.slope	The slope of the linear model $rh_x \sim \sqrt{x} \mid x = [0,100]$
coef.int	The intercept of the linear model $rh_x \sim \sqrt{x} \mid x = [0,100]$
rsdVar	The variance of the residuals of the linear model $rh_x \sim \sqrt{x} \mid x = [0,100]$
Metrics processed from L1B using waveformLidar (Zhou & Popescu, 2019)	
rough	The distance in meters between the start of the waveform and the presumed first waveform peak
frontS	The tilt from the beginning of the waveform to its first presumed peak
grInt	Area under the waveform component(s) that represent the ground area
vlnt	Area under the waveform component(s) that represent the detected vegetation
abgInt	Area under the waveform component(s) that lies above the presumed ground elevation
rtIntVG	The ratio of vlnt/abgInt
Additional metrics processed from L1B	
b5_maxMin	Number of waveform's peaks and valleys below 2 m above the identified ground height
b4_maxMin	Number of waveform's peaks and valleys between 2-5 m above the identified ground height
b3_maxMin	Number of waveform's peaks and valleys between 5-10 m above the identified ground height
b2_maxMin	Number of waveform's peaks and valleys between 10-15 m above the identified ground height
b1_maxMin	Number of waveform's peaks and valleys above 15 m above the identified ground height
grAmpl	The amplitude of the waveform at the lowest detected mode (ground)
ampl90	The amplitude of the waveform at the rh90 point (crown)
ampRatio	The ratio between the ampl90/grAmpl

3.2.4. Vegetation Classification

Reference data were collected through the interpretation of very high-resolution (VHR) satellite images in the Collect Earth platform (Bey *et al.* 2016) that provided access to different VHR satellite image providers, such as Google Earth, Google Earth Engine, and Planet data. In total, we

successfully identified the vegetation types across 1,062 granules. A Random Forest classifier was then developed to link these five classes to the GEDI metrics. The model was developed using the R package `randomForest` and was implemented in R. The number of trees to grow was set to 500, with the minimum size of terminal nodes set to 4. The classification accuracy was assessed based on the out-of-bag error estimate and the confusion matrix, both of which are the standard procedures in assessing Random Forest classifier performance. Once developed, we applied the classifier to identify the vegetation types across 78,436 GEDI granules.

Variable importance was measured as the mean decrease in classification accuracy and the mean decrease in Gini. The former indicates the change in the classifier's error when the variable under evaluation is omitted. Random Forest classifier utilizes decision trees, which consist of nodes where observations are split according to the associated variable's values. Each split is expected to reduce the heterogeneity or impurity within groups, which is directly proportional to the Gini coefficient. Variables associated with a higher decrease in Gini have more significant contributions to the homogeneity of the nodes and leaves in the classification and hence, have relatively higher importance given other variables in the classifier.

3.2.5. Comparison with Land Cover Data

We compared the vegetation classification developed solely using GEDI metrics with a land cover map that depicted the most recent state of the study area. The land cover was developed using a fusion of Sentinel-2 and Sentinel-1 images acquired between June 2019 – March 2020 in a Random Forest classification on Google Earth Engine (Gorelick *et al.* 2018). The land cover mapped 11 classes, including different types of anthropogenic land cover classes along with natural and semi-natural vegetation types, with an overall accuracy of 83.45% (Table 2.4). We reclassified some of the mapped land cover classes to match the vegetation types of interest: the “other tree-dominated” class, which covered only a tiny part of the study area, was reclassified as rubber plantation; “paddy”, “mango” or “cashew plantation” classes were reclassified as cropland. The land cover at each of the GEDI footprints was extracted from the map using the footprint centroid's coordinates.

3.3. Results

3.3.1. Vegetation classification using GEDI metrics

The Random Forest classifier developed only using GEDI metrics distinguished five different vegetation types with moderate accuracy (Table 3.2). The out-of-bag estimate of error rate was 18.08%, with the deciduous forest as the class with the lowest accuracy (user's accuracy = 73%, producer's accuracy = 69%). The vegetation type with the highest accuracy was the "AnnCrop" class that comprises annual crops, grassland, and shrubland.

Table 3.2. Accuracy matrix of the GEDI-only classification

Predicted\Ref.	AnnCrop	DecidF	Evergreen	FloodFor	Rubber	Col. sum	User's acc.
AnnCrop	413	28	6	12	4	463	89%
DecidF	11	160	46	0	3	220	73%
Evergreen	7	39	157	0	8	211	74%
FloodFor	10	1	1	90	0	102	88%
Rubber	12	4	0	0	50	66	76%
Row sum	453	232	210	102	65	1062	
Producer's acc.	91%	69%	75%	88%	77%	Overall accuracy = 81.9%	

Out-Of-Bag (OOB) estimate of error rate: 18.1%

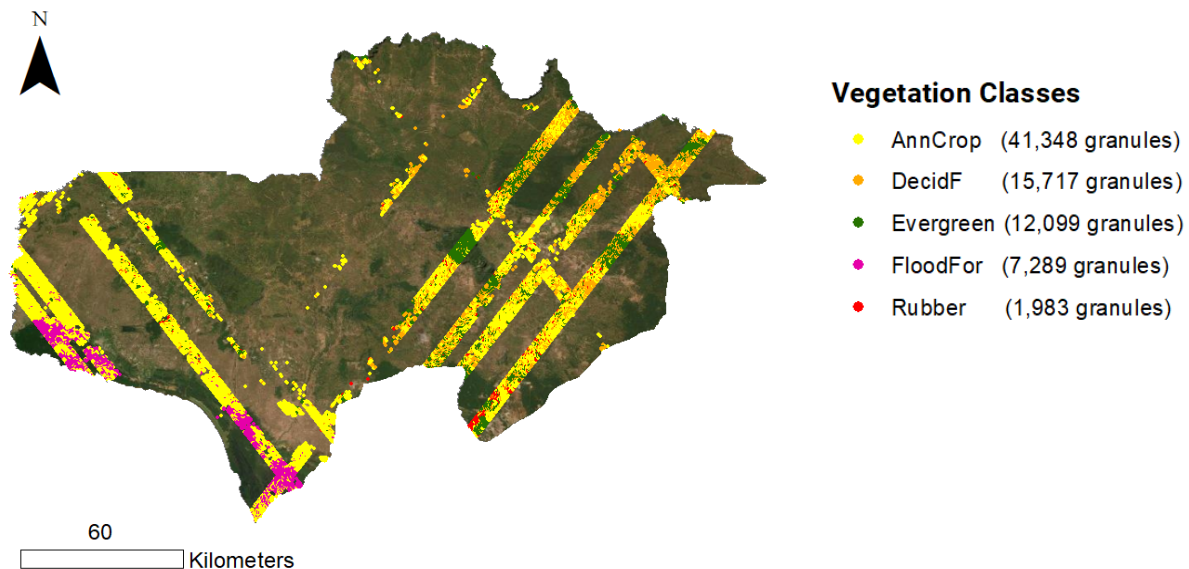


Figure 3.2. Classification result showing the distribution of the vegetation classes in the study area.

The map of the classified granules depicts the vegetation types' spatial distribution (Figure 3.2). Cropland/grassland was prevalent across the landscape, while the other vegetation types were clustered at certain parts of the study area. The seasonally inundated forest granules were concentrated in the area surrounding the Tonle Sap Lake in the Southwest. According to the land cover data, the drier area in the Northeast hosted most of the deciduous forest, which is the dominant land cover type in the study area. Evergreen forest and rubber plantation distribution were confined to the protected and concession area, respectively.

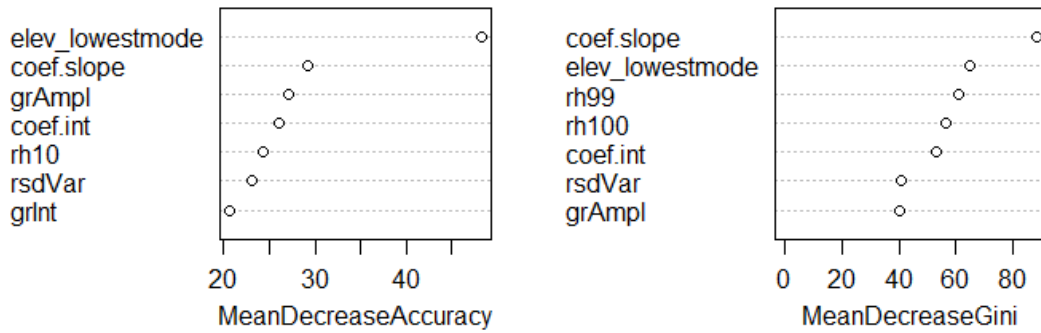


Figure 3.3. Variable importance as indicated by the Mean Decrease Accuracy and Mean Decrease Gini.

Topographic information provided by GEDI measurement, the relative elevation of the lowest mode (*elev_lowestmode*), was the metric with the highest mean decrease accuracy when omitted (Figure 3.3). The variables we derived from linear models of the *rh* metrics all had high importance values. The slope of the linear model (*coef.slope*) has the highest Mean Decrease Gini value. The metrics derived from Level-1B data with the highest importance values were the *grInt* and *grAmpl*, representing the ground portion of the waveform.

3.3.2. Comparison between GEDI-only classification and optical-radar classification

The comparison of the classified GEDI granules with the land cover data revealed that the two are in agreement in most cases (Figure 3.4). In many cases where the GEDI-only classification disagreed with the land cover map, the GEDI-only classification tends to be closer to the reference than the land cover map.

We found some exceptions to this when GEDI-only prediction misidentified evergreen forest as deciduous and deciduous as evergreen. The misclassification may be related to the moderate error rate in distinguishing the two natural forest classes, as shown in Table 2. The land cover map

distinguished evergreen from the deciduous forest more accurately because the major difference between the two vegetation types is more phenological rather than structural; the former is detectable by optical data. The reference data collection, which relied on the interpretation of high-resolution optical images, was also a potential factor that explains the higher agreement between the reference data and land cover data than with GEDI-only prediction for the deciduous and evergreen forest.

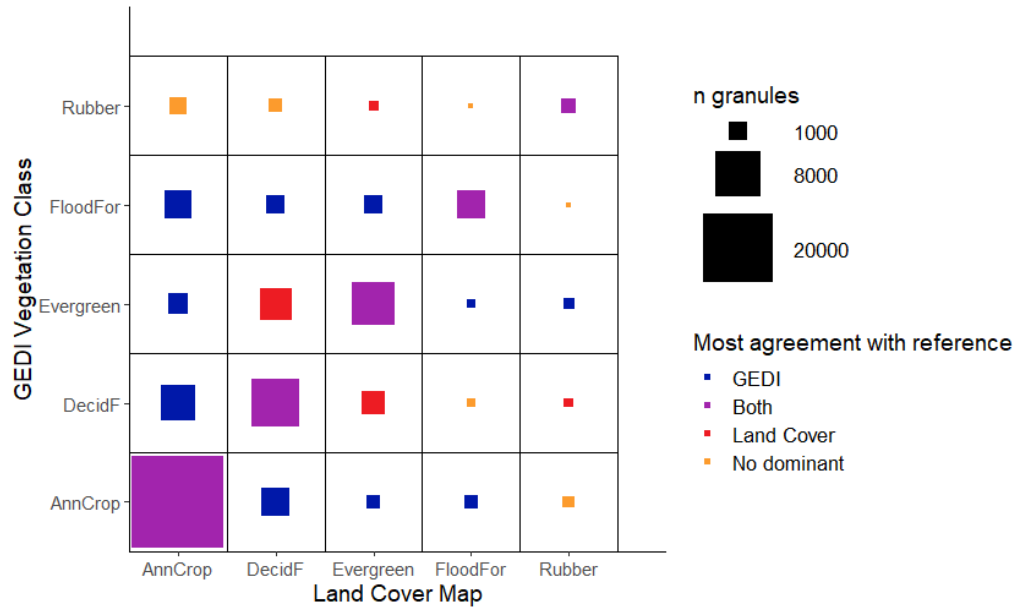


Figure 3.4. Comparison of the classification result and 2019 land cover map

3.4. Discussion

We demonstrated the potential of GEDI metrics derived from Level-1B and Level-2A in identifying vegetation types in Cambodia, where intensively managed types coexist with natural or semi-natural vegetation types. The application expanded the previously studied application of waveform lidar data in a landscape dominated by natural vegetation types (Marselis *et al.*, 2018). Without any implementation of post-classification correction, as seen in many other studies (e.g. (Marselis *et al.* 2018, Poortinga *et al.* 2019), our classification that was based solely on GEDI Level-1B and 2A metrics was sufficient in mapping the distribution of the major vegetation classes in the study area.

Errors in the classification were related to several possible factors. The high classification error between the deciduous forest and the evergreen forest in our results may indicate the presence of

a mixed forest class, where the tree species of the two forest types coexist (Rollet 1972, Food and Agriculture Organization of the United Nations 2020). However, the definition of a mixed forest class, which requires information on the tree species composition, was not feasible in this study due to the absence of overlaps between our ground reference points with the GEDI footprints. Both GEDI-only classification and land cover were prone to commission error in mapping the rubber plantations, which had limited distribution and was concentrated in concession areas. The locations where GEDI-only results and land cover data agree can improve the specificity of the rubber plantation distribution mapping.

Our classifier mapped the seasonally flooded forest with better accuracy than the land cover map. In this case, the elevation information measured by GEDI (`elev_lowestmode`) provides valuable information that was missing in the land cover map that was developed using a fusion of Sentinel-1 and Sentinel-2 data. The elevation data that was measured relative to a reference ellipsoid (Dubayah *et al.* 2020) contributed to the identification of the areas that were prone to flooding. Hence, incorporating the variable as a predictor improved the classifier's ability to identify the seasonally flooded forest.

The most important variables in the GEDI-only Random Forest classifier consisted of topographic, intensity, and height metrics derived from Level 1-B and Level 2-A GEDI data. The strong association of topographic information with vegetation classification was consistent with previous studies, e.g. Marselis *et al.* 2018. Two variables that were expected to have higher importance, according to Zhou and Popescu (2019), i.e., the front slope angle and the slope distance from the waveform beginning to the highest peak, had low importance in the classification.

A similar approach can be applied to version 2 GEDI data that were recently released. The increased geolocation accuracy of GEDI version 2 may increase the vegetation classification accuracy. However, our preliminary observation showed that version 2 waveforms were preprocessed using a different approach from version 1 and thus have “smoother” waveforms. Therefore, some modifications may be necessary to replicate our framework using the version 2 GEDI data.

3.5. Conclusion

GEDI Level 1-B and Level 2-A provide topographic and vegetation structure information that is useful in vegetation types identification in the study area. The GEDI-only classification was able to identify vegetation types with good agreement with the land cover map derived from the fusion of Sentinel-2 and Sentinel-1 images. The GEDI measured elevation data was particularly useful in mapping the seasonally inundated forest. The mapping of rubber plantations, which was uncommon in the study area, could be refined by combining the GEDI-only result with the land cover map to see where the two agree.

These findings emphasize the potential of GEDI waveform lidar data in developing the policy-relevant map, especially in data-scarce environments. Given the limited extent of the GEDI data coverage, GEDI waveform information can be combined with the high-resolution optical images that are commonly used in reference data collection platforms to reduce uncertainty in vegetation types identification. This approach can improve the reference data quality, as demonstrated in this research by the accurate identification of vegetation types by the GEDI-only classification.

As more data are acquired through the lifetime of the GEDI mission, more classes can be incorporated into the classification. This may also reduce the classification error associated with the high intra-class variation. Further research is needed to systematically assess the classification accuracy provided by a classifier developed based on the waveform metrics provided by GEDI Level 1-B and Level 2-A dataset, combined with Level 2-B dataset, for example, the canopy cover and Plant Area Index (PAI).

Chapter 4. Conclusion

In this thesis, I demonstrated the application of novel remote sensing technologies in improving land cover mapping in the landscape of Siem Reap and Preah Vihear. The fusion between Sentinel-1 and Sentinel-2 satellite data slightly improved the land cover classification accuracy (1.6% overall accuracy increase) relative to Sentinel-2-only classification. Even after some correction efforts, the noises recorded in Sentinel-1 radar data seemed to be one of the factors that caused the lower-than-expected increase in the overall accuracy of the radar-optical fusion approach. The availability of analysis-ready Sentinel-1 data, as proposed by Small *et al.* (2021), will reduce the noise and inconsistencies in the mosaic of Sentinel-1 scenes, and therefore, increase the performance of the optical-radar fusion approach for accurate land cover mapping in larger landscapes that require several observation swaths and rows to obtain full coverage.

Despite the moderate level of overall accuracy, the produced map successfully distinguished different natural or semi-natural forest types in the study area at a finer resolution than an existing land cover map in the area. The most significant variables were the ones containing seasonal/phenological information. A similar finding on the significance of phenological information in mapping vegetation types in the area was reported in a recent study that used the phenological indices derived from Landsat's Enhanced Vegetation Index (EVI) time-series data. Based on my results, remotely sensed seasonal/phenological indicators (*VV_p10*, *VV_p90*, *NDVI_p10*, *NDVI_p90*) are useful in mapping vegetations in tropical landscapes with pronounced seasonal temperature and precipitation variation. In areas with frequent cloud cover, radar predictors are particularly useful as the microwave signals used in radar instruments can penetrate through clouds.

The land cover change analysis identified a substantial decline of deciduous dipterocarp forest cover within the study period while commodity cropland expanded. A significant relationship between the explanatory variables and deciduous dipterocarp forest dynamics indicates the significance of anthropogenic factors in driving land cover change in the area. For instance, the land cover change model showed how land designations defined the most likely change in the affected area. The land cover change model I developed can provide a baseline for scenario development that supports sustainable landscape planning and monitoring in the area.

The second part of my thesis demonstrated how topographic and vegetation structure information from the new GEDI data could further improve vegetation classification accuracy in complex tropical landscapes. GEDI-only classification could identify the seasonally inundated forest better than the land cover map derived from the fusion of optical-radar data. These findings indicate the potential of structural information in mapping complex landscapes' land cover with an enhanced level of detail and accuracy. Airborne laser scanning can be deployed in the study area to obtain complete “wall-to-wall” lidar coverage.

Bibliography

- De Alban, J. D. T., Connette, G. M., Oswald, P., & Webb, E. L. (2018). Combined Landsat and L-band SAR data improves land cover classification and change detection in dynamic tropical landscapes. *Remote Sensing*, 10(2). <https://doi.org/10.3390/rs10020306>
- Allan, J. A. (1980). Remote sensing in land and land use studies. *Geography*, 65(1), 35–43.
- Bey, A., Díaz, A. S. P., Maniatis, D., Marchi, G., Mollicone, D., Ricci, S., et al. (2016). Collect earth: Land use and land cover assessment through augmented visual interpretation. *Remote Sensing*, 8(10), 1–24. <https://doi.org/10.3390/rs8100807>
- Bey, A., Jetimane, J., Lisboa, S. N., Ribeiro, N., Siteo, A., & Meyfroidt, P. (2020). Mapping smallholder and large-scale cropland dynamics with a flexible classification system and pixel-based composites in an emerging frontier of Mozambique. *Remote Sensing of Environment*, 239(November 2019), 111611. <https://doi.org/10.1016/j.rse.2019.111611>
- Blackburn, G. A. (1998). Quantifying chlorophylls and carotenoids at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sensing of Environment*, 66(3), 273–285. [https://doi.org/10.1016/S0034-4257\(98\)00059-5](https://doi.org/10.1016/S0034-4257(98)00059-5)
- Bonham-Carter, G. F. (1994). Tools for Map Analysis: Multiple Maps. In *Geographic Information Systems for Geoscientists: Modelling with GIS* (pp. 267–337). Ottawa, Ontario, Canada: Pergamon. <https://doi.org/10.1016/b978-0-08-041867-4.50014-x>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1201/9780429469275-8>
- Buschmann, C., & Nagel, E. (1993). In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *International Journal of Remote Sensing*, 14(4), 711–722.
- Chaplin-Kramer, R., Sharp, R. P., Mandle, L., Sim, S., Johnson, J., Butnar, I., et al. (2015). Spatial patterns of agricultural expansion determine impacts on biodiversity and carbon storage. *Proceedings of the National Academy of Sciences of the United States of America*, 112(24), 7402–7407. <https://doi.org/10.1073/pnas.1406485112>

- Chhetri, P. K., & Thai, E. (2019). Remote sensing and geographic information systems techniques in studies on treeline ecotone dynamics. *Journal of Forestry Research*, 30(5), 1543–1553. <https://doi.org/10.1007/s11676-019-00897-x>
- Daughtry, C. S. T., Walthall, C. L., Kim, M. S., De Colstoun, E. B., & McMurtrey, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74(2), 229–239. [https://doi.org/10.1016/S0034-4257\(00\)00113-9](https://doi.org/10.1016/S0034-4257(00)00113-9)
- Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., et al. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, 1(January), 100002. <https://doi.org/10.1016/j.srs.2020.100002>
- Elz, I., Tansey, K., Page, S. E., & Trivedi, M. (2015). Modelling deforestation and land cover transitions of tropical peatlands in Sumatra, Indonesia using remote sensed land cover data sets. *Land*, 4(3), 670–687. <https://doi.org/10.3390/land4030670>
- ESA. (2020). SNAP - ESA Sentinel Application Platform v8.0.0. Retrieved from <https://step.esa.int/main/snap-8-0-released/>
- Fagan, M. E., Morton, D. C., Cook, B. D., Masek, J., Zhao, F., Nelson, R. F., & Huang, C. (2018). Mapping pine plantations in the southeastern U.S. using structural, spectral, and temporal remote sensing data. *Remote Sensing of Environment*, 216(June), 415–426. <https://doi.org/10.1016/j.rse.2018.07.007>
- Fitzherbert, E. B., Struebig, M. J., Morel, A., Danielsen, F., Brühl, C. A., Donald, P. F., & Phalan, B. (2008). How will oil palm expansion affect biodiversity? *Trends in Ecology and Evolution*, 23(10), 538–545. <https://doi.org/10.1016/j.tree.2008.06.012>
- Gao, B.-C. (1996). NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water From Space. *Remote Sensing of Environment*, 58(3), 257–266. <https://doi.org/10.24059/olj.v23i3.1546>
- Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P., Ramankutty, N., & Foley, J. A. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s

- and 1990s. *Proceedings of the National Academy of Sciences*, 107(38), 16732–16737. <https://doi.org/10.1073/pnas.0910275107>
- Glinskis, E. A., & Gutiérrez-Vélez, V. H. (2019). Quantifying and understanding land cover changes by large and small oil palm expansion regimes in the Peruvian Amazon. *Land Use Policy*, 80(October 2018), 95–106. <https://doi.org/10.1016/j.landusepol.2018.09.032>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Grogan, K., Pflugmacher, D., Hostert, P., Mertz, O., & Fensholt, R. (2019). Unravelling the link between global rubber price and tropical deforestation in Cambodia. *Nature Plants*, 5(1), 47–53. <https://doi.org/10.1038/s41477-018-0325-4>
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850–853. <https://doi.org/10.1126/science.1244693>
- Heupel, K., Spengler, D., & Itzerott, S. (2018). A Progressive Crop-Type Classification Using Multitemporal Remote Sensing Data and Phenological Information. *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 86(2), 53–69. <https://doi.org/10.1007/s41064-018-0050-7>
- Hofton, M., Blair, B., Story, S., & Yi, D. (2019). Algorithm Theoretical Basis Document (ATBD) for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products. Greenbelt: Goddard Space Flight Center.
- Huete, A.R. (1988). A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309.
- Huete, Alfredo R., Justice, C., & Leeuwen, W. van. (1999). *MODIS Vegetation Index (MOD 13) Algorithm Theoretical Basis*.
- Hurni, K., Schneider, A., Heinemann, A., Nong, D. H., & Fox, J. (2017). Mapping the expansion

- of boom crops in Mainland Southeast Asia using dense time stacks of landsat data. *Remote Sensing*, 9(4). <https://doi.org/10.3390/rs9040320>
- Ingalls, M. L., Diepart, J., Truong, N., Hayward, D., Neil, T., Phomphakdy, C., et al. (2018). *State of Land in the Mekong Region*. Bern.
- Jia, X., Khandelwal, A., Carlson, K., Gerber, J. S., West, P. C., & Kumar, V. (2019). Plantation Mapping in Southeast Asia. *Frontiers in Big Data*, 2(December), 1–12. <https://doi.org/10.3389/fdata.2019.00046>
- Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112(10), 3833–3845. <https://doi.org/10.1016/j.rse.2008.06.006>
- Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., et al. (2016). A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, 8(1), 1–23. <https://doi.org/10.3390/rs8010070>
- Kavitha, A. V., Srikrishna, A., & Satyanarayana, C. (2021). A Review on Detection of Land Use and Land Cover from an Optical Remote Sensing Image. *IOP Conference Series: Materials Science and Engineering*, 1074(1), 012002. <https://doi.org/10.1088/1757-899x/1074/1/012002>
- Key, C. H., Benson, N., Ohlen, D., Howard, S., McKinley, R., & Z., Z. (2002). The normalized burn ratio and relationships to burn severity: ecology, remote sensing and implementation. In J. D. Greer (Ed.), *Proceedings of the Ninth Forest Service Remote Sensing Applications Conference*. San Diego, CA: American Society for Photogrammetry and Remote Sensing.
- Kong, R., Castella, J. C., Suos, V., Leng, V., Pat, S., Diepart, J. C., et al. (2021). Investigating farmers' decision-making in adoption of conservation agriculture in the Northwestern uplands of Cambodia. *Land Use Policy*, 105(March). <https://doi.org/10.1016/j.landusepol.2021.105404>
- Leinenkugel, P., Kuenzer, C., Oppelt, N., & Dech, S. (2013). Characterisation of land surface phenology and land cover based on moderate resolution satellite data in cloud prone areas -

- A novel product for the Mekong Basin. *Remote Sensing of Environment*, 136, 180–198. <https://doi.org/10.1016/j.rse.2013.05.004>
- Leinenkugel, P., Wolters, M. L., Oppelt, N., & Kuenzer, C. (2015). Tree cover and forest cover dynamics in the Mekong Basin from 2001 to 2011. *Remote Sensing of Environment*, 158, 376–392. <https://doi.org/10.1016/j.rse.2014.10.021>
- Lin, E., Sprunger, C. D., & Hwang, J. (2021). The farmer’s battlefield: traditional ecological knowledge and unexploded bombs in Cambodia. *Agriculture and Human Values*, (0123456789). <https://doi.org/10.1007/s10460-021-10195-0>
- De Lopa, T. T. (2001). Deforestation in Cambodia: A stakeholder management approach. *International Journal of Sustainable Development and World Ecology*, 8(4), 380–394. <https://doi.org/10.1080/13504500109470095>
- Loveridge, R., & Srun, T. (2015). *Ten-year species action plan for the Giant Ibis *Thaumatibis gigantea* in Cambodia*. Phnom Penh, Cambodia.
- Magliocca, N. R., Khuc, Q. Van, De Bremond, A., & Ellicott, E. A. (2020). Direct and indirect land-use change caused by large-scale land acquisitions in Cambodia. *Environmental Research Letters*, 15(2). <https://doi.org/10.1088/1748-9326/ab6397>
- Mahanty, S., & Milne, S. (2016). Anatomy of a boom: Cassava as a ‘gateway’ crop in Cambodia’s north eastern borderland. *Asia Pacific Viewpoint*, 57(2), 180–193. <https://doi.org/10.1111/apv.12122>
- Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., & Gascon, F. (2017). Sen2Cor for Sentinel-2. In *Image and Signal Processing for Remote Sensing XXIII* (p. 1042704). International Society for Optics and Photonics. <https://doi.org/10.1117/12.2278218>
- Marselis, S. M., Tang, H., Armston, J. D., Calders, K., Labrière, N., & Dubayah, R. (2018). Distinguishing vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: A case study in Lopé National Park, Gabon. *Remote Sensing of Environment*, 216(July), 626–634. <https://doi.org/10.1016/j.rse.2018.07.023>

- McFeeters, S. K. (1996). 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.
- Mendes, F. de S., Baron, D., Gerold, G., Liesenberg, V., & Erasmi, S. (2019). Optical and SAR remote sensing synergism for mapping vegetation types in the endangered Cerrado/Amazon ecotone of Nova Mutum-Mato Grosso. *Remote Sensing*, 11(10). <https://doi.org/10.3390/rs11101161>
- Meyfroidt, P., Carlson, K. M., Fagan, M. E., Gutiérrez-Vélez, V. H., Macedo, M. N., Curran, L. M., et al. (2014). Multiple pathways of commodity crop expansion in tropical forest landscapes. *Environmental Research Letters*, 9(7). <https://doi.org/10.1088/1748-9326/9/7/074012>
- Milne, S. (2015). Cambodia's Unofficial Regime of Extraction: Illicit Logging in the Shadow of Transnational Governance and Investment. *Critical Asian Studies*, 47(2), 200–228. <https://doi.org/10.1080/14672715.2015.1041275>
- Mulia, R., Hoan Trong Do, Van Thanh Pham, T. Q. N., Sonya Dewi, Andre Ekadinata, Adrian Dwiputra, Alfa Nugraha, F. J., & Khanh Quoc Nguyen, D. K. T. N. (2019). *Green Growth Action Plan for Lam Dong Province for the Period of 2021 – 2030, Vision to 2050*. Ha Noi.
- Neef, A., Touch, S., & Chiengthong, J. (2013). The Politics and Ethics of Land Concessions in Rural Cambodia. *Journal of Agricultural and Environmental Ethics*, 26(6), 1085–1103. <https://doi.org/10.1007/s10806-013-9446-y>
- Nomura, K., & Mitchard, E. T. A. (2018). More than meets the eye: Using Sentinel-2 to map small plantations in complex forest landscapes. *Remote Sensing*, 10(11). <https://doi.org/10.3390/rs10111693>
- Poortinga, A., Tenneson, K., Shapiro, A., Nquyen, Q., Aung, K. S., Chishtie, F., & Saah, D. (2019). Mapping plantations in Myanmar by fusing Landsat-8, Sentinel-2 and Sentinel-1 data along with systematic error quantification. *Remote Sensing*, 11(7), 1–19. <https://doi.org/10.3390/rs11070831>

- Portillo-Quintero, C., Hernández-Stefanoni, J. L., Reyes-Palomeque, G., & Subedi, M. R. (2021). The road to operationalization of effective tropical forest monitoring systems. *Remote Sensing*, 13(7), 1–9. <https://doi.org/10.3390/rs13071370>
- Potapov, P., Tyukavina, A., Turubanova, S., Talero, Y., Hernandez-Serna, A., Hansen, M. C., et al. (2019). Annual continuous fields of woody vegetation structure in the Lower Mekong region from 2000-2017 Landsat time-series. *Remote Sensing of Environment*, 232(June), 111278. <https://doi.org/10.1016/j.rse.2019.111278>
- Potapov, Peter, Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., et al. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, 253(November 2020), 112165. <https://doi.org/10.1016/j.rse.2020.112165>
- R Core Team. (2021). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from <https://www.r-project.org/>
- Ramankutty, N., Mehrabi, Z., Waha, K., Jarvis, L., Kremen, C., Herrero, M., & Rieseberg, L. H. (2018). Trends in Global Agricultural Land Use: Implications for Environmental Health and Food Security. *Annual Review of Plant Biology*, 69(1), 789–815. <https://doi.org/10.1146/annurev-arplant-042817-040256>
- Reis, M. S., Dutra, L. V., Escada, M. I. S., & Sant’Anna, S. J. S. (2020). Avoiding invalid transitions in land cover trajectory classification with a compound maximum a posteriori approach. *IEEE Access*, 8, 98787–98799. <https://doi.org/10.1109/ACCESS.2020.2997019>
- Rock, B. N., Vogelmann, J. E., Williams, D. L., Vogelmann, A. F., & Hoshizaki, T. (1986). Remote Detection of Forest Damage. *BioScience*, 36(7), 439–445. <https://doi.org/10.2307/1310339>
- Roe, S., Streck, C., Obersteiner, M., Frank, S., Griscom, B., Drouet, L., et al. (2019). Contribution of the land sector to a 1.5 °C world. *Nature Climate Change*, 9(11), 817–828. <https://doi.org/10.1038/s41558-019-0591-9>
- Roeun, V. (2017, April). New Land Law Would Levy Fines for Fallow Farms: Minister. *The*

Cambodia Daily. Retrieved from <https://english.cambodiadaily.com/news/new-land-law-would-levy-fines-for-fallow-farms-minister-127478/>

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. In *Third Earth Resources Technology Satellite-I Symposium Volume I: Technical Presentations Section A* (pp. 309–317). Washington, D.C.: National Aeronautics and Space Administration. <https://doi.org/10.1021/jf60203a024>

Saah, D., Tenneson, K., Matin, M., Uddin, K., Cutter, P., Poortinga, A., et al. (2019). Land Cover Mapping in Data Scarce Environments: Challenges and Opportunities. *Frontiers in Environmental Science*, 7(November). <https://doi.org/10.3389/fenvs.2019.00150>

Saah, D., Tenneson, K., Poortinga, A., Nguyen, Q., Chishtie, F., Aung, K. S., et al. (2020). Primitives as building blocks for constructing land cover maps. *International Journal of Applied Earth Observation and Geoinformation*, 85(June 2019), 101979. <https://doi.org/10.1016/J.JAG.2019.101979>

Sabins, F. F. (1997). *Remote Sensing: Principles and Interpretation* (3rd ed.). New York: W.H. Freeman and Co.

Sánchez-Cuervo, A. M., de Lima, L. S., Dallmeier, F., Garate, P., Bravo, A., & Vanthomme, H. (2020). Twenty years of land cover change in the southeastern Peruvian Amazon: implications for biodiversity conservation. *Regional Environmental Change*, 20(1). <https://doi.org/10.1007/s10113-020-01603-y>

Schott, J. R., Salvaggio, C., & Volchok, W. J. (1988). Radiometric Scene Normalization Using Pseudoinvariant Features. *Remote Sensing of Environment*, 16, 1–16. Retrieved from https://ac.els-cdn.com/0034425788901162/1-s2.0-0034425788901162-main.pdf?_tid=7a0b8319-e802-4130-acbb-b1d7d3f77f21&acdnat=1547269812_9b544851bbd42aeb1544eb31a99ba1ff

Sentinel Hub. (2021). s2cloudless. Retrieved from <https://github.com/sentinel-hub/sentinel2-cloud-detector>

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., & Lucas, R. (2014).

- New global forest/non-forest maps from ALOS PALSAR data (2007-2010). *Remote Sensing of Environment*, 155, 13–31. <https://doi.org/10.1016/j.rse.2014.04.014>
- Singh, M. (2020). Evaluating the impact of future climate and forest cover change on the ability of Southeast (SE) Asia’s protected areas to provide coverage to the habitats of threatened avian species. *Ecological Indicators*, 114(July 2019), 106307. <https://doi.org/10.1016/j.ecolind.2020.106307>
- Singh, M., Evans, D., Chevance, J. B., Tan, B. S., Wiggins, N., Kong, L., & Sakhoeun, S. (2019). Evaluating remote sensing datasets and machine learning algorithms for mapping plantations and successional forests in Phnom Kulen National Park of Cambodia. *PeerJ*, 2019(10). <https://doi.org/10.7717/peerj.7841>
- Small, D., Rohner, C., Miranda, N., Ruetschi, M., & Schaepman, M. E. (2021). Wide-Area Analysis-Ready Radar Backscatter Composites. *IEEE Transactions on Geoscience and Remote Sensing*, 1–14. <https://doi.org/10.1109/TGRS.2021.3055562>
- Soares-Filho, B., Alencar, A., Nepstad, D., Cerqueira, G., Del Carmen Vera Diaz, M., Rivero, S., et al. (2004). Simulating the response of land-cover changes to road paving and governance along a major Amazon highway: The Santarém-Cuiabá corridor. *Global Change Biology*, 10(5), 745–764. <https://doi.org/10.1111/j.1529-8817.2003.00769.x>
- Somvanshi, S. S., & Kumari, M. (2020). Comparative analysis of different vegetation indices with respect to atmospheric particulate pollution using sentinel data. *Applied Computing and Geosciences*, 7(June), 100032. <https://doi.org/10.1016/j.acags.2020.100032>
- Song, X. P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., & Townshend, J. R. (2018). Global land change from 1982 to 2016. *Nature*, 560(7720), 639–643. <https://doi.org/10.1038/s41586-018-0411-9>
- Tellman, B., Magliocca, N. R., Turner, B. L., & Verburg, P. H. (2020). Understanding the role of illicit transactions in land-change dynamics. *Nature Sustainability*, 3(March), 175–181. <https://doi.org/10.1038/s41893-019-0457-1>
- Traganos, D., Poursanidis, D., Aggarwal, B., Chrysoulakis, N., & Reinartz, P. (2018). Estimating

satellite-derived bathymetry (SDB) with the Google Earth Engine and sentinel-2. *Remote Sensing*, 10(6), 1–18. <https://doi.org/10.3390/rs10060859>

Venkatappa, M., Sasaki, N., Shrestha, R. P., Tripathi, N. K., & Ma, H. O. (2019). Determination of vegetation thresholds for assessing land use and land use changes in Cambodia using the Google Earth Engine cloud-computing platform. *Remote Sensing*, 11(13). <https://doi.org/10.3390/rs11131514>

Wales, N. A. (2020). An examination of forest cover change at Angkor, Cambodia, using satellite imagery, interviews and interpretation of historical events. *Applied Geography*, 122(February 2019), 102276. <https://doi.org/10.1016/j.apgeog.2020.102276>

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., et al. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688), 333–336. <https://doi.org/10.1038/nature25181>

Wohlfart, C., Wegmann, M., & Leimgruber, P. (2014). Mapping threatened dry deciduous dipterocarp forest in South-east Asia for conservation management. *Tropical Conservation Science*, 7(4), 597–613. <https://doi.org/10.1177/194008291400700402>

Appendix

A.1. A list of Sentinel scenes for compositing in land cover classification

a. Sentinel 2

No.	PRODUCT_ID	landCover_year
1	S2A_MSIL2A_20151201T033112_N9999_R018_T48PUA_20210417T004805	2015
2	S2A_MSIL2A_20151201T033112_N9999_R018_T48PUV_20210417T010921	2015
3	S2A_MSIL2A_20151201T033112_N9999_R018_T48PVA_20210417T013414	2015
4	S2A_MSIL2A_20151201T033112_N9999_R018_T48PVV_20210417T013504	2015
5	S2A_MSIL2A_20151201T033112_N9999_R018_T48PWA_20210417T013544	2015
6	S2A_MSIL2A_20151208T033242_N9999_R118_T48PVA_20210417T014909	2015
7	S2A_MSIL2A_20151208T033242_N9999_R118_T48PVV_20210417T014945	2015
8	S2A_MSIL2A_20151208T033242_N9999_R118_T48PWA_20210417T015022	2015
9	S2A_MSIL2A_20151208T033242_N9999_R118_T48PWV_20210417T021031	2015
10	S2A_MSIL2A_20151211T034152_N9999_R018_T48PUA_20210417T023144	2015
11	S2A_MSIL2A_20151211T034152_N9999_R018_T48PUV_20210417T025329	2015
12	S2A_MSIL2A_20151211T034152_N9999_R018_T48PVA_20210417T031718	2015
13	S2A_MSIL2A_20151211T034152_N9999_R018_T48PWA_20210417T031802	2015
14	S2A_MSIL2A_20151218T032142_N9999_R118_T48PVV_20210417T033327	2015
15	S2A_MSIL2A_20151218T032142_N9999_R118_T48PWV_20210417T033358	2015
16	S2A_MSIL2A_20151221T034152_N9999_R018_T48PUA_20210417T034839	2015
17	S2A_MSIL2A_20151221T034152_N9999_R018_T48PUV_20210417T040908	2015
18	S2A_MSIL2A_20151221T034152_N9999_R018_T48PWA_20210417T043054	2015
19	S2A_MSIL2A_20151228T033242_N9999_R118_T48PVA_20210417T044322	2015
20	S2A_MSIL2A_20151228T033242_N9999_R118_T48PVV_20210417T044359	2015
21	S2A_MSIL2A_20151228T033242_N9999_R118_T48PWA_20210417T044435	2015
22	S2A_MSIL2A_20151228T033242_N9999_R118_T48PWV_20210417T050423	2015
23	S2A_MSIL2A_20151231T033142_N9999_R018_T48PUA_20210417T052408	2015
24	S2A_MSIL2A_20151231T033142_N9999_R018_T48PUV_20210417T054353	2015
25	S2A_MSIL2A_20151231T033142_N9999_R018_T48PVA_20210417T060351	2015
26	S2A_MSIL2A_20151231T033142_N9999_R018_T48PVV_20210417T060435	2015
27	S2A_MSIL2A_20151231T033142_N9999_R018_T48PWA_20210417T060516	2015
28	S2A_MSIL2A_20160107T033242_N9999_R118_T48PVA_20210417T061736	2015
29	S2A_MSIL2A_20160107T033242_N9999_R118_T48PWA_20210417T061809	2015
30	S2A_MSIL2A_20160107T033242_N9999_R118_T48PWV_20210417T063855	2015
31	S2A_MSIL2A_20160110T034152_N9999_R018_T48PWA_20210417T065928	2015
32	S2A_MSIL2A_20160120T034152_N9999_R018_T48PVA_20210417T071129	2015
33	S2A_MSIL2A_20160120T034152_N9999_R018_T48PWA_20210417T071215	2015
34	S2A_MSIL2A_20160130T033002_N9999_R018_T48PUA_20210417T072457	2015
35	S2A_MSIL2A_20160130T033002_N9999_R018_T48PUV_20210417T072459	2015

36	S2A_MSIL2A_20160130T033002_N9999_R018_T48PVA_20210417T074608	2015
37	S2A_MSIL2A_20160130T033002_N9999_R018_T48PVV_20210417T074648	2015
38	S2B_MSIL2A_20191202T032109_N0213_R118_T48PVA_20191202T072407	2019
39	S2B_MSIL2A_20191202T032109_N0213_R118_T48PVB_20191202T072407	2019
40	S2B_MSIL2A_20191202T032109_N0213_R118_T48PVV_20191202T072407	2019
41	S2B_MSIL2A_20191202T032109_N0213_R118_T48PWA_20191202T072407	2019
42	S2B_MSIL2A_20191202T032109_N0213_R118_T48PWB_20191202T072407	2019
43	S2B_MSIL2A_20191202T032109_N0213_R118_T48PWV_20191202T072407	2019
44	S2B_MSIL2A_20191205T033119_N0213_R018_T48PUA_20191205T070306	2019
45	S2B_MSIL2A_20191205T033119_N0213_R018_T48PUB_20191205T070306	2019
46	S2B_MSIL2A_20191205T033119_N0213_R018_T48PUV_20191205T070306	2019
47	S2B_MSIL2A_20191205T033119_N0213_R018_T48PVA_20191205T070306	2019
48	S2B_MSIL2A_20191205T033119_N0213_R018_T48PVB_20191205T070306	2019
49	S2B_MSIL2A_20191205T033119_N0213_R018_T48PVV_20191205T070306	2019
50	S2B_MSIL2A_20191205T033119_N0213_R018_T48PWA_20191205T070306	2019
51	S2B_MSIL2A_20191205T033119_N0213_R018_T48PWB_20191205T070306	2019
52	S2A_MSIL2A_20191207T032121_N0213_R118_T48PVA_20191207T071912	2019
53	S2A_MSIL2A_20191207T032121_N0213_R118_T48PVB_20191207T071912	2019
54	S2A_MSIL2A_20191207T032121_N0213_R118_T48PVV_20191207T071912	2019
55	S2A_MSIL2A_20191207T032121_N0213_R118_T48PWA_20191207T071912	2019
56	S2A_MSIL2A_20191207T032121_N0213_R118_T48PWB_20191207T071912	2019
57	S2A_MSIL2A_20191207T032121_N0213_R118_T48PWV_20191207T071912	2019
58	S2A_MSIL2A_20191210T033131_N0213_R018_T48PUA_20191210T070124	2019
59	S2A_MSIL2A_20191210T033131_N0213_R018_T48PUB_20191210T070124	2019
60	S2A_MSIL2A_20191210T033131_N0213_R018_T48PUV_20191210T070124	2019
61	S2A_MSIL2A_20191210T033131_N0213_R018_T48PVA_20191210T070124	2019
62	S2A_MSIL2A_20191210T033131_N0213_R018_T48PVB_20191210T070124	2019
63	S2A_MSIL2A_20191210T033131_N0213_R018_T48PVV_20191210T070124	2019
64	S2A_MSIL2A_20191210T033131_N0213_R018_T48PWA_20191210T070124	2019
65	S2A_MSIL2A_20191210T033131_N0213_R018_T48PWB_20191210T070124	2019
66	S2B_MSIL2A_20191212T032129_N0213_R118_T48PVA_20191212T065639	2019
67	S2B_MSIL2A_20191212T032129_N0213_R118_T48PVB_20191212T065639	2019
68	S2B_MSIL2A_20191212T032129_N0213_R118_T48PVV_20191212T065639	2019
69	S2B_MSIL2A_20191212T032129_N0213_R118_T48PWA_20191212T065639	2019
70	S2B_MSIL2A_20191212T032129_N0213_R118_T48PWB_20191212T065639	2019
71	S2B_MSIL2A_20191212T032129_N0213_R118_T48PWV_20191212T065639	2019
72	S2B_MSIL2A_20191215T033139_N0213_R018_T48PUA_20191215T073452	2019
73	S2B_MSIL2A_20191215T033139_N0213_R018_T48PUB_20191215T073452	2019
74	S2B_MSIL2A_20191215T033139_N0213_R018_T48PUV_20191215T073452	2019
75	S2B_MSIL2A_20191215T033139_N0213_R018_T48PVA_20191215T073452	2019
76	S2B_MSIL2A_20191215T033139_N0213_R018_T48PVB_20191215T073452	2019
77	S2B_MSIL2A_20191215T033139_N0213_R018_T48PVV_20191215T073452	2019
78	S2B_MSIL2A_20191215T033139_N0213_R018_T48PWA_20191215T073452	2019

79	S2B_MSIL2A_20191215T033139_N0213_R018_T48PWB_20191215T073452	2019
80	S2A_MSIL2A_20191217T032131_N0213_R118_T48PVA_20191217T064024	2019
81	S2A_MSIL2A_20191217T032131_N0213_R118_T48PVB_20191217T064024	2019
82	S2A_MSIL2A_20191217T032131_N0213_R118_T48PVV_20191217T064024	2019
83	S2A_MSIL2A_20191217T032131_N0213_R118_T48PWA_20191217T064024	2019
84	S2A_MSIL2A_20191217T032131_N0213_R118_T48PWB_20191217T064024	2019
85	S2A_MSIL2A_20191217T032131_N0213_R118_T48PWV_20191217T064024	2019
86	S2A_MSIL2A_20191220T033141_N0213_R018_T48PUA_20191220T070111	2019
87	S2A_MSIL2A_20191220T033141_N0213_R018_T48PUB_20191220T070111	2019
88	S2A_MSIL2A_20191220T033141_N0213_R018_T48PUV_20191220T070111	2019
89	S2A_MSIL2A_20191220T033141_N0213_R018_T48PVA_20191220T070111	2019
90	S2A_MSIL2A_20191220T033141_N0213_R018_T48PVB_20191220T070111	2019
91	S2A_MSIL2A_20191220T033141_N0213_R018_T48PVV_20191220T070111	2019
92	S2A_MSIL2A_20191220T033141_N0213_R018_T48PWA_20191220T070111	2019
93	S2A_MSIL2A_20191220T033141_N0213_R018_T48PWB_20191220T070111	2019
94	S2B_MSIL2A_20191222T032139_N0213_R118_T48PVA_20191222T072524	2019
95	S2B_MSIL2A_20191222T032139_N0213_R118_T48PVB_20191222T072524	2019
96	S2B_MSIL2A_20191222T032139_N0213_R118_T48PVV_20191222T072524	2019
97	S2B_MSIL2A_20191222T032139_N0213_R118_T48PWA_20191222T072524	2019
98	S2B_MSIL2A_20191222T032139_N0213_R118_T48PWB_20191222T072524	2019
99	S2B_MSIL2A_20191222T032139_N0213_R118_T48PWV_20191222T072524	2019
100	S2B_MSIL2A_20191225T033139_N0213_R018_T48PUA_20191225T064801	2019
101	S2B_MSIL2A_20191225T033139_N0213_R018_T48PUB_20191225T064801	2019
102	S2B_MSIL2A_20191225T033139_N0213_R018_T48PUV_20191225T064801	2019
103	S2B_MSIL2A_20191225T033139_N0213_R018_T48PVA_20191225T064801	2019
104	S2B_MSIL2A_20191225T033139_N0213_R018_T48PVB_20191225T064801	2019
105	S2B_MSIL2A_20191225T033139_N0213_R018_T48PVV_20191225T064801	2019
106	S2B_MSIL2A_20191225T033139_N0213_R018_T48PWA_20191225T064801	2019
107	S2B_MSIL2A_20191225T033139_N0213_R018_T48PWB_20191225T064801	2019
108	S2A_MSIL2A_20191227T032131_N0213_R118_T48PVA_20191227T065256	2019
109	S2A_MSIL2A_20191227T032131_N0213_R118_T48PVB_20191227T065256	2019
110	S2A_MSIL2A_20191227T032131_N0213_R118_T48PVV_20191227T065256	2019
111	S2A_MSIL2A_20191227T032131_N0213_R118_T48PWA_20191227T065256	2019
112	S2A_MSIL2A_20191227T032131_N0213_R118_T48PWB_20191227T065256	2019
113	S2A_MSIL2A_20191227T032131_N0213_R118_T48PWV_20191227T065256	2019
114	S2A_MSIL2A_20191230T033141_N0213_R018_T48PUA_20191230T071815	2019
115	S2A_MSIL2A_20191230T033141_N0213_R018_T48PUB_20191230T071815	2019
116	S2A_MSIL2A_20191230T033141_N0213_R018_T48PUV_20191230T071815	2019
117	S2A_MSIL2A_20191230T033141_N0213_R018_T48PVA_20191230T071815	2019
118	S2A_MSIL2A_20191230T033141_N0213_R018_T48PVB_20191230T071815	2019
119	S2A_MSIL2A_20191230T033141_N0213_R018_T48PVV_20191230T071815	2019
120	S2A_MSIL2A_20191230T033141_N0213_R018_T48PWA_20191230T071815	2019
121	S2A_MSIL2A_20191230T033141_N0213_R018_T48PWB_20191230T071815	2019

b. Sentinel 1

No	PRODUCT_ID	landCover_year
1	S1A_IW_GRDH_1SDV_20151224T224411_20151224T224440_009190_00D3 D4_9187	2015
2	S1A_IW_GRDH_1SDV_20151224T224440_20151224T224505_009190_00D3 D4_4610	2015
3	S1A_IW_GRDH_1SDV_20151229T225219_20151229T225244_009263_00D5 E6_123B	2015
4	S1A_IW_GRDH_1SDV_20151229T225244_20151229T225309_009263_00D5 E6_6000	2015
5	S1A_IW_GRDH_1SDV_20191220T225234_20191220T225259_030438_037B D7_9BDB	2019
6	S1A_IW_GRDH_1SDV_20191220T225259_20191220T225324_030438_037B D7_1933	2019
7	S1A_IW_GRDH_1SDV_20191220T225324_20191220T225349_030438_037B D7_50CC	2019
8	S1A_IW_GRDH_1SDV_20191227T224427_20191227T224452_030540_037F5 E_4F99	2019
9	S1A_IW_GRDH_1SDV_20191227T224452_20191227T224517_030540_037F5 E_A755	2019

A.2. Formulas of the spectral indices used in land cover classification

Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974)

$$NDVI = \frac{NIR - red}{NIR + red}$$

Soil Adjusted Vegetation Index (SAVI) (A.R. Huete, 1988; Somvanshi & Kumari, 2020)

$$SAVI = \frac{NIR - red}{NIR + red + 0.5} \times 1.5$$

Normalized Difference Water Index 1 (NDWI) (Gao, 1996)

$$NDWI = \frac{NIR - SWIR1}{NIR + SWIR1}$$

Normalized Difference Water Index 2 (NDWI₂) (McFeeters, 1996)

$$NDWI_t = \frac{green - NIR}{green + NIR}$$

Normalized Burn Ratio (NBR) (Key et al., 2002)

$$NBR = \frac{NIR - SWIR2}{NIR + SWIR2}$$

Enhanced Vegetation Index (EVI) (Alfredo R. Huete et al., 1999)

$$EVI = \frac{NIR - red}{NIR + 6 \times red - 7.5 \times blue + 1} \times 2.5$$

Green Normalized Difference Vegetation Index (GNDVI) (Buschmann & Nagel, 1993)

$$GNDVI = \frac{NIR - green}{NIR + green}$$

Enhanced Vegetation Index 2 (EVI2) (Jiang et al., 2008)

$$EVI2 = \frac{NIR - red}{NIR + 2.4 \times red + 1} \times 2.5$$

Moisture Stress Index (MSI) (Rock et al., 1986)

$$MSI = \frac{SWIR1}{NIR}$$

Modified Chlorophyll Absorption in Reflectance Index (MCARI) (Daughtry et al., 2000)

$$MCARI = \frac{(red1 - red - 0.2 \times (red1 - green)) \times red1}{red}$$

Pigment Specific Simple Ratio (PSSR) (Blackburn, 1998)

$$PSSR = \frac{NIR}{red}$$

Associated Sentinel-2 bands

blue = band 2 of Sentinel-2 Level-2A median composite

green = band 3 of Sentinel-2 Level-2A median composite

red = band 4 of Sentinel-2 Level-2A median composite

red1 = band 5 of Sentinel-2 Level-2A median composite

NIR = band 8 of Sentinel-2 Level-2A median composite

SWIR1 = band 11 of Sentinel-2 Level-2A median composite

SWIR2 = band 12 of Sentinel-2 Level-2A median composite