



ENHANCING THE MEKONG RIVER COMMISSION'S LAND USE AND LAND COVER 2020 MAPPING PRODUCTS



The MRC is funded by contributions from its Member Countries and Development Partners, including Australia, the European Union, Finland, Flanders/Belgium, France, Germany, Japan, Luxembourg, the Netherlands, New Zealand, Sweden, Switzerland, and the United States of America.

Copyright © Mekong River Commission, 2023

First published (2023)

Some rights reserved.

This work is a product of the Mekong River Commission Secretariat (MRCS). While all efforts have been made to present accurate information, the Mekong River Commission (MRC) does not guarantee the accuracy of the data included in this work. The boundaries, colours, denomination and other information shown on any map in this work do not imply any judgement on the part of the MRC concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Nothing herein shall constitute or be considered to be a limitation upon or waiver of the privileges and immunities of the MRC, all of which are specifically reserved.

This publication may be reproduced, in whole or in part and in any form, for educational or non-profit purposes without special permission from the copyright holder provided that the MRC is acknowledged as the source and that notification is sent to the MRC. The MRCS would appreciate receiving a copy of any publication that uses this publication as a source. This publication cannot be used for sale or for any other commercial purpose whatsoever without permission in writing from the MRCS.

Title: Enhancing the MRC land use and land cover 2020 mapping products

DOI: <https://doi.org/10.52107/mrc.aqrsbr>

Keywords: land use/land cover/maps/Lower Mekong Basin

For bibliographic purposes, this volume may be cited as:

Mekong River Commission. (2023). *Enhancing the MRC land use and land cover 2020 mapping products*. Vientiane: Mekong River Commission Secretariat. <https://doi.org/10.52107/mrc.aqrsbr>

Information on MRC publications and digital products are available at:
www.mrcmekong.org/publications

All queries on rights and licenses should be addressed to:
Mekong River Commission
Documentation and Learning Centre
184 Fa Ngoum Road, Unit 18, Ban Sithane Neua, Sikhottabong District, Vientiane 01000, Lao PDR
Telephone: +856-21 263 263 | E-mail: mrcs@mrcmekong.org | www.mrcmekong.org

Citation

Mekong River Commission. (2023). *Enhancing the MRC land use and land cover 2020 mapping products*. Vientiane: Mekong River Commission Secretariat.

<https://doi.org/10.52107/mrc.aqrsbr>

Authors

Project management

Dr Winai Wangpimool, Director of Technical Support Division (2019–2022)

Mr Tran Minh Khoi, Director of Technical Support Division (2022–Now)

Technical experts

Mekong River Commission Secretariat

Mr Kittiphong Phongsapan, Remote Sensing and GIS Specialist

National technical experts

Mr Teng Peng Seang, Phnom Penh Geoinformatics Education Center, Cambodia

Mr Soukkanh bounthabandid, Forest Inventory and Planning Division, Department of Forestry, Lao PDR

Dr Totsanat, Land Development Department, Ministry of Agriculture and Cooperative, Thailand

Ms Pham Thi Nhi, Sub-National Institute of Agricultural Planning and Projection (Sub-NIAPP), Viet Nam

International technical expert

Dr Ate Poortinga, SERVIR-Mekong, Thailand

Contents

Figures	vi
Tables	vii
Abbreviations and acronyms	viii
Executive Summary	ix
1 Introduction	1
1.1 Background.....	1
1.2 Overall goal and objectives	2
1.3 Project outputs.....	3
2 Land cover category and materials	4
2.1 Land cover classification	4
2.2 Analysis of the MRC land change detection.....	5
2.3 Data used for LULC 2020 mapping.....	15
2.4 Google Earth Engine	23
3 Methodology on development LULC	24
3.1 Satellite image composites	25
3.2 Random forest classification and create primitives.....	29
3.3 Assemblage and land cover mapping	31
4 Results of the Land Use and Land Cover Map 2020	33
4.1 Satellite image catalogues and accessibilities.....	33
4.2 Probability classified primitive maps	33
4.3 Accuracy assessment of primitive models	34
4.4 Accuracy assessment for the LULC 2020 product.....	39
4.5 Updating of the land cover map of 2020	39
4.6 Land use change analysis	47
5 Conclusions and recommendations	49
5.1 Conclusions and recommendations.....	49
5.2 Lesson learned.....	49
Annex	51
Annex 1: Field Data Collection Form.....	51
Annex 2: GEE Code to Generate Satellite Image Composite and Download.....	54
Annex 3: Example of Field Data Collection Images.....	55

Annex 4: Access to Training Datasets	56
Annex 5: List of GEE Code of Primitive Models for the LULC Map of 2020	57
References	59

Figures

Figure 1. Land Cover Map 2010.....	2
Figure 2. Homogenized MRC Land Cover Distribution, 1997, 2003, 2010	7
Figure 3. Homogeneous types of land cover, 1997.....	8
Figure 4. Homogeneous type of land cover, 2003	9
Figure 5. Homogeneous types of land cover, 2010.....	10
Figure 6. Probability change maps for the ‘from’ category for the different land cover types	11
Figure 7. Probability change maps for the ‘to’ category for the different land cover types ..	12
Figure 8. Distribution of validation change pixels in the ‘Specific to Generic’ category.....	13
Figure 9. Distribution of validation change pixels in the ‘Generic to Specific’ category.....	13
Figure 10. Map of land cover change	14
Figure 11. Planned locations for field data collection	17
Figure 12. Field data collection by PGEC	18
Figure 13. Field data collection by FIPD.....	19
Figure 14. Field data collection map.....	21
Figure 15. Number of reference data points per class.....	21
Figure 16. Spatial distribution of reference data points per land cover class.....	22
Figure 17. The Google Earth Engine user interface	23
Figure 18. Satellite data processing.....	24
Figure 19. Development of primitives and assemblage	31
Figure 20. Decision tree used to generate the final land cover map	32
Figure 21. Probability distributions of 14 primitives from the Random Forest algorithm in the LMB.....	34
Figure 22. Probability distribution of primitive models (aquaculture and bamboo)	35
Figure 23. Probability distribution of primitive models (bare soil and cropland).....	35
Figure 24. Probability distribution of primitive models (deciduous and evergreen forest) ...	36
Figure 25. Probability distribution of primitive models (flooded forest and grassland).....	36
Figure 26. Probability distribution of primitive models (industrial plantation andmarsh/swamp)	37
Figure 27. Probability distribution of primitive models (orchard and paddy rice)	37
Figure 28. Probability distribution of primitive models (urban area and water body).....	38
Figure 29. LULC 2020 area	41
Figure 30. LULC map of 2020 in the Lower Mekong Basin.....	42
Figure 31. LULC of 2020 in Lower Mekong Basin for Cambodia	43
Figure 32. LULC map of 2020 in Lower Mekong Basin for Lao PDR	44
Figure 33. LULC of 2020 in Lower Mekong Basin for Thailand.....	45
Figure 34. LULC map of 2020 in Lower Mekong Basin for Viet Nam.....	46
Figure 35. Area comparison of land changing in 2003, 2010, and 2020.....	48

Tables

Table 1. Land cover categories definition	4
Table 2. Homogenized MRC land cover distribution, 1997, 2003, 2010.....	6
Table 3. Composite satellite imagery uses for the development of LULC 2020	15
Table 4. Detailed number of the hotspots (result) for targeting the field data collection	16
Table 5. Overview of the band combinations.....	27
Table 6. Confusion matrix for binary classification to predict croplands.....	30
Table 7. Confusion matrix of primitive models	38
Table 8. Confusion matrix for accuracy assessment	39
Table 9. LULC 2020 classification area.....	40
Table 10. Comparison of land changing in the LMB.....	47

Abbreviations and acronyms

API	Application programming interface
BDS	Basin Development Strategy
EO	Earth Observation
FAO	Food and Agriculture Organization of the United Nations
FN	False Negative
FP	False Positive
GEE	Google Earth Engine
GIS	Geographic Information System
IWRM	Integrated Water Resources Management
LCCS	Land Cover Classification System
LMB	Lower Mekong Basin
LULC	Land use and land cover
MRC	Mekong River Commission
MRC-IS	Mekong River Commission - Information System
SAR	Synthetic Aperture Radar
SP	Strategic Plan
TA	Target area
TN	True Negative
TP	True Positive

Executive Summary

To support Mekong River Commission (MRC) Strategic Plan 2021–2025 and Basin Development Strategy (BDS) 2021–2030, the MRC Secretariat (MRCS) has proposed the project to update the 2020 Land Cover and Land Use (LULC) map of the Mekong Basin by applying the most recent Earth Observation technology via the machine learning algorithm from the Google Earth Engine (GEE) platform, which was approved by the MRC Joint Committee. The LULC 2020 product has generated important baseline data that can benefit for water resource planning, floodplain management and other activities, such as hydrological and climate change studies together with environmental modelling.

This report summarizes the process starting from the background of the project, field data collection, and most importantly, the process of applying the EO technology via machine learning using the GEE platform and the approach agreed by the expert group from the Member Countries to update the LULC map of 2020 for the Lower Mekong Basin (LMB) region. A set of satellite images in 2020 including Landsat 8, Sentinel-1, and Sentinel-2 Planet image was utilized to prepare the LULC 2020 map. All of these satellite images were pre-processed with atmospheric correction, cloud removal, and topographic correction to create the image composite and were made them available for download for the Member Countries.

The method was customized to create the products using biophysical (primitive) layers according to the definition of land cover types of the MRC. Primitive layers are a suite of biophysical and end- member objects for land cover mapping, i.e. tree canopy cover and evergreen. Land cover primitives represent raw information needed to make decisions in a dichotomous key for land cover classification. The primitive layers for LULC of the LMB were created using machine learning models with reference data from field data collection and additional extra reference data collected by the SERVIR-Mekong team. By using these primitive layers as the input for the machine learning model into the decision tree logic with the Monte Carlo simulation, the LULC 2020 mapping was carried out. This technical report clearly details the process of data use, including the methodology used for developing the product.

The updated LULC 2020 data set is very important to support decision-making. The data, information and image interpretation are also useful for those engaged in studies and research in the LMB.

1 Introduction

1.1 Background

In 1997, the Mekong River Commission (MRC) completed a project to produce forest cover maps in the Lower Mekong Basin (LMB). In 2010, the MRC Secretariat (MRCS) implemented the updated land cover map 2009/2010 products, covering the LMB. The project aimed to generate new data sets of land cover map 2009/2010 products, both dry and wet season, including an annual data set. The project was implemented as a core activity of the former Information and Knowledge Management Programme during the MRC Strategic Plan 2011–2015.

Currently, the Mekong River Basin has been facing a multitude of challenges from climate changes to rapid development in recent decades, resulting in significant changes in land use/land cover (LULC). The LULC changes are caused by key factors, such as agricultural expansion and intensification, deforestation, increased river damming, increased urbanization, growing human populations, and the expansion of industrial forest plantations, as well as frequent natural disasters from both flooding and drought.

Therefore, it is crucial to deeply understand the updated status of the LULC change in the Mekong River Basin to particularly support the implementation of the MRC-Basin Development Strategies (BDS) and Strategic Plan (SP).

Under the MRC Strategic Plan 2016–2020, a set of core functions was implemented and activities carried out to identify practical knowledge on monitoring responsibilities, which provided an opportunity to build ownership and capacity among the Member Countries' agencies. One targeted outcome under this SP is the updating of the MRC's Land Use and Land Cover (LULC) Mapping 2020, under the activity, "Regularly update and maintain the MRC-IS and its tool and functionalities". It also links to the new MRC-BDS and SP 2021–2025, focusing on "BDS Output 4.1.2: Integrated data and information systems for more effective basin-wide data management and sharing".

To support the MRC BDS and SP 2021–2025, it is necessary to provide more informative services to Member Countries regarding the MRC SP activity: upgrade remote sensing and satellite imagery repository and develop the capacity of handling and using satellite products for water resource application. This will entail carrying out an inventory and prioritizing the remote sensing data and products for the development of the region's decision support systems, including the upgrading of MRC's Decision Support Framework. The work will be aligned with the implementation on reinvigorating the MRC's data, information, modelling, forecasting, and communication systems by focusing on the project to update the Land Cover and Land Use (LULC) Map of the Lower Mekong Basin of 2020 by applying the Earth Observatory (EO) technology via machine learning from the Google Earth Engine (GEE) platform by satellite-based LULC Mapping and Monitoring in the LMB, which was approved by the MRC Joint Committee Members.

The project generates important baseline data of LULC in LMB for water resource planning, floodplain management, and other activities such as hydrological and climate change studies, together with environmental modelling for the GIS database. Understanding the status of the up-to-date LULC data set is very significant to support decision-making and the BDS. The data, information, and image interpretation are also useful for those engaged in studies and research in the LMB.

A wide range of stakeholders recognizes the importance of LULC for sustaining human-dependent livelihoods and maintaining the ecological integrity of the Basin, as reflected in the widespread use of LULC types as the basis for agro-ecological zones. This implementation plan must also be an indispensable part of the knowledge base on the EO technology through the satellite data in the LMB, both of which should be regularly monitored for sustainable uses.

In this context, the MRC Secretariat implemented of the “Enhancing MRC’s Land Use/Land Cover Monitoring System to provide service for the LMB – Phase 2” for developing and updating LULC Map 2019/2020 of LMB region.

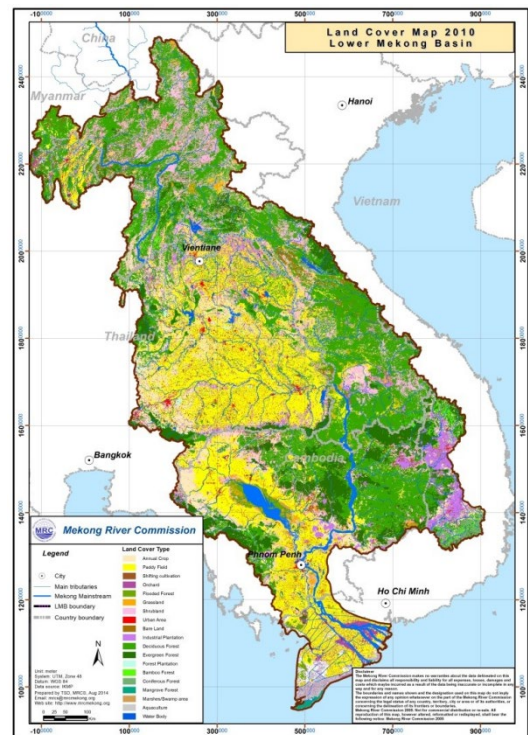


Figure 1. Land Cover Map 2010

1.2 Overall goal and objectives

The goal of the updating the MRC’s LULC Map 2020 products for the LMB by satellite-based approach is under the Activity 6.2.6, which is mainly aimed at implementing and generating new data sets of LULC map 2020, including their information catalogue on annual data sets for disseminating through the Mekong River Commission-Information System (MRC-IS).

The objectives include sharing reliable data for achieving the Sustainable Development Goals (SDGs), as well as necessary capacity building for Member Countries. Moreover, the EO technology and the remote sensing approach efficiently contributed the overall figures, which were the key information in the decision support system for all governmental and intergovernmental organizations for dimensions of Integrated Water Resources Management (IWRM), which will be considered a key achievement of the MRCS under SP 2016–2020.

The updated LULC Map 2020 products in the LMB will make it easier to better inform the dynamic state of the Basin in terms of agricultural, socio-economic, and climate adaptation sectors by building a tangible capacity in relevant national institutes by making a LULC Map 2020 products and evaluating the sustainability of water use in for all sectors as well as IWRM-based Basin Development Strategy. The updating LULC data set is very important to support decision-making. The data, information, and image interpretation are also useful for those

engaged in studies and research in the LMB. In addition, this report seeks to raise awareness among stakeholders of the significance and ecological functions of LULC in the LMB.

In addition, the project aims to update and improve MRC-IS by integrating the updated remote sensing and GIS tools into the MRC-IS. It will also establish a baseline for producing updated and qualified LULC data sets with a land use information catalogue for basin-wide management. The main objectives are to collect, compile, and provide accurate and reliable information to help interpret satellite imagery by applying the EO technology via a machine learning from the GEE platform, conduct achievable accuracy assessments, and establish a land use information catalogue for the LMB and disseminate it via MRC-IS.

1.3 Project outputs

The project aims to improve the information system of the MRC and will also establish a baseline for producing an updated and qualified LULC data set with land use information catalogue for the basin-wide management. The main outputs are:

- a data set of 2020 satellite images, which covers Lower Mekong Basin (LMB) consisting of Landsat-8, Sentinel-2, Sentinel-1, and Planet images;
- a GEE script to generate annual image composite for Landsat-8, Sentinel-1, and Sentinel-2;
- an updated LULC map for 2020 for the LMB region;
- a GEE script to generate the LULC 2020 map.

2 Land cover category and materials

2.1 Land cover classification

Land cover is defined by the Food and Agriculture Organization of the United Nations (FAO) as “the observed (bio) physical cover on the earth's surface”. This is not to be confused with land use, which is characterized by the people’s arrangements, activities, and inputs in a certain land cover type to produce, change or maintain it.

Potential applications for up-to-date land cover data include fish production models, development planning, wetlands monitoring, climate change scenarios, as well as flood and land management. The only means for efficiently producing land cover maps of large areas such as the LMB is to draw on remote sensing data and technologies for interpreting these data. Various land cover classification systems based on remote sensing data are available, some prominent examples of which are:

- CORINE Land Cover by the European Environmental Agency (EEA);
- FAO’s and United Nations Environment Programme’s Land Cover Classification System (LCCS);
- the Land Cover Classification System of the US Geological Survey (USGS).

The LCCS, which was applied in updating the land cover maps of the LMB, is currently the only universally applicable classification system in operational use. The Technical Committee ISO/TC211 has therefore adopted LCCS Version 3 as the ISO 19144 standard, which consists of two parts namely, the Classification System Structure and the Land Cover Classification System. LCCS has been set up to allow for a high degree of flexibility and is applicable in all climatic zones and environmental conditions. It is also compatible with other classification systems and has been widely adopted at the national level throughout Africa, Asia, the Near East, and Latin America. MRC could draw on the experience of previous land cover classification projects that had been carried out in cooperation with FAO, the last of which was completed in 2003.

The project to update the Land Cover and Land Use Map of the Lower Mekong Basin of 2020 was designed to produce three data sets based on the single-class land cover based on the FAO LCCS, which the MRC has followed since 2010 (Kityuttachai, 2016), as shown in Table 1.

Table 1. Land cover categories definition

LCC	Land Cover Types	Description
4	Annual crop	Mandatory herbaceous growth forms; cultivated and managed vegetation
5	Paddy field	Mandatory Gramineae; cultivated and managed vegetation with field size:0.2-2.0 ha; species of rice
6	Shifting cultivation	Temporary sequence between herbaceous growth forms of cultivated and managed vegetation and woody growth forms of Natural / semi-natural vegetation; sequence length: 3 to 7 years

7	Orchard	Mandatory trees; cultivated and managed vegetation of orchard and other plantation; field size:1–3 ha
8	Flooded forest	Multi-stratum of mandatory trees/shrubs/herbaceous; natural or semi-natural vegetation; cover 10–100%.
9	Grassland	Mandatory herbaceous growth forms; Natural or Semi Natural vegetation; cover 10–100%
10	Shrubland	Mandatory shrubs; natural or semi-natural vegetation; cover: 10–80%
11	Urban area	Mandatory built-up surface
12	Bare land	Mandatory bare soil; optional coarse mineral fragments (stone:1–40%)
13	Industrial plantation	Mandatory of woody growth form; cultivated and managed vegetation of orchard and other plantation; species of industrial crops
14	Deciduous forest	Mandatory trees; cover 10–100%; natural or semi-natural vegetation; deciduous and broadleaved
15	Evergreen forest	Mandatory trees; cover 10–100%; natural or semi-natural vegetation; evergreen
16	Forest plantation	Mandatory trees; cultivated and managed vegetation of forest plantation; broadleaved
17	Bamboo forest	Mandatory woody growth forms; cover: 10–100%; height: 4–15 m; natural or semi-natural vegetation; species of bamboo
18	Coniferous forest	Mandatory trees; cover 10–100%; natural or semi-natural vegetation; evergreen and needle leaved; species of coniferous sp.
19	Mangrove forest	Mandatory woody growth forms; cover: 10–100%; natural or semi-natural vegetation; broadleaved and evergreen.
20	Marsh/swamp area	Mandatory herbaceous growth forms
21	Aquaculture	Mandatory artificial water body above surface; general aquaculture
23	Water body	Mandatory periodic variations water body above surface; fresh water

2.2 Analysis of the MRC land change detection

In order to support the LULC 2020 product development for the component of field data collection, the location for the field data collection study was identified during the phase 1 (2019) of the project which technical collaboration with the Asian Disaster Preparedness Center (ADPC).

The result of the analysis is presented in two parts. In the first part, the analysis focuses on comparing the change of the land based on the MRC land cover archive data of 1997, 2003, and 2010. The second part, the analysis focus on the hotspot and change detection analysis for 2000–2018. The LULC map, which was used for the analysis, was produced with the Regional Land Cover Monitoring System (RLCMS) tool developed by SERVIR-Mekong, ADPC.¹ The result of the analysis shows the map of land cover change, which was divided into different levels indicating the degree of the probability of land use changes.

2.2.1 Spatial and temporal change in MRC land cover

The land cover distribution for 1997, 2003, and 2010 derived from the maps is presented in Table 2.

¹ USAID. Regional Land Cover Monitoring System. <https://servir.adpc.net/tools/regional-land-cover-monitoring-system>

Table 2. Homogenized MRC land cover distribution, 1997, 2003, 2010

Land cover type	Area 1997		Area 2003		Area 2010	
	km ²	Percentage	km ²	Percentage	km ²	Percentage
Agriculture	252,440	40.88	42,217	6.84	51,560	8.35
Aquaculture			2,071	0.34	6,828	1.11
Bamboo forest	3,647	0.59	9,202	1.49	5,585	0.90
Bare soil	2,385	0.39	2,843	0.46	3,804	0.62
Flooded forest	3,828	0.62	4,510	0.73	4,881	0.79
Forest	215,770	34.94	314,050	50.86	243,757	39.47
Forest plantation	1,690	0.27	490	0.08	1,506	0.24
Grassland	11,649	1.89	13,935	2.26	8,298	1.34
Mangrove	342	0.06	1,808	0.29	1,099	0.18
Marsh/swamp area	1,461	0.24	914	0.15	2,600	0.42
Orchard/industrial plantation			8,451	1.37	37,182	6.02
Other	3,413	0.55	163	0.03		0.00
Paddy rice			154,359	25.00	140,207	22.70
Shifting cultivation			13,639	2.21	10,681	1.73
Shrubland	109,980	17.81	21,057	3.41	69,525	11.26
Urban area	697	0.11	15,716	2.54	15,785	2.56
Water body	10,230	1.66	12,107	1.96	14,234	2.30

Figure 2 represents the static area of each land cover category for each study year, i.e. 1997, 2003 and 2010. In 1997, agricultures occupied the largest percentage of land cover, at 41% . However, part of this agriculture category is classified as paddy rice, with an area of 25% and 23% in 2003 and 2010, respectively. From 2003 to 2010, there was a slight increase in area of agriculture, from 6% to 8%.

It can be observed that the urban area increased significantly, from 0.11% in 1997, to 2.5% in 2003, and to 2.6% in 2010. This is due to the urban growth and development, which consists of different aspects or dimensions, such as population growth, physical growth, and economic growth, especially in developing countries.

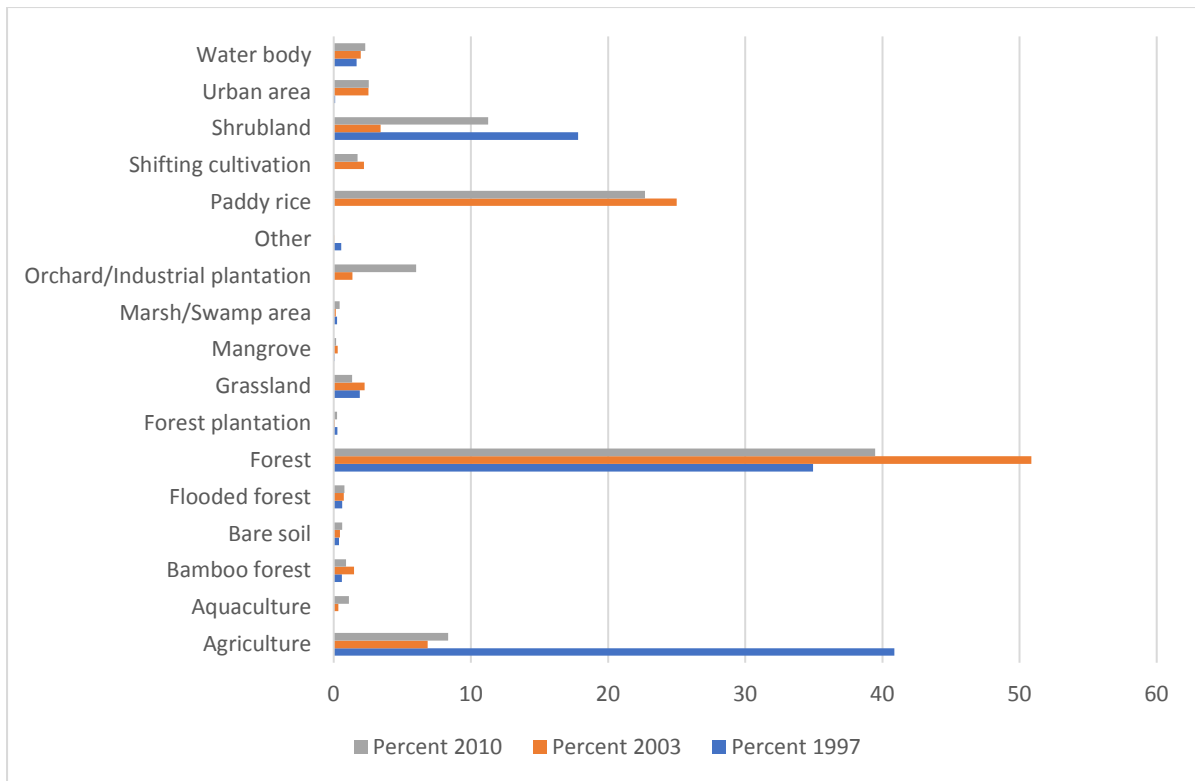


Figure 2. Homogenized MRC land cover distribution, 1997, 2003, 2010

Land cover 1997

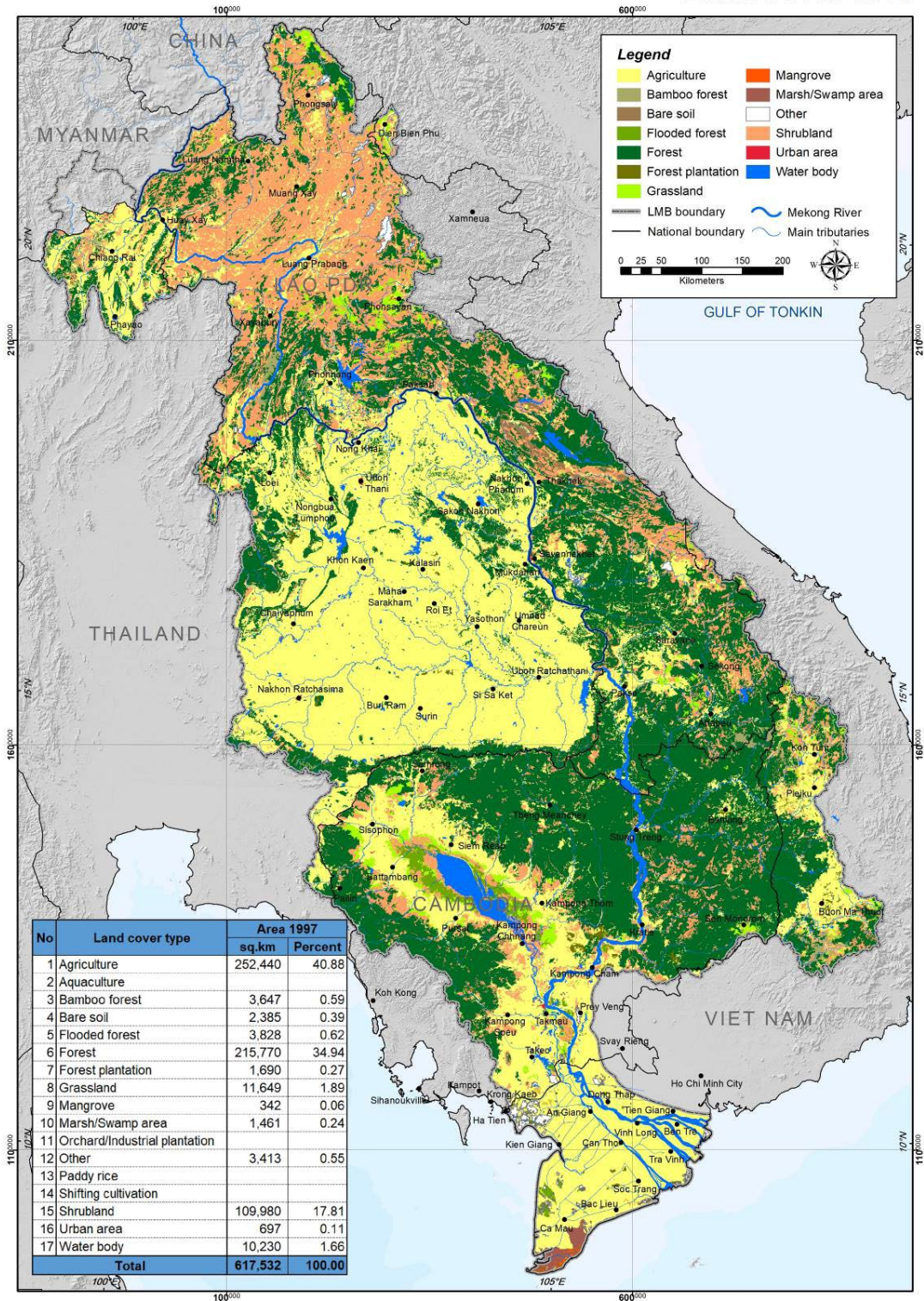


Figure 3. Homogeneous types of land cover, 1997

Land cover 2003

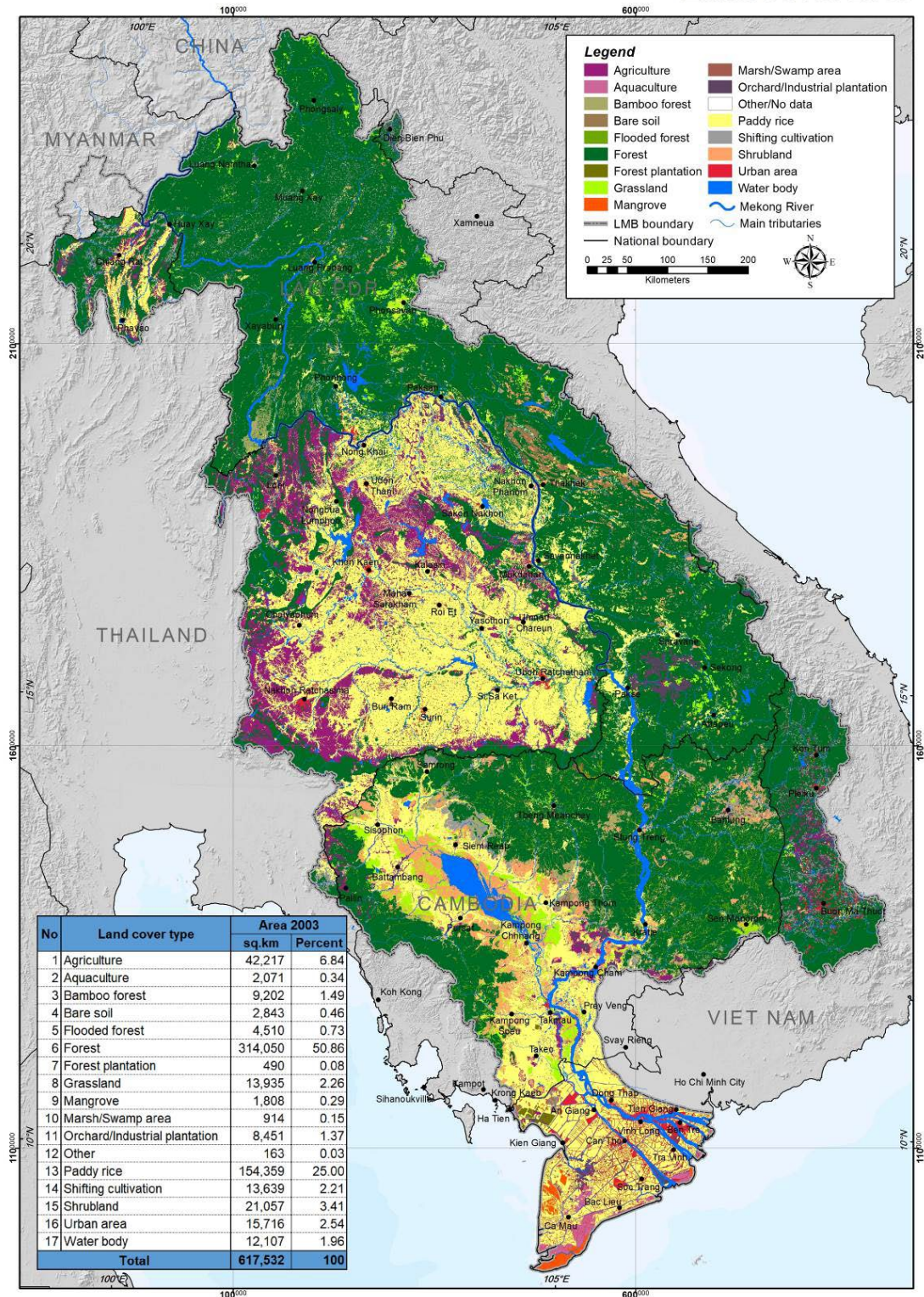


Figure 4. Homogeneous type of land cover, 2003

Land cover 2010

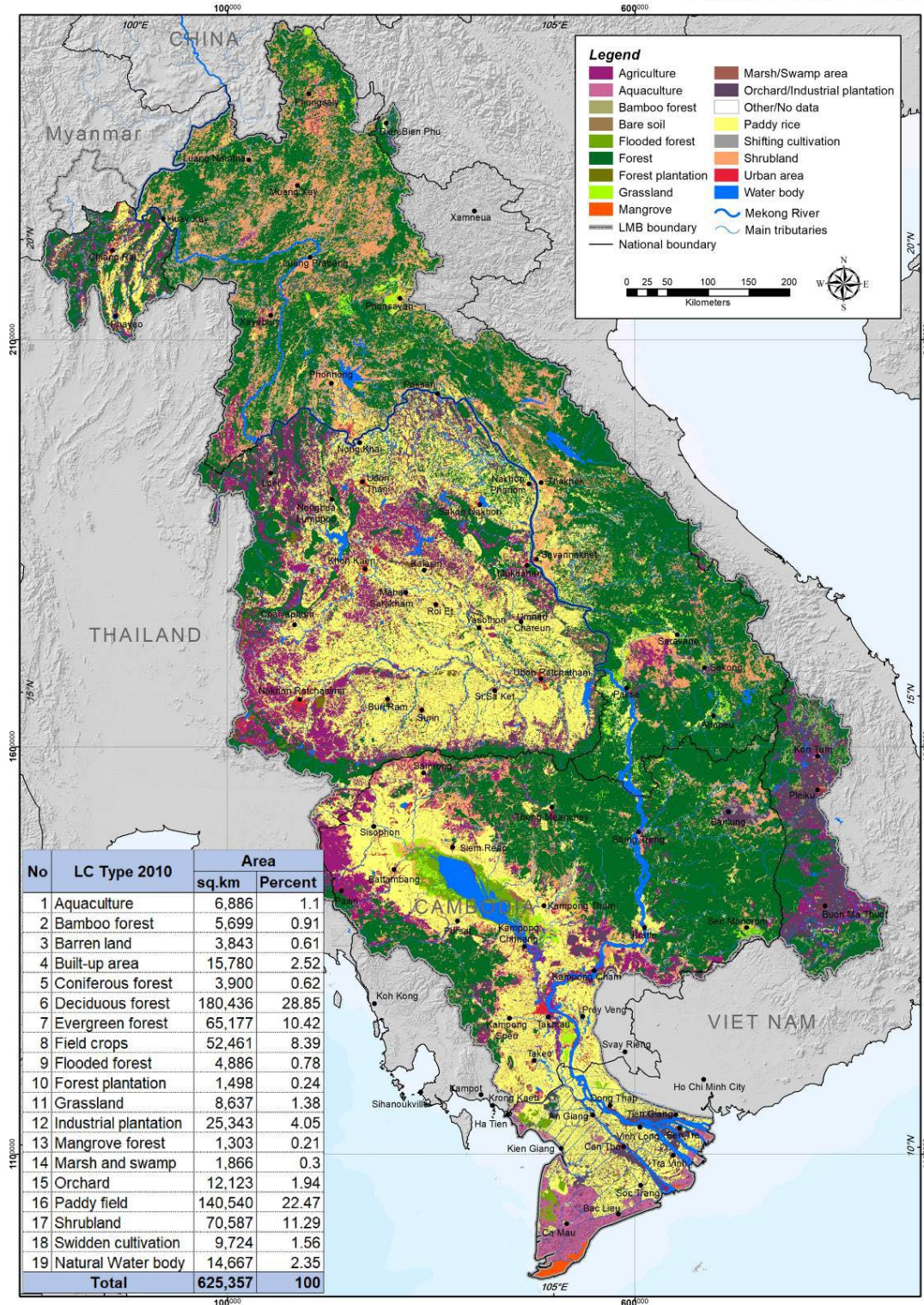


Figure 5. Homogeneous types of land cover, 2010

2.2.2 Hotspot and change detection analysis

For the hotspot analysis, an independent data set of change pixels was produced from 2000 to 2018. The change hotspot maps for 2018 were developed. As a result of the hotspot analysis, the maps show the probability that a pixel will change into any category. The probability that the following categories will change into another 'generic' category was calculated: aquaculture, barren, cropland, flooded forest, forest, mangroves, orchards and plantations (Figure 6). Additionally, we calculated the probability that other 'generic' categories will change into aquaculture, barren, cropland, flooded forest, forest, orchard and plantations, wetlands and urban and built-up (Figure 7).

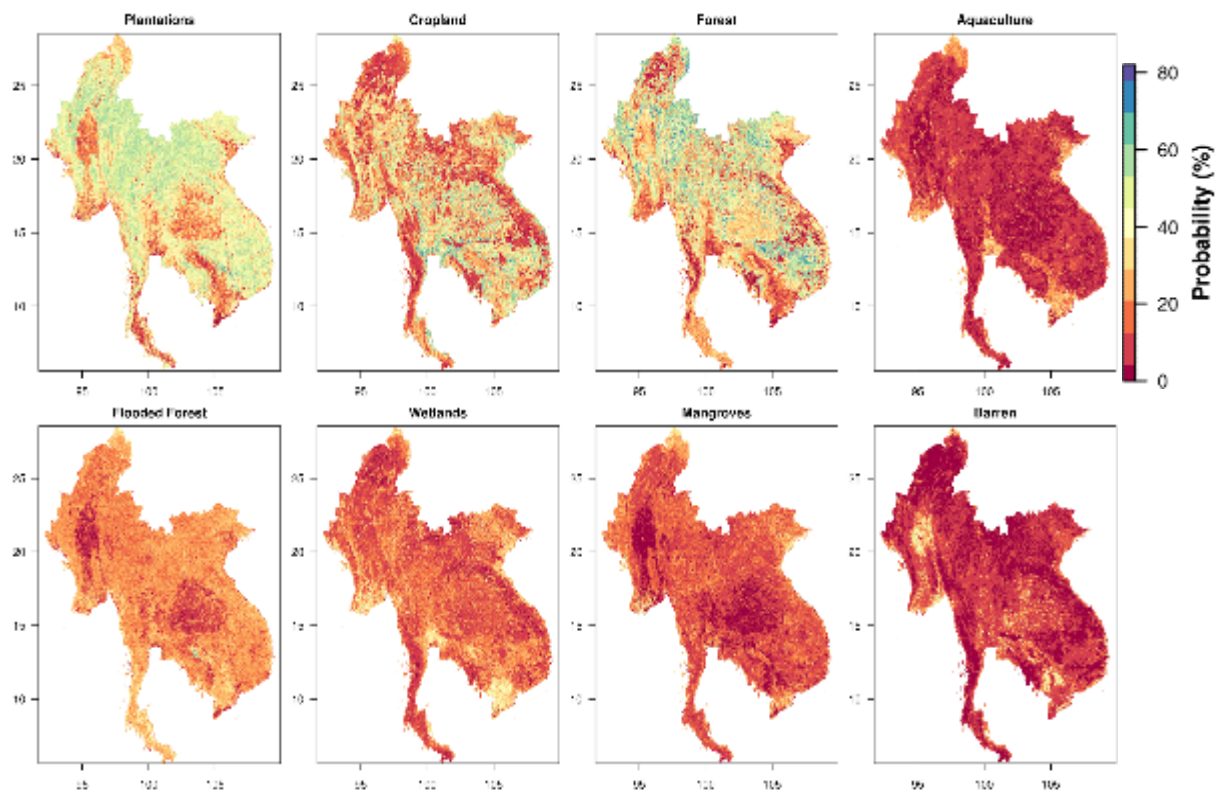


Figure 6. Probability change maps for the ‘from’ category for the different land cover types

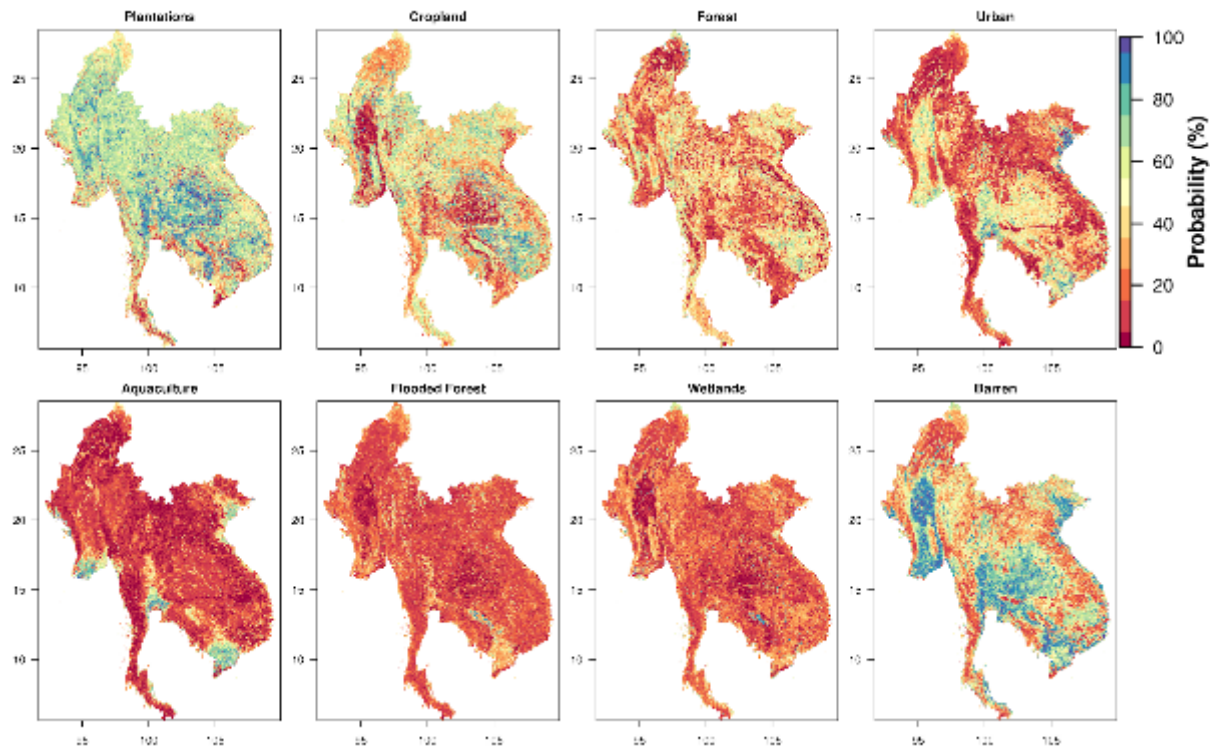


Figure 7. Probability change maps for the 'to' category for the different land cover types

The 2018 change probability maps are shown in Figure 6 and 7. The change probabilities in anthropogenic land cover types such as plantations, cropland, urban, aquaculture and barren were high, while other types of land cover were low. In the plantation land cover category, high change probabilities are shown around agricultural area, as well as in aquaculture. Higher probabilities in cropland distributed near the coast. In urban, high probabilities were found near population centres. Forest has also shown relatively high change probabilities. Higher change values are specifically notable in the Red River Delta in northern Viet Nam. For flooded forest and wetlands, there were high change probabilities around Tonle Sap, which is likely caused by the natural dynamics between them.

By observing each map, it is possible to see the land cover dynamics. For instance, large areas in Cambodia have a high probability of change in the 'from forest' category and also a high probability of change in the 'to cropland' category, while the croplands have a high probability to change into plantations.

Although throughout the region, the map shows that there is a high probability of change only from cropland to barren, overall there is a low probability of change for the other land cover categories overall. This is most likely because many croplands have a spectral signature of bare land during part of the year. Changes from cropland into any other category including bare land occurs infrequently in the training data. However, changes from any other category to bare land has a clear spectral signal of bare land.

Figure 8 and 9 show the probability distribution of pixels that have changed in the past. In the 'Specific to Generic' category, aquaculture had the largest coverage, and in the 'Generic to Specific' category, barren had the largest coverage. The median values are all greater than 80%, and all lower quantiles greater than 70%; wetlands and flooded forests have the highest median and smallest range in both categories respectively.

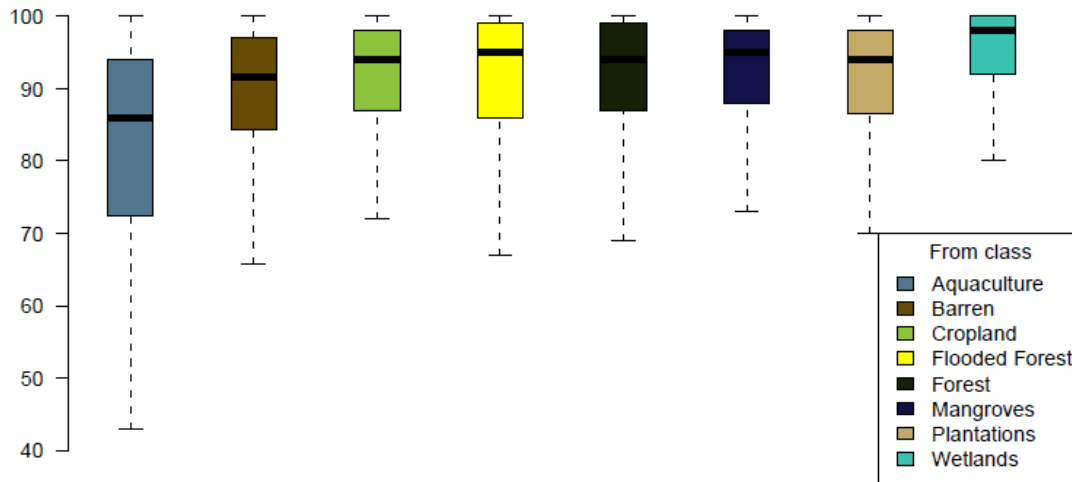


Figure 8. Distribution of validation change pixels in the 'Specific to Generic' category

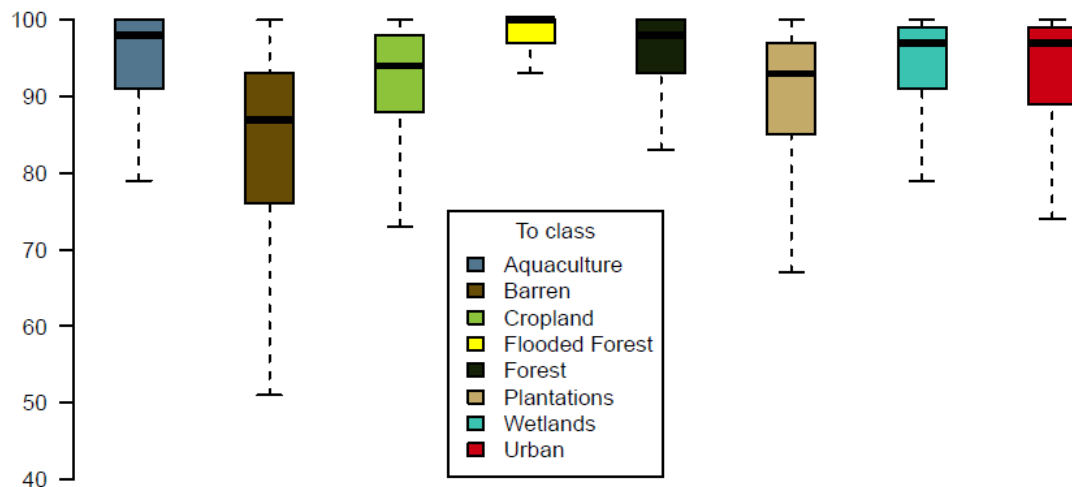


Figure 9. Distribution of validation change pixels in the 'Generic to Specific' category

Land cover change

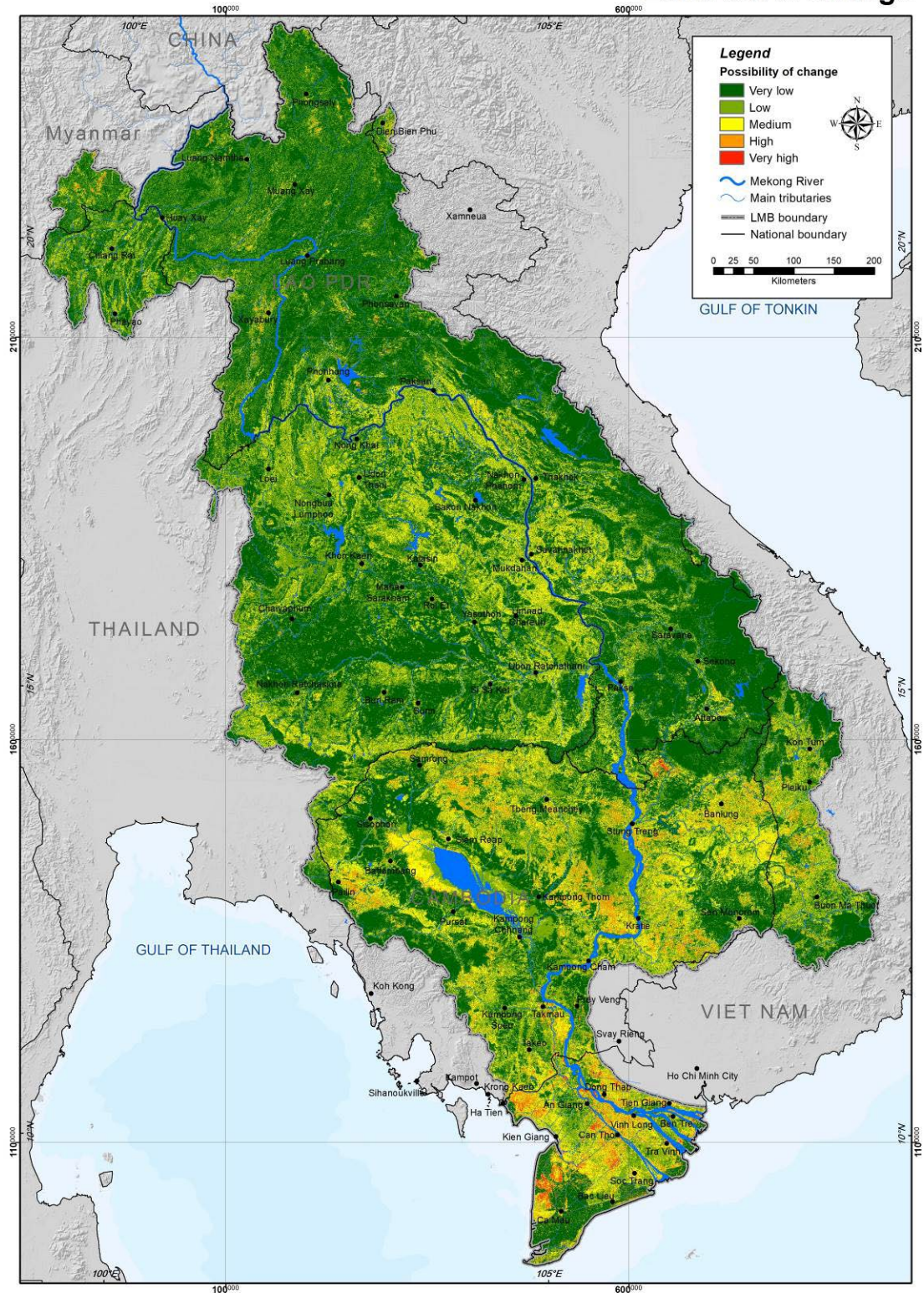


Figure 10. Map of land cover change

2.3 Data used for LULC 2020 mapping

2.3.1 Satellite image

Satellite data used for the land cover mapping was a combination of optical and Synthetic Aperture Radar (SAR) image data, including Landsat-8, Sentinel-2, Planet, and Sentinel-1. The data set was compiled with a pre-processing process, which is described in the Satellite image composites (3.1.)

The available information of the composite satellite imagery used for the development of LULC 2020 is shown in Table 3. To download the composite image, the GEE code and the instructions are provided in Annex 2.

Table 3. Composite satellite imagery uses for the development of LULC 2020

Satellite Imagery	Type of Sensor	Products	Resolution	Band available
Landsat 8	Passive Sensor	Annual	30 m	Red, Green, Blue, NIR, SWIR1, SWIR2
Sentinel 1 (SAR)	Active Sensor	Annual, monthly	10 m	VH, VV
Planet data	Passive Sensor	Jun, Sep, Oct, Nov, Dec 2020	4.7 m	Red, Green, Blue, NIR
Sentinel 2-A	Passive Sensor	Annual	60 m	B1 Aerosols
			10 m	B2 Blue
			10 m	B3 Green
			10 m	B4 Red
			20 m	B5 Red Edge 1
			20 m	B6 Red Edge 2
			20 m	B7 Red Edge 3
			10 m	B8 NIR
			20 m	B8A Red Edge 4
			60 m	B9 Water vapor
			60 m	B10 Cirrus
			20 m	B11 SWIR 1
20 m	B12 SWIR 2			

2.3.2 Field data collection

Referring to the land change analysis in chapter 2.2, the hotspot of land changing was identified. The field data collection was established using this information as a basis to identify the location for collecting the data. The objective was to collect the different classification and related LULC data in detail. The questionnaire for field data collection (Annex 1) was developed and shared to the Member Countries since then. According to the memorandum of understanding (MoU) and Terms of Reference (ToR) for the Member Countries, the field data collection activity was carried out during April to November 2020 for all Member Countries. The output of this activity consists of a full set of the data and information collected from the field related to LULC types.

The required number of data collected from the Change Detection Analysis for assigning the targeted area is shown in Table 4. The field data collection teams must collect the ground-truth points from the hotspots level: ‘very high’ level is the mandatory assignment, ‘high’ is the second priority, and ‘medium’ is the optional task to be determined by the expert team whether to collect their ground truth.

Table 4. Detailed number of hotspots disaggregated by probability of change for targeting field data collection

No.	Country	Probability of change (2000–2019)			Total
		Medium	High	Very high	
1	Cambodia	301	348	240	889
2	Lao PDR	131	112	49	292
4	Thailand	244	97	31	372
5	Viet Nam	124	238	476	838

Note: Drawn the results of the Change Detection Analysis on satellite data 2010 and 2018/2019.

Field data collection mainly focused on hotspot areas, i.e. the areas with changes in land cover of high significance. Data were collected in each of the Member Countries by national field data collection and land cover experts.

The planned location for the field data collection was shared with the Member Countries for their reference, as illustrated in Figure 11.

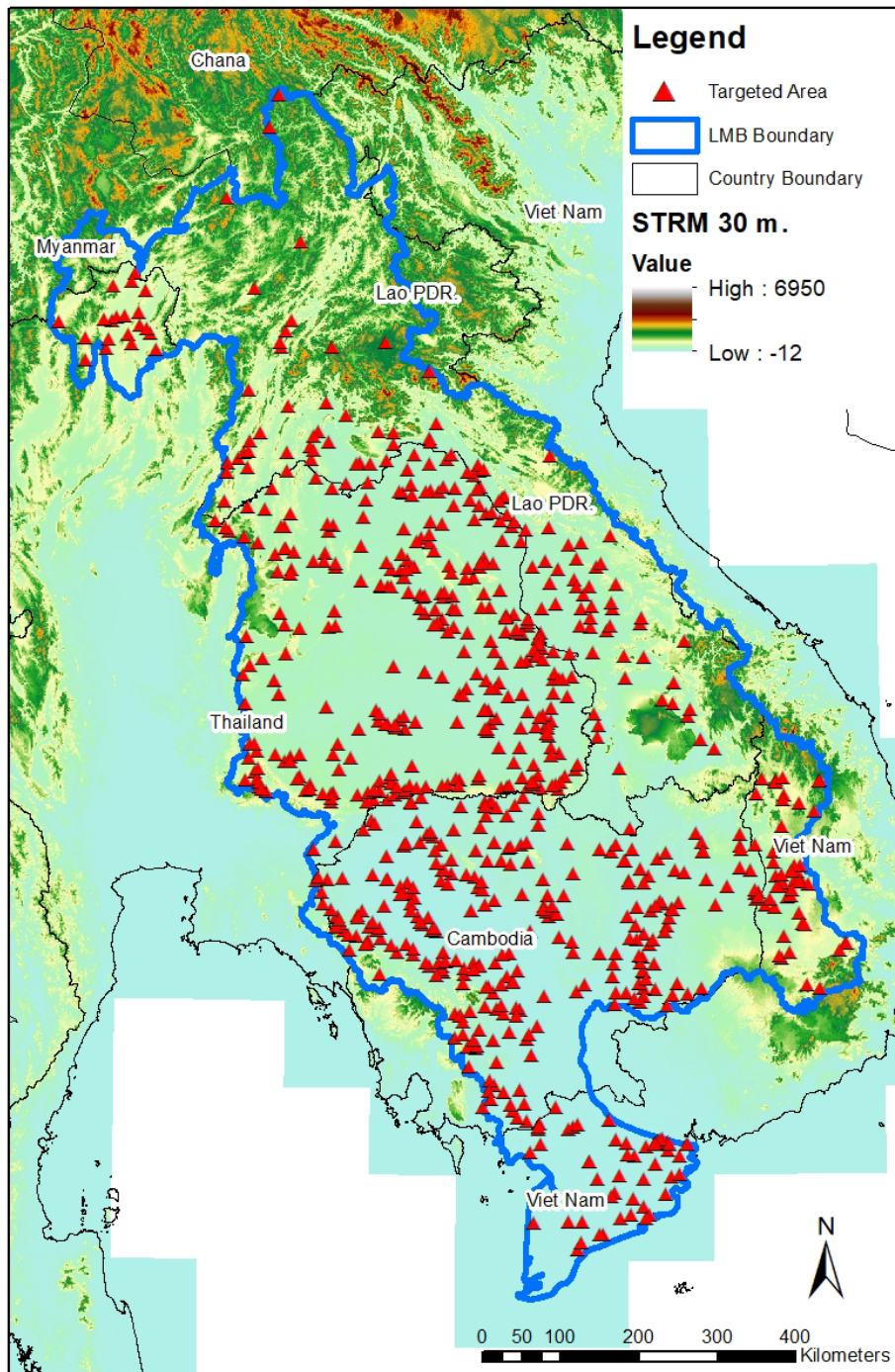


Figure 11. Planned locations for field data collection

After the field data collection was completed, field data collection for each of the Member Countries was summarized, as shown below. The forms, photos and detailed activity of each countries' data collection is shown in Annex 3.

Cambodia

The Phnom Penh Geoinformatics Education Center (PGEC) was the agency who response for field data collection. There are 240 locations of target areas (TAs) marked as "very high". With this TA location, the PGEC team generated about 2,142 Pois for its work. A total of 8,000

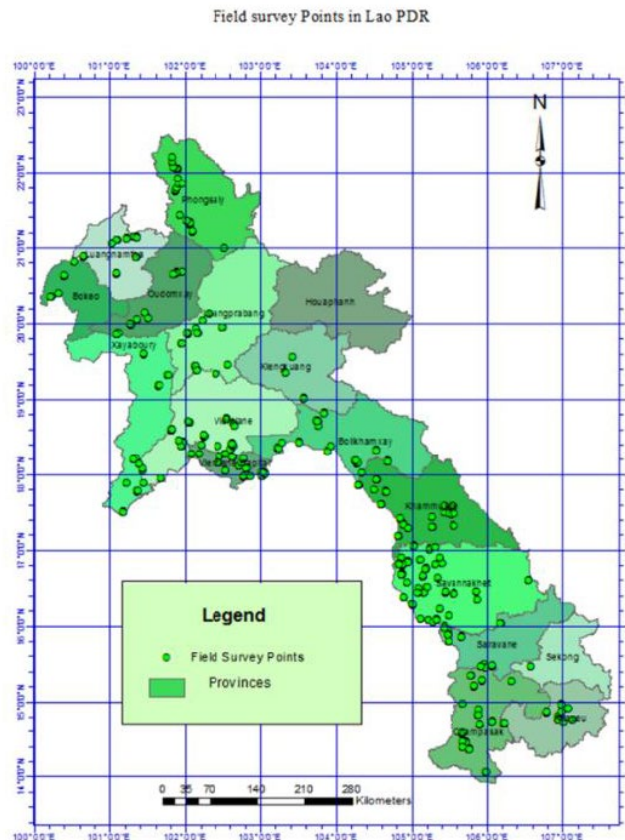


Figure 13. Field data collection by FIPD

Thailand

The Land Development Department was tasked with work covering field data collection, including carrying out a desk study to prepare for field trips and survey. There were 343 TAs out of 372 TAs, which had been accessed, accounting for 92.20%. There was a total 2,239 photos for 343 TAs in the field data collection and 1,661 LULC descriptions for targeting sites that were observed. In addition to the team’s experience in the region and in the TAs, information from first data, such as current land use maps, were collected. interviewing/discussing with local people/local government have helped the team find the best ways to obtain TAs and/or targeting sites and to choose appropriate LULC. Moreover, tools such as Google Earth, Google Maps, and apps for taking photos with necessary information helps significantly, especially to same time in the field.

Viet Nam

The Sub-National Institute of Agricultural Planning and Projection (Sub-NIAPP) was the agency who response for field data collection. Out of 476 TAs, 474 TAs (99.58%) were accessed and 2 TAs could not be accessed. According to the MoU, for each TA, there were eight targeting sites and for each targeting site, four photos were taken in four directions (N, E, S, W).

There was a total of 20,463 photos for 474 TAs in the field data collection, an average of 43 photos per TA. The reasons for this number of photos were: (i) In some TAs that are diverse

in LULC, more than eight targeting sites were chosen. (ii) The Sub-NIAPP used the application of *timestamp camera* to take photos and in some devices, our team had set up the mode “keep original photo”, which led to double photos in several targeting sites; and (iii) In some cases, when the team proceeded to the TAs, they also took photos for data reference. These three main reasons led to a higher number of photos in the field data collection than in the mentioned in the MoU. Figure 14 indicates the location of the data collection for all Member Countries.

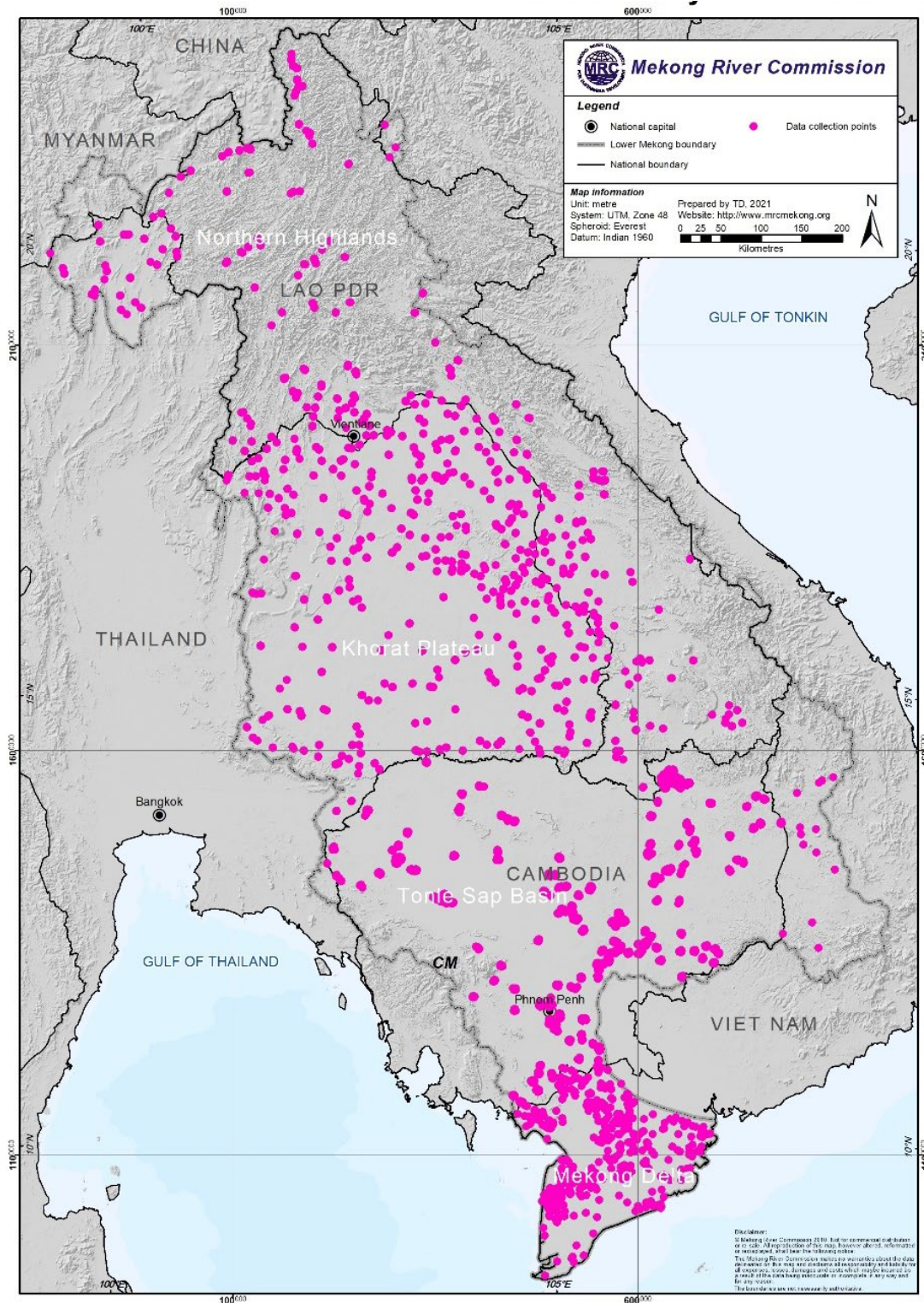


Figure 14. Field data collection map

2.3.3 Other reference and training data and resources

We used building blocks, also called primitives, to create land cover maps based on the MRC’s definitions and hierarchies. The primitive map layers are a suite of biophysical and end member maps, such as canopy height and percent canopy cover. Land cover primitives represent raw information needed to make decisions in a dichotomous key for land cover typing. These primitive layers are reassembled to create a final land cover map product according to a decision logic that results in land cover classes corresponding to the desired land cover typology (Saah, 2019).

The primitive maps were created by applying the random forest machine learning method. The authors trained with the primitive models using reference data (training data) that were collected from different resources including observations recorded in the field and high-resolution imagery provided by MRC Member Countries (Cambodia, Lao PDR, Thailand, Viet Nam) and existing time-series of training data from Regional Land Cover Monitoring System (RLCMS) of SERVIR-Mekong. The authors used a random selection of 161,000 points for each primitive to generate land cover maps of the LMB region. The number of reference points per land cover type is specified in Figure 15, and a detailed number of each type is provided in Annex 4. The spatial distribution of training data points is shown in Figure 16.

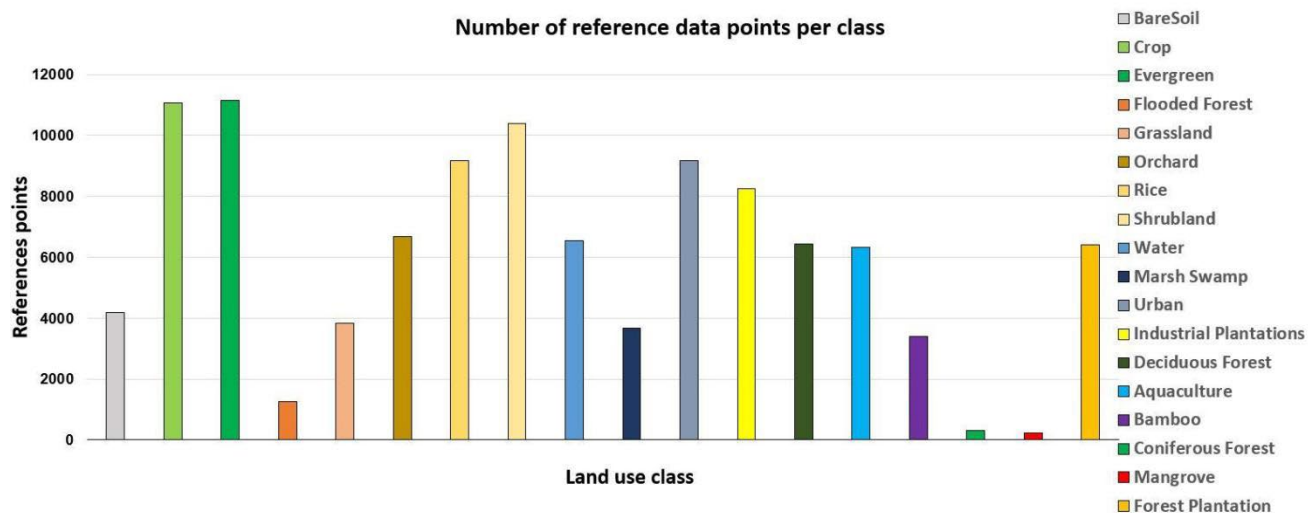


Figure 15. Number of reference data points per class

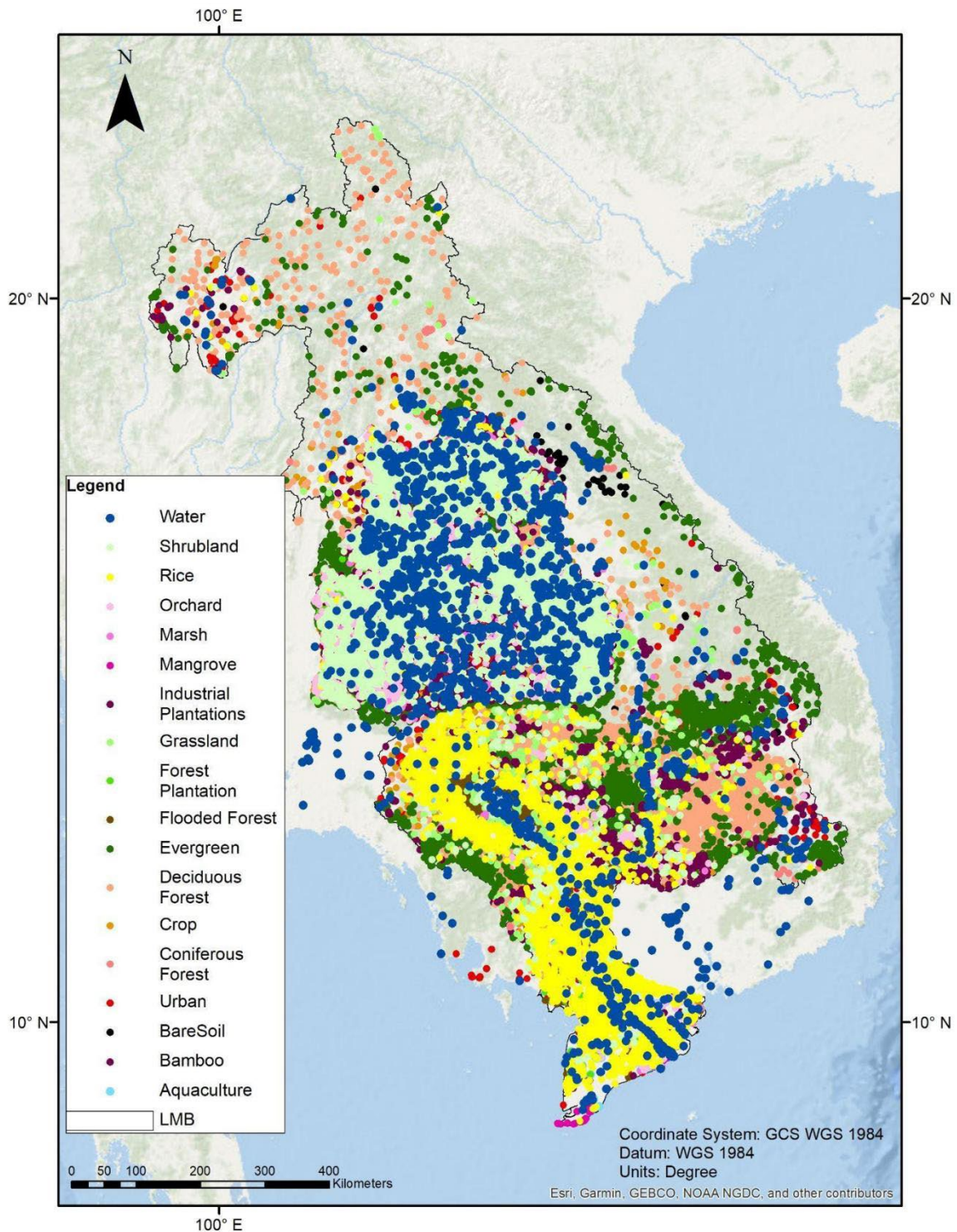


Figure 16. Spatial distribution of reference data points per land cover class

These collected training data were cleaned and regrouped into land use classes as per the MRC's land use classification. Finally, we merged data sets related to similar classes from different sources and generated training data for each class to train the model for the MRC land cover type.

2.4 Google Earth Engine

GEE is a cloud-based platform for planetary-scale geospatial analysis that brings Google's massive computational capabilities to bear on a variety of high-impact social issues, including deforestation, drought, disaster, disease, food security, water management, climate monitoring and environmental protection. It is unique in the field as an integrated platform designed to empower not only traditional remote sensing scientists, but also a much wider audience that lacks the technical capacity needed to utilize traditional supercomputers or large-scale commodity cloud computing resources.

Google Earth Engine consists of a multi-petabyte, analysis-ready data catalogue co-located with a high-performance, intrinsically parallel computation service. It is accessed and controlled through an Internet-accessible application programming interface (API) and an associated web-based interactive development environment that enables rapid prototyping and visualization of results. The data catalogue houses a large repository of publicly available geospatial data sets, including observations from a variety of satellite and aerial imaging systems in both optical and non-optical wavelengths, environmental variables, weather and climate forecasts and hindcasts, land cover, and topographic and socio-economic data sets. All of these data are pre-processed to a ready-to-use but information-preserving form that allows efficient access and removes many barriers associated with data management. Users can access and analyse data from the public catalogue as well as their own private data using a library of operators provided by the Earth Engine API. These operators are implemented in a large parallel processing system that automatically subdivides and distributes computations, providing high-throughput analysis capabilities. Users access the API either through a thin client library or a web-based, interactive development environment built on top of this library.

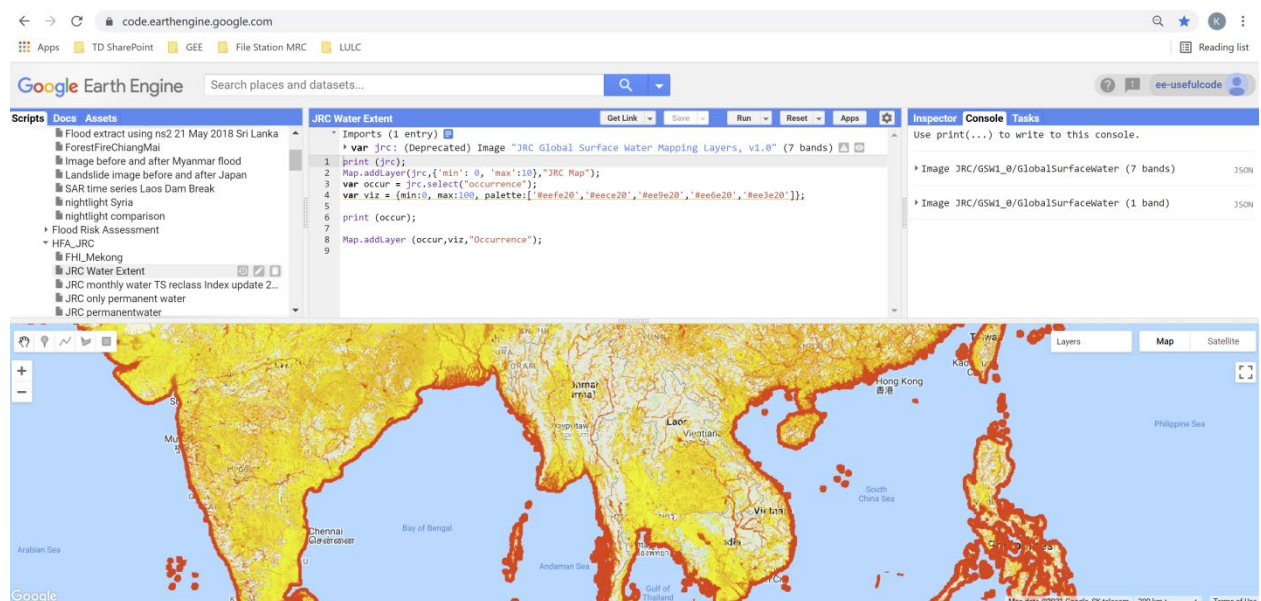


Figure 17. The Google Earth Engine user interface

3 Methodology on development LULC

There are three phases in the creation of the land cover map for the LMB region includes: (i) data input preparation and pre-processing after defining a land cover classification typology; (ii) supervised classification to create the primitive (or biophysical) layers; and (iii) the assemblage of biophysical layers into a customized land cover map, and an accuracy assessment. These steps are outlined in Figure 18.

Optical imagery from Landsat 8, Sentinel 2, SAR images from Sentinel-1, and Planet high-resolution images were processed and combined into a single stack of images for the LMB region. This image stack was used as a predictor in the classification process. The training sample was created by assigning the available reference data the coincident image values. A random forest algorithm was applied to the training sample and then used to calculate biophysical probability layers. Primitives were used in a decision tree to create the final assemblage with plantations. We discuss the specifics in more detail in the following sections.

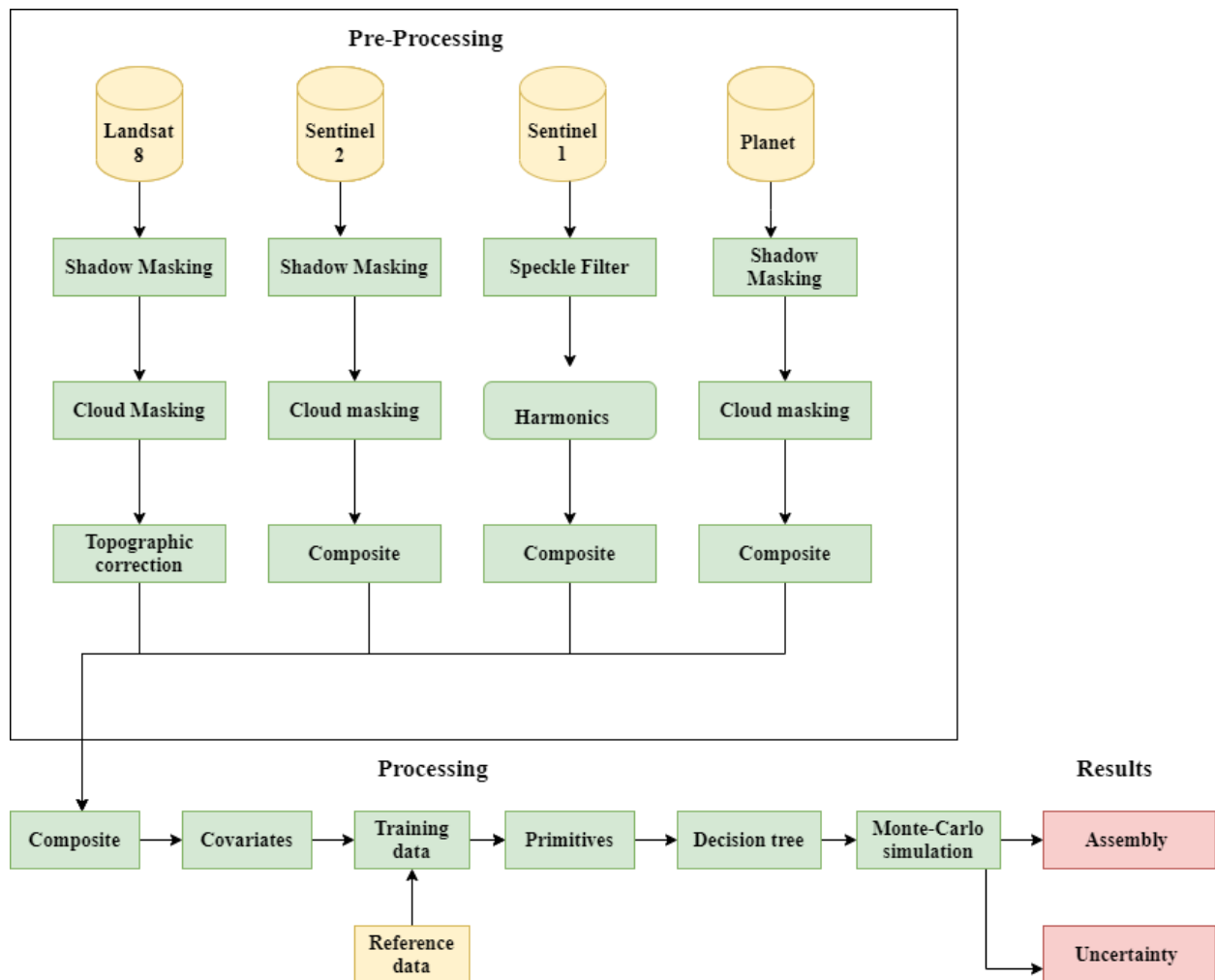


Figure 18. Satellite data processing

3.1 Satellite image composites

The data processing was conducted in GEE, a cloud-based computing environment that includes access to the full archive of Landsat and Sentinel imagery (. The GEE combines a large data archive of satellite imagery with a computational platform. The platform enables scientists to conduct research on environmental issues on a variety of spatial and temporal scales (Gorelick, 2017; Chen, 2017; Markert, 2018; Poortinga, 2019). Image processing for different satellite image data sets is described in detail below.

The first data set we used is the United State Geological Survey (USGS) Landsat 8 surface reflectance product. This product contains atmospherically corrected, orthorectified surface reflectance data. The images have been atmospherically corrected using the Landsat Surface Reflectance Code (LaSRC) (Roy, 2016; Vermote, 2016) and also contains the data produced by CFMASK (Zhu, 2012). The Landsat 8 has a spatial resolution of 30 m. The images with more than 40% cloud cover were excluded from the analysis due to issues with haze.

The second data set is the Copernicus Sentinel-2 mission, which provides high spatial resolution images. We used the Sentinel-2 image collection in GEE, which contains spectral bands representing Top-of-Atmosphere (TOA) reflectance from the Sentinel 2a and 2b satellites. The spatial resolution for Sentinel-2 varies for the different bands. The blue, green, red, and near-infrared bands have a resolution of 10 m, the red-edge and shortwave-infrared bands have a resolution of 20 m, and all of the others have a resolution of 60 m.

The third data set is the Copernicus Sentinel-1 mission, which provides Synthetic Aperture Radar (SAR) images with high temporal and spatial resolution. We also used the Sentinel-1 data from the GEE archive. GEE makes scenes available after the GEE team has applied thermal noise removal, radiometric calibration, and terrain correction using the Sentinel-1 Toolbox processing algorithms. Pre-processing includes thermal noise removal, radiometric calibration, and terrain correction. We used the available polarization bands, which were the VV and VH dual bands. A filter was applied to de-speckle the image (Lee, 2008). The Sentinel-1 has a spatial resolution of 10 m.

The fourth data set is Planet image, which provides multispectral high-resolution images. We used the Planet image was provided by Norway's International Climate and Forests Initiative (NICFI) data delivery portal, and the data was processed in GEE with cloud and shadows removal, and topographic corrections. The Planet image has spatial resolution of 5.7m. For the optical imagery, composites were created annually. A monthly composite was used for the SAR imagery. Composites included a medoid, 20th and 80th percentile from the annual image collection. Processing steps for Sentinel-2 included shadow and cloud removal.

Cloud shadow masking

Cloud shadow removal is an essential step because of the negative influence cloud shadow can have on data analysis (Zhai, 2018). We used the Temporal Dark Outlier Mask (TDOM) algorithm (Housman, 2018). This algorithm identifies pixels that are dark in the infrared bands but are found to not always be dark in past and/or future observations. It detects statistical

outliers with respect to the sum of the infrared bands. Next, dark pixels are identified by the sum of the infrared bands (NIR, SWIR1, and SWIR2). The pixel quality attributes generated from the CFMASK algorithm (pixel-qa band) was also used for Landsat-8 shadow masking (Zhu, 2012).

Cloud removal is another essential step in optical remote sensing. Clouds were removed using the quality bands for Landsat-8 and Sentinel-2. Furthermore, the GEE cloud score algorithm was used. The algorithm uses the spectral and thermal properties of clouds to identify and remove pixels with cloud cover from the imagery. The algorithm finds bright and cold pixels and compares them to the spectral properties of snow. The Normalized Difference Snow Index (NDSI) is calculated to prevent snow from being masked. The algorithm uses the visible, near-infrared, and shortwave infrared for a scaled cloud-score and then takes the minimum.

3.1.1 Topographic correction

The slope, aspect, and elevation can cause variations in reflectance for similar features with different terrain positions (Colby, 1991; Riaño, 2003; Shepherd, 2003); topographic correction is the process to reduce these effects. The Modified Sun-Canopy-Sensor Topographic Correction method as described by Soenen (2005) was applied to account for these effects. The method uses the sun-canopy-sensor (SCS) (Gu, 1998) with a semi-empirical moderator to account for diffuse radiation (Smith, 1980; Justice, 1981; Teillet, 1982). The ALOS global digital surface model was used (Tadono, 2014; Takaku, 2014).

The image repository on GEE available for users to download is provided below; data specific for each country can be found in Annex 2.

Landsat 8 – Image composite:

<https://code.earthengine.google.com/0d7c768838254bca899de0052c2452f8>

Sentinel – 2 Image composite:

<https://code.earthengine.google.com/402877f25da39dac19e3645451efc1b7>

Asset of all available images:

https://code.earthengine.google.com/6c4b45b2e3b39ed8a1210ac11866fffc?accept_repo=users%2FUsefulCode%2Fmrc

3.1.2 Covariates

The seasonal Landsat-8, Sentinel-1, and Sentinel-2 and Planet seasonal composites were combined into an image stack, where the different images are represented as bands. A series of indices were calculated from the optical satellites, we created multiple indices for each primitive model. We applied a random forest classifier to create probability maps for each land cover class using the probability model on these images.

3.1.2.1 Development of covariates

A variety of covariates was calculated from Landsat and Sentinel image composites, and contained all the bands and derived indices. Covariates are used as the composite band values and derived indices that are used as predictors in the random forest algorithm. They serve as additional information and important in developing the classification models. Table 3 provides an overview of the band combinations used in the normalized difference (Equation 1) calculations. For some combinations, there are more familiar names, such as the Normalized Difference Water Index (NDWI) (McFeeters, 1996), the Normalized Burn Ratio (Key, 1999), the Normalized Difference Snow Index (NDSI) (Salomonson, 2004), and the Normalized Difference Vegetation Index (NDVI) (Rouse, 1973) Normalized difference metrics (ND; Equation 1) were calculated for every Landsat-8 and Sentinel-2 image. Table 5 provides an overview of the band combinations.

Equation 1

$$ND(band1,band2) = \frac{band1 - band2}{band1 + band2}$$

Table 5. Overview of the band combinations

Blue	Green	Red	NIR	SWIR1
Green	Red	SWIR1	Red	SWIR2
Red	NIR	SWIR2	SWIR1	
NIR	SWIR1		SWIR2	
SWIR1	SWIR2			
SWIR2				

Two ratio (R) bands were included, calculated by dividing one band by another. Similarly, SWIR1 and NIR bands as well as the red and SWIR1 were also calculated. The Enhanced Vegetation Index (EVI, Equation 2) (Jiang, 2008) was also included, as well as the Soil-adjusted Vegetation Index (SAVI), Equation 3 using L = 0.5) (Huete, 1979). The Index-based Built-Up Index (IBI) (Xu, 2008) was calculated from Equation 4.

Equation 2

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

Equation 3

$$SAVI = \frac{(1 + L)(NIR - red)}{NIR + red + L}$$

Equation 4

$$IBI = \frac{NIR}{(NIR + red)} + \frac{green}{(green + swir1)}$$

The total number of bands in the final image stack was 142. We sampled the complete stack of covariates for each land cover class and evaluated the importance of the covariates in R (Breiman, 2001; Liaw, 2002; Team, 2018). A smaller number of covariates reduces the computational expense and eliminates noise. Whereas the bands have different spatial resolutions, sampling was carried out on a 10 m spatial resolution, the smallest spatial resolution of all bands.

*3.1.2.2 Additional data layers as covariates***Tree canopy cover and height**

Yearly tailor-made products mapping fractional tree canopy cover and tree canopy height derived from summary statistics of annual Landsat surface reflectance products and global sub-pixel training data (Hansen, 2011, 2016) were included as covariates to map the primitive layers. The process of creating these annual products includes temporal smoothing using linear regression and median filters for inter-annual variation. Links to the assets are shown below:

Tree canopy cover:

https://code.earthengine.google.com/?asset=projects/servir-mekong/yearly_primitives_smoothed/tree_canopy

Tree canopy height:

https://code.earthengine.google.com/?asset=projects/servir-mekong/yearly_primitives_smoothed/tree_height

Water persistence metrics

The JRC Global Surface Water Mapping data set contains maps of the location and temporal distribution of surface water from 1984 to 2015, and provides statistics on the extent and change of those water surfaces (Pekel, 2016). The mapping layers product consists of one image containing six bands. It maps different facets of the spatial and temporal distribution of surface water over the last 32 years.

Terrain indices

We used terrain properties as inputs in creating primitives. The five terrain properties include elevation, slope, aspect, and two aspect derivatives. These include a measure of the deviation from east (the sine of aspect) and deviation from north (the cosine of the aspect). All terrain indices were derived from the digital elevation data from the 1-arc-second (approximately 30 m ground resolution) Shuttle Radar Topography Mission data set.

3.2 Random forest classification and create primitives

Each supervised classification was set up to predict the primitive class of interest. The other land cover classes were aggregated into an absent class. For example, for the settlement primitive we used reference data that were assigned a label as either settlement or not-settlement. Another random training sample of about equal size was created from the other reference data. In order to create good primitive models, the reference data selected for each primitive needed to meet requirements such as having a good spatial distribution over the whole LMB region; an equal number of points for training the target class and others types; and a high proportion of data points in the other class for classes that look similar. See Annex 4 for a detailed number of training data to build a model of each primitive.

The table with training data was used in a random forest model in GEE to select the most important covariates (Breiman, 2001; Liaw, 2002; Team, 2018). Random forest is a widely used method, which can handle large data sets and categorical as well as continuous data. The random forest method is relatively easy to interpret and understand by observing the trees, and is not prone to overfitting.

Various methods are available to select a parsimonious set of metrics to use as predictors in a model, including dimensional reduction of the data, such as a principal component analysis or canonical correlation analysis. We used the information on the variable importance measures to select the covariate list. The random forest classifier was then applied in GEE using the selected most important covariates and the training data. The classifier was trained with 10 trees in the probability mode. The classifier was then applied to each class, applying variable selection for 100 trees, resulting in the probability map for the class. In this LULC map development exercise, we built 15 primitive models for 14 land cover categories. For three primitives, including mangrove, industrial plantation, and plantation forest, we used binary layers from 2010 in the assemblage.

Primitives assessment/accuracy assessment of primitive models using a 50% threshold

We evaluated the accuracy and Kappa using a 50% threshold. We then applied a confusion error matrix to the binary classification of the primitive model, as class 0 (other classes) and class 1 (primitive class). An example of a confusion matrix for binary classification of the 'cropland' primitive model is shown in Table 6 (TN = True Negative; FP = False Positive; FN = False Negative, TP = True Positive).

Table 6. Confusion matrix for binary classification to predict croplands

		Predicted	
		Negative	Positive
Negative	Other class (0)	TN	FP
Positive	Cropland (1)	FN	TP

The accuracy of the model is calculated using the given **Equation 5** and **Equation 6**, as shown below.

Equation 5

$$\text{Overall accuracy} = (TN+TP) / (TN+FP+FN+TP)$$

Equation 6

$$\frac{\text{Total number of correctly classified samples}}{\text{Total number of reference samples chosen}}$$

Overall accuracy =

The user’s accuracy and the producer’s accuracy of the classified map were calculated to find the accuracies of the primitive model (Equations 7 and 8).

Equation 7

$$\text{Producer's accuracy} = \frac{\text{Number of samples correctly classified in a given class}}{\text{Total number of samples chosen for that class}} \times 100$$

Equation 8

$$\text{User's accuracy} = \frac{\text{Number of samples correctly classified in a given class from the selected samples in that group}}{\text{Total number of samples classified in that group out of entire samples selected}} \times 100$$

Kappa statistics reflect the difference between the actual agreement and the agreement expected by chance. For example, Kappa of 0.58 of the crop primitive model means a 58% better agreement than by chance alone. The formula to calculate kappa is shown below (Equation 9). Observed accuracy is determined by diagonal in error matrix. Chance agreement incorporates off-diagonal, sum of [Product of row and column totals for each class].

Equation 9

$$\hat{K} = \frac{\text{observed accuracy} - \text{chance agreement}}{1 - \text{chance agreement}}$$

The confusion matrix, accuracies, and kappa value of each primitive model are shown in Table 7 of chapter 4.2, *Probability classified primitive maps*.

3.3 Assemblage and land cover mapping

In the final phase to generate the land cover map, we combined all the primitive data layers into a land cover map using an assemblage logic. The overall process is described in Figure 19.

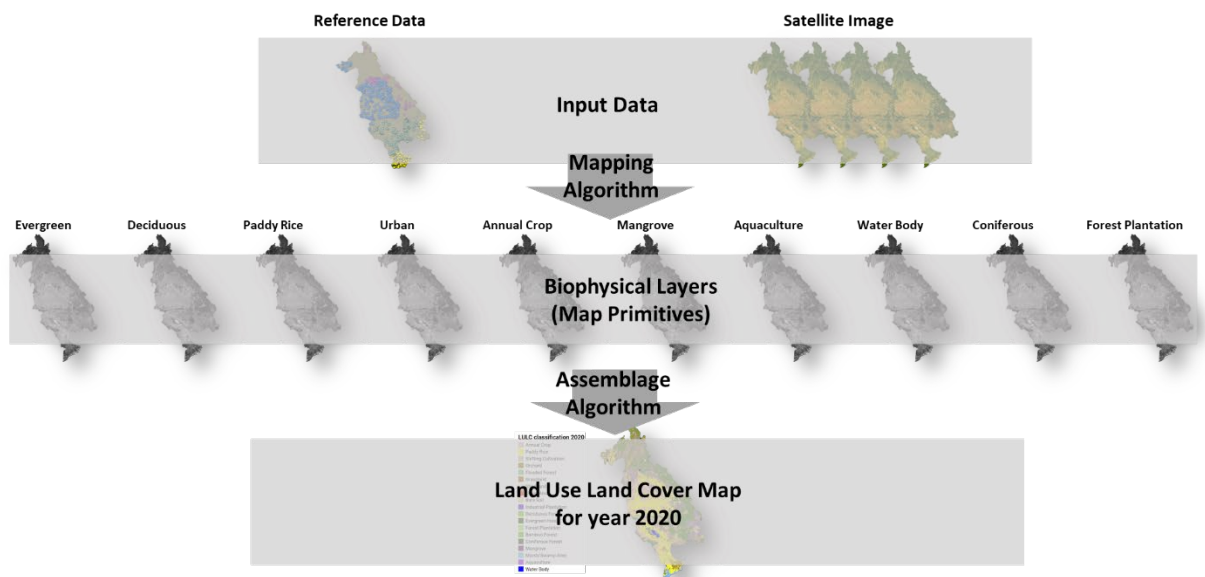


Figure 19. Development of primitives and assemblage

Assemblage is a method for combining different primitive layers into a land cover map while incorporating users’ land cover definitions and priorities and preserving uncertainty information. We used a hierarchical, decision tree structure (dichotomous key) to prioritize the integration from the primitive layers. We propagated uncertainty along the decision tree by running a series of Monte Carlo simulations. During each iteration of the simulation, primitive layers are randomly generated according to the accuracy of the primitive layer. The simulated values are then passed through the user-defined decision tree to generate a series of land cover predictions. These predictions are aggregated to produce a final land cover map, as detailed below.

In comparison with alternative probabilistic classification methodologies, such as Bayesian networks or fuzzy logic, Monte Carlo sampling over deterministic decision trees enables end-users to construct bifurcating decision trees by posing yes/no land cover-related questions. Because the primitives are probabilistic, sampling from them independently within the group-designed decision trees and collecting aggregate statistics allow to retain as much uncertainty as possible in the input data in the final products. Although a Bayesian network approach may provide a clearer uncertainty propagation scheme to those familiar with probabilistic mathematics, the Monte Carlo methodology is considered more easily and widely understood, and therefore more appropriate.

The assemblage is constructed from the user defined rule set. We constructed a decision tree to generate the final 18 class typologies. Placements near the top of the tree are more likely to be classified, while primitives lower in the decision tree are more likely to be masked. The

tree was designed based on the priority of the land cover types, but the accuracy was taken into consideration.

A decision tree was applied for all primitives to create the final land cover assemblage of the LULC 2020 of the LMB region. It was used to set the order and thresholds used to combine each of the primitives together into one final land cover map. According to the IPCC definition of forest, a forest includes land of more than 0.5 ha with trees higher than 5 m and a canopy cover of more than 10 percent. Figure 20 shows that the decision tree that was applied for the LULC 2020 is divided into two main branches based on the information of the minimum mapping unit (MMU) with tree canopy cover and height characters. One branch for forest land and plantation land cover types with tree canopy cover less than 10% and the other branch for non-forest land cover types with tree canopy cover less than 10%.

Figure 20 describes in greater detail the tree and threshold of each primitive to generate the final land cover map.

This decision tree was run 100 times with a Monte-Carlo simulation process (Binder, 1993). The process entailed adding a grid with random numbers to each of the primitives. It produces a final land cover map, which is the mode of the 100 simulations, and a probability map, which is the count of the mode divided by the total number of model runs.

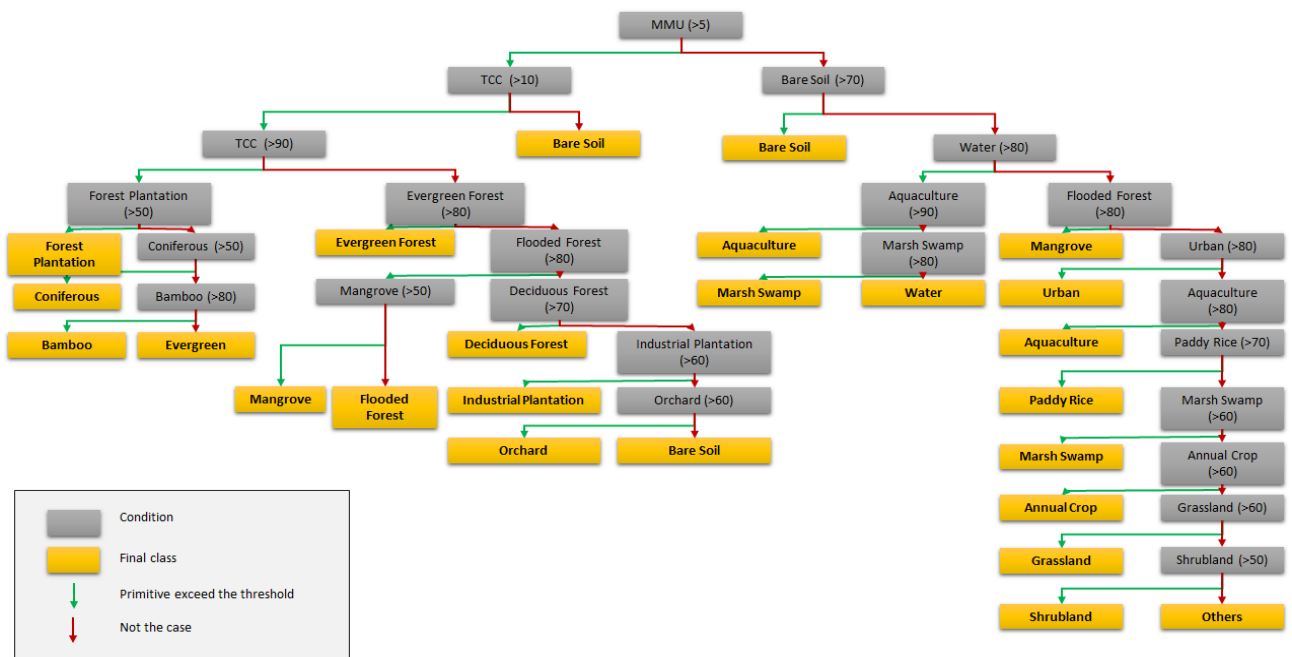


Figure 20. Decision tree used to generate the final land cover map

4 Results of the Land Use and Land Cover Map 2020

4.1 Satellite image catalogues and accessibilities

The satellite image composite of the LMB region generated seasonal composites for wet and dry season in 2020 for four satellite images including: Landsat, Sentinel-2, Sentinel-1, and Planet image. The MRC Member Countries can access and download these data sets using the GEE link, which is provided in Annex 5.

GEE codes to generate the satellite image composites of Landsat, Sentinel-1, and Sentinel-2 is also included in the GEE code repository for MRC.

https://code.earthengine.google.com/?accept_repo=users/UsefulCode/mrc

4.2 Probability classified primitive maps

The output of the random forest model (primitive) is the probability map for the class. Figure 21 shows classified maps using probability classification for 14 classes. Colour patterns indicate low probability to high probability of the class. These primitive classes will be used in the assemblage process for final LULC 2020 mapping.

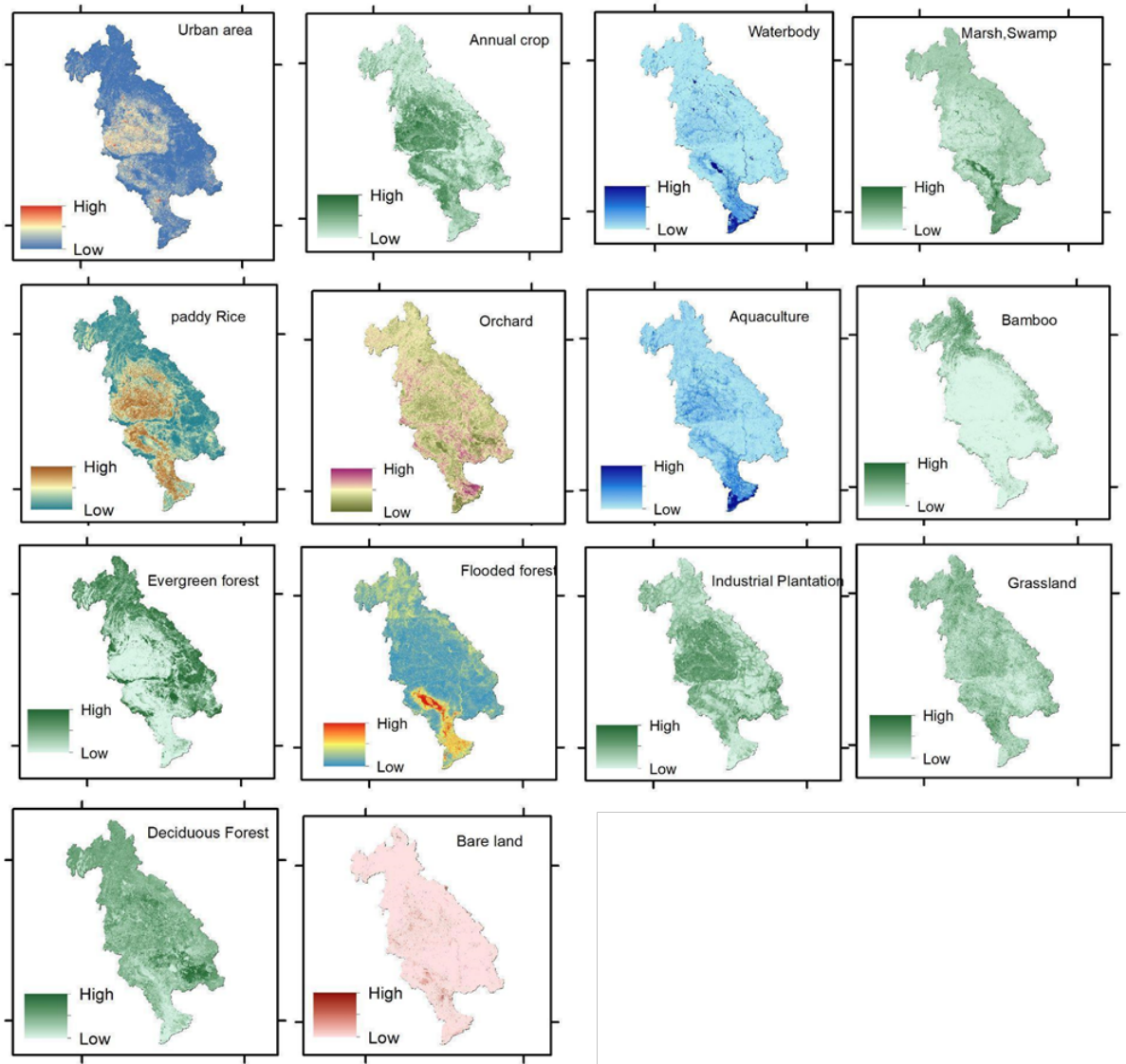


Figure 21. Probability distributions of 14 primitives from the Random Forest algorithm in the LMB

4.3 Accuracy assessment of primitive models

The study compared the reference data labels to the estimated primitive probabilities. Figure 28 shows the histograms of probability distribution from 14 land cover classes; there are two histograms in each class. From the first histogram, the distribution of the primitive probability is shown, whereas the probability distribution of the other classes is shown in the second histogram. The cumulative probability over the horizontal axis in these two graphs indicates the separation between and combination of the class and all other classes. It can be observed that there is a very sharp division between the land cover classes: grassland, industrial plantations, bamboo, aquaculture, waterbody, urban, evergreen, and flooded forest. Crop, deciduous forest, orchard, rice, and marsh and swamp show a greater mixing of the classes.

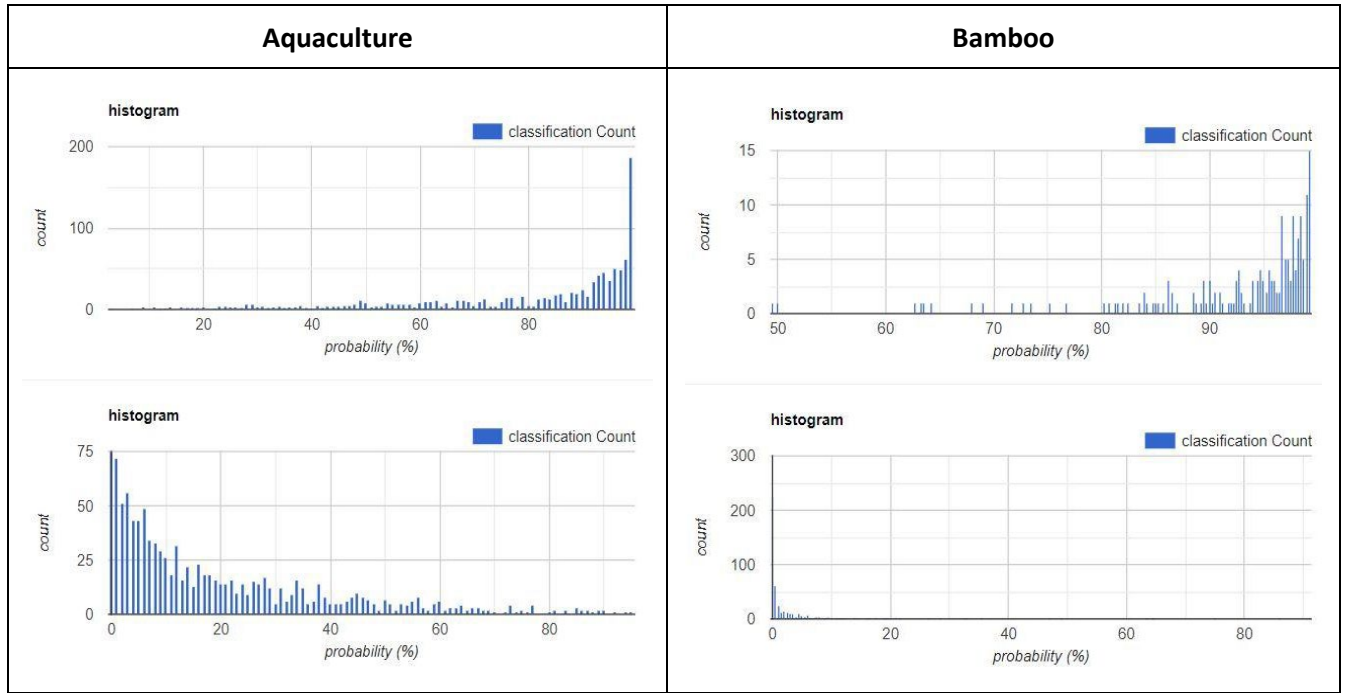


Figure 22. Probability distribution of primitive models (aquaculture and bamboo)

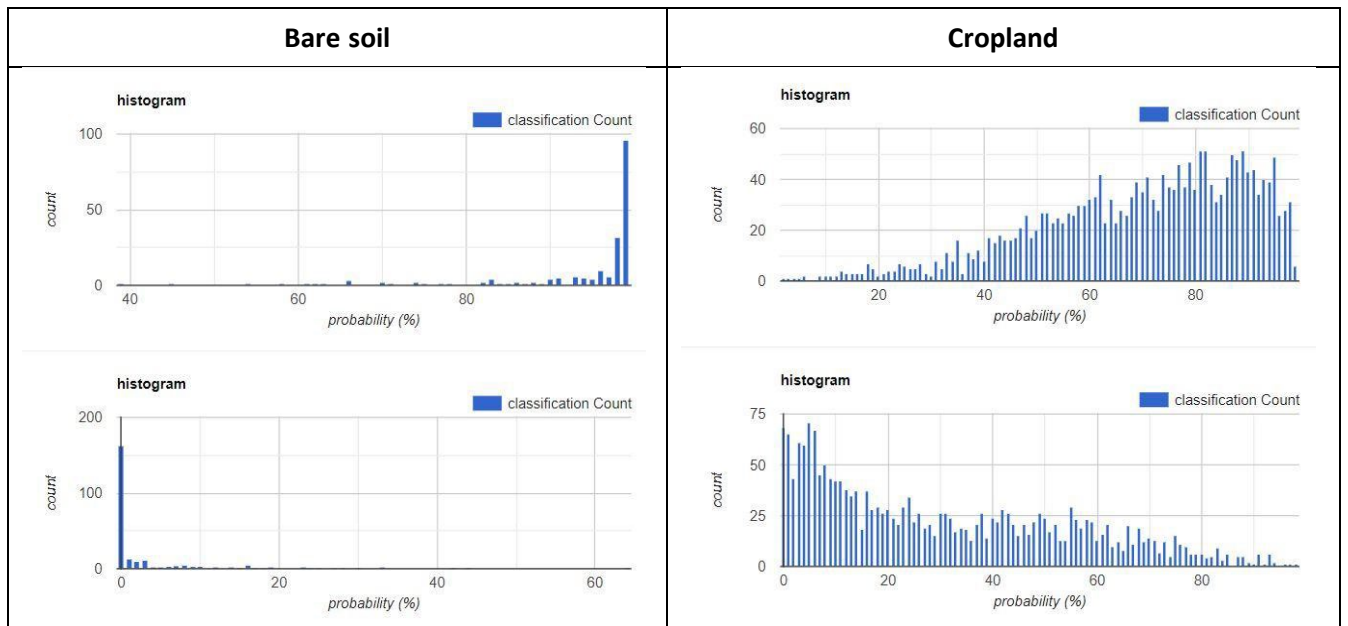


Figure 23. Probability distribution of primitive models (bare soil and cropland)

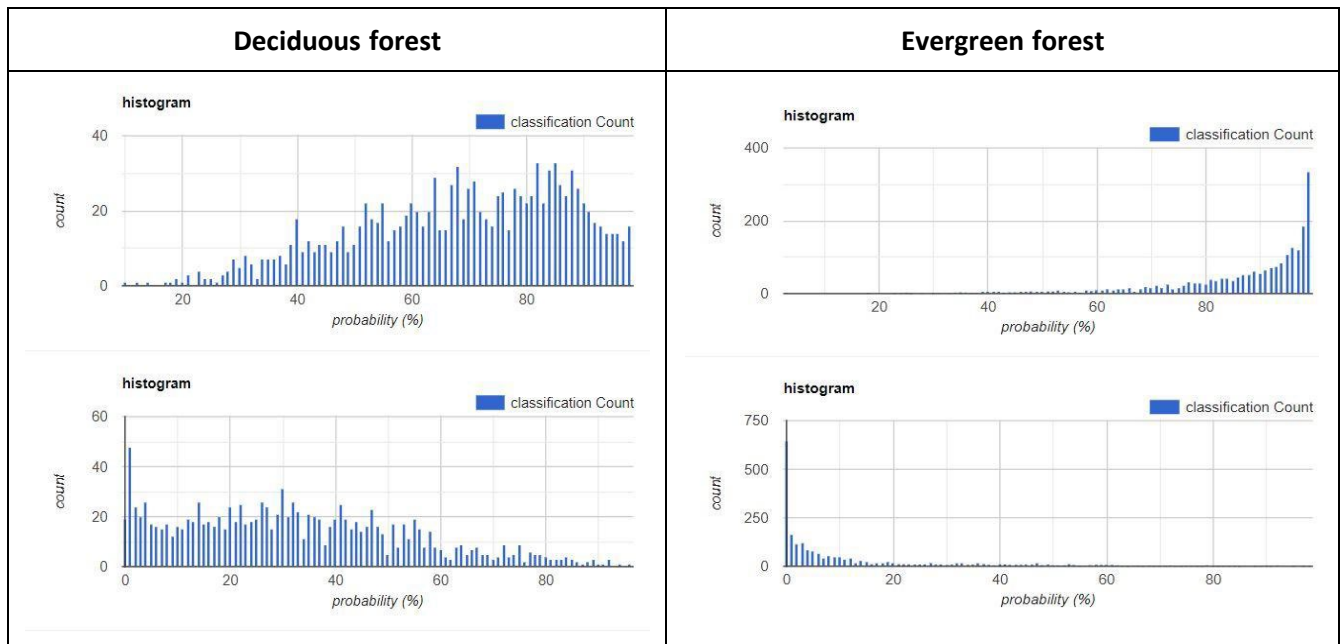


Figure 24. Probability distribution of primitive models (deciduous and evergreen forest)

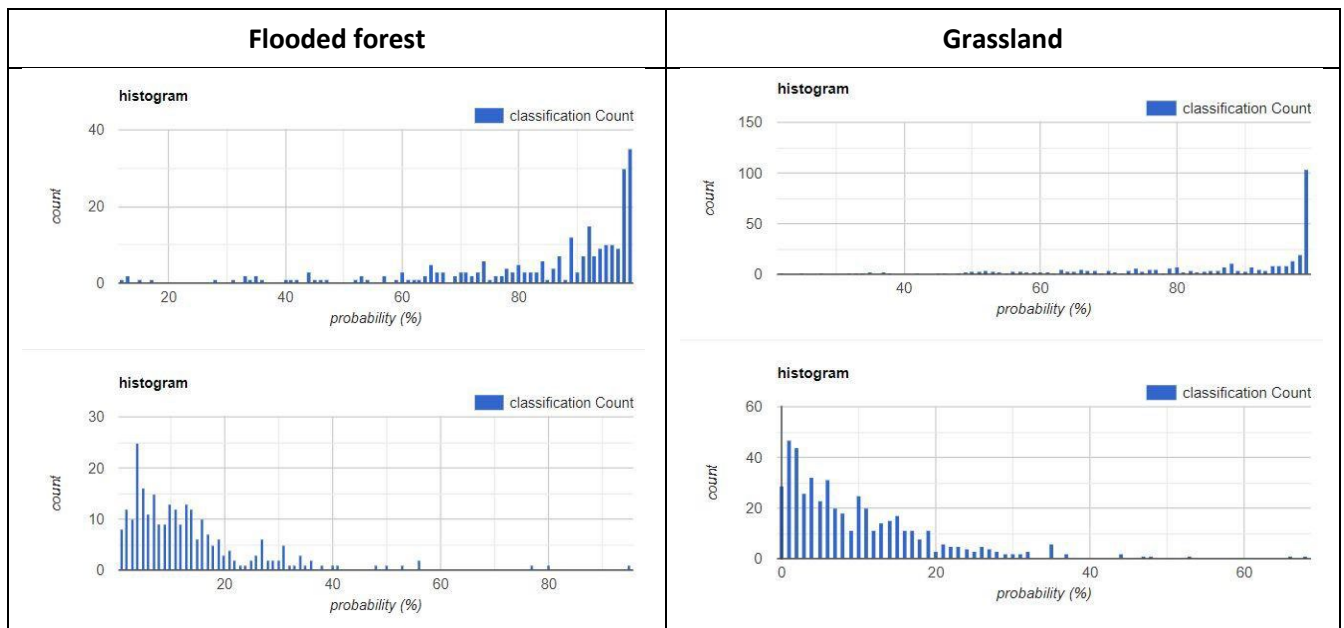


Figure 25. Probability distribution of primitive models (flooded forest and grassland)

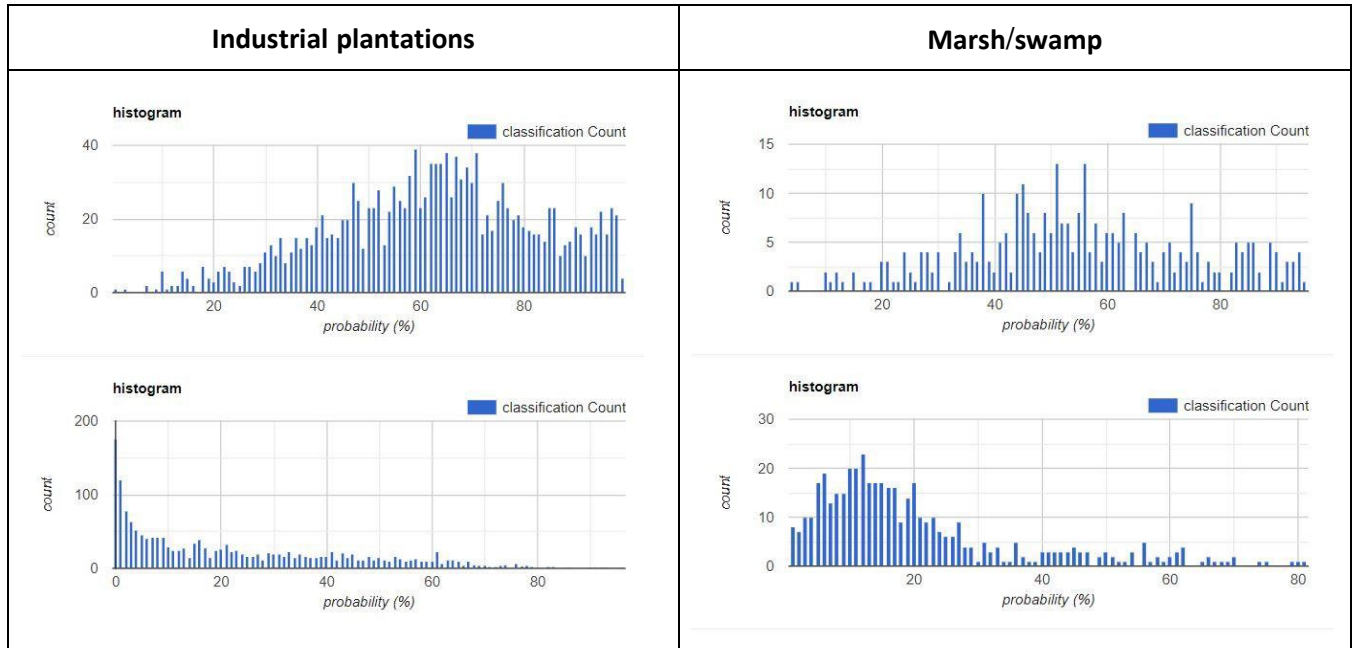


Figure 26. Probability distribution of primitive models (industrial plantation and marsh/swamp)

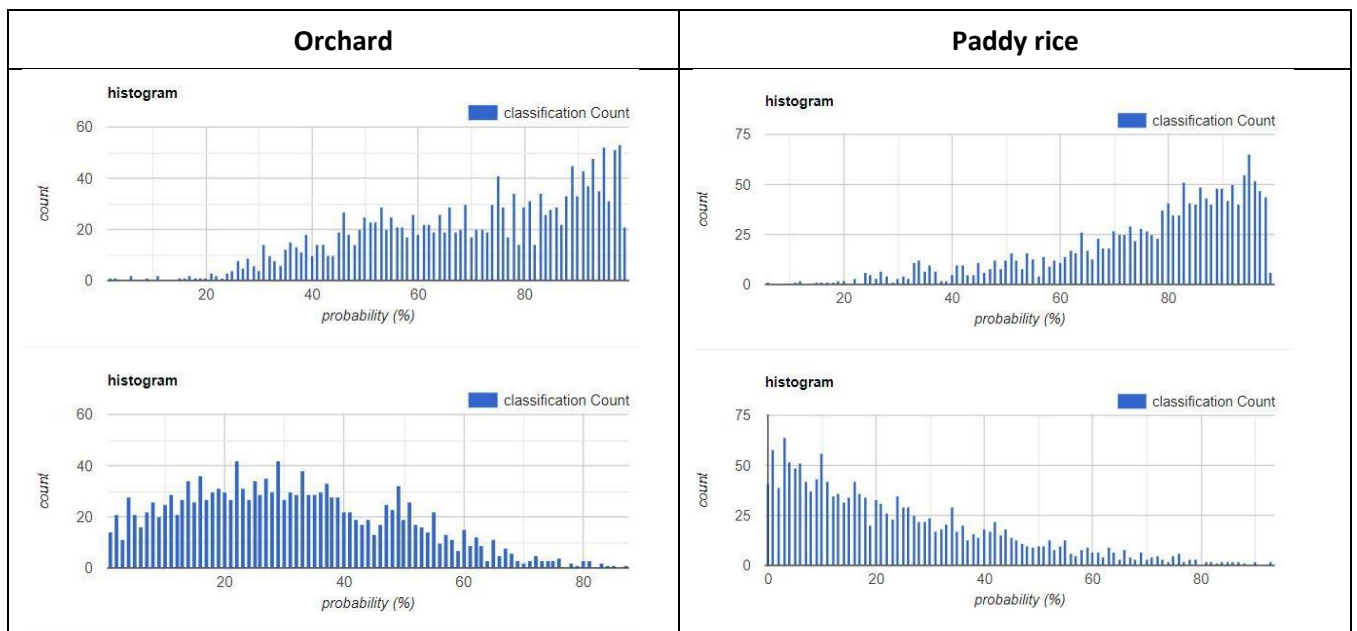


Figure 27. Probability distribution of primitive models (orchard and paddy rice)

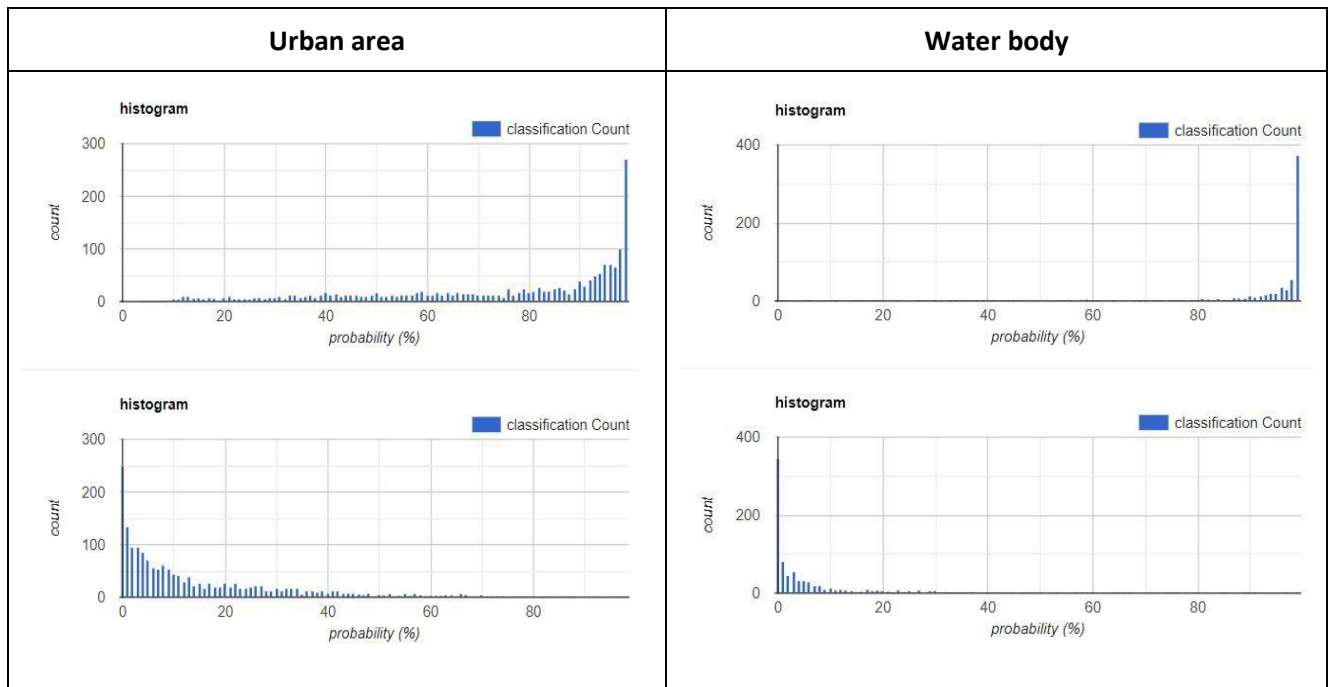


Figure 28. Probability distribution of primitive models (urban area and water body)

We evaluated all primitive images using the accuracy and Kappa at 50% validation thresholds. Table 7 shows the accuracies. There was an overall accuracy of 0.88, with the highest accuracy for bamboo and the lowest for marsh/swamp, , which was often confused with orchards.

Table 7. Confusion matrix of primitive models

Class name	Producer's accuracy (%)	User's accuracy (%)	Kappa	Overall accuracy (%)
Annual crop	0.84	0.80	0.63	0.81
Urban	0.83	0.94	0.77	0.88
Paddy rice	0.87	0.88	0.76	0.88
Deciduous	0.79	0.79	0.60	0.80
Evergreen	0.94	0.93	0.94	0.87
Marsh/swamp	0.56	0.85	0.52	0.77
Aquaculture	0.87	0.90	0.77	0.88
Grassland	0.89	0.99	0.90	0.95
Orchard	0.84	0.86	0.70	0.85
Bare soil	1.00	0.96	0.96	0.98
Flooded forest	0.90	0.98	0.95	0.89
Water	0.96	0.97	0.94	0.97
Industrial plantation	0.74	0.86	0.63	0.82
Bamboo	1.00	0.97	0.98	0.99

4.4 Accuracy assessment for the LULC 2020 product

The MRC conducted the accuracy assessment to ensure that the quality of the product was acceptable. Ten % of the field data was collected separately at the beginning of the work. The 2,198 points were randomly extracted from the field data collection points, which has been shared by the Member Countries since December 2020.

The resulting Confusion Matrix is depicted in Table 8. The overall accuracy was 87.1% with the Kappa Coefficient was 0.856, which indicates high agreement between the classified map and the ground-truth data.

Table 8. Confusion matrix for accuracy assessment

Confusion matrix		Classified Map																	Number of Classified pixel	
		Annual Crop	Aquaculture	Bamboo Forest	Bare Soil	Coniferous Forest	Deciduous Forest	Evergreen Forest	Flooded Forest	Forest Plantation	Grassland	Industrial Plantation	Mangrove	Marsh/Swamp Area	Orchard	Paddy Rice	Shrubland	Urban Area		Water Body
Ground truth data	Annual Crop	202	0	0	0	0	0	7	8	0	0	1	3	7	4	9	5	5	0	251
	Aquaculture	1	96	0	0	0	0	0	1	0	0	0	0	1	0	8	0	0	7	114
	Bamboo Forest	0	0	30	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	33
	Bare Soil	1	0	0	43	0	0	0	0	0	0	0	0	0	0	1	0	0	0	45
	Coniferous Forest	0	0	0	0	15	1	1	0	0	0	0	0	0	0	0	0	0	0	17
	Deciduous Forest	15	0	0	0	0	135	12	0	0	0	0	0	0	1	2	5	0	0	170
	Evergreen Forest	2	0	0	0	0	10	190	1	0	0	0	0	2	0	2	0	0	0	207
	Flooded Forest	0	1	0	0	0	0	0	31	0	0	0	0	1	0	1	0	0	0	34
	Forest Plantation	0	0	0	0	0	0	5	2	132	0	0	0	0	0	2	5	0	0	146
	Grassland	0	0	0	1	0	2	0	1	0	19	0	2	1	0	5	1	0	0	32
	Industrial Plantation	0	0	0	0	0	9	3	0	0	1	89	0	0	1	2	4	0	0	109
	Mangrove	0	0	0	0	0	1	0	0	0	0	0	45	0	0	3	0	0	1	50
	Marsh/Swamp Area	0	1	0	0	0	0	0	1	0	0	0	1	22	0	2	0	0	1	28
	Orchard	2	0	0	0	0	3	5	0	0	2	2	3	12	189	2	11	2	0	233
	Paddy Rice	5	7	0	0	0	0	0	0	0	1	0	0	5	1	481	1	3	3	507
	Shrubland	0	0	0	0	0	0	0	0	0	0	0	0	0	4	2	15	0	0	21
Urban Area	0	0	0	0	0	0	0	0	0	0	0	0	2	4	8	3	45	0	62	
Water Body	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	136	139	
Number of ground truth data		228	106	30	44	15	163	224	45	132	23	92	54	53	204	531	50	56	148	2198

4.5 Updating of the LULC 2020 map

The LULC 2020 map represents the LMB at a regional scale, at a resolution of 10 m. The LULC 2020 classification can be broadly classified into major components based on the area: forest, rice paddy, and annual crops.

Forest is the prevalent land cover in the LMB. The forest comprises two main types of classification, evergreen and deciduous forest, whose area covers 28.11% and 12.13%, respectively. Most of the area is in Lao PDR and the eastern part of Cambodia.

Approximately 22.39% of the LMB is dedicated to rice paddy production (yellow areas on the map), which dominates the vast low-lying alluvial plains of the Chi-Mun Basin in northeast Thailand, the Vientiane plain in Lao PDR, the Tonlé Sap Basin in Cambodia, and the Delta in

southern Viet Nam. Narrower river valleys in northern Thailand and central and southern Lao PDR are also utilized for paddy. Smaller paddy areas adjacent to rivers and streams in northern Lao PDR are too small to map but are nonetheless highly important economically in the otherwise steep topography of northern Lao PDR.

Annual crops is the third most prevalent land use type that covers the LMB. The orange areas on the map represent the annual cropland in 2020. The total area of the annual crop land in the entire basin is calculated at 15%, i.e. approximately 92,747 km². The distribution area of the annual crops in LMB are 29,104 km² in Cambodia, 18,070 km² in Lao PDR, 37,765 km² in Thailand, and 7,808 km² in Viet Nam.

Mangrove forest is only found in Viet Nam, at 0.15% of the area of LMB, since the boundary in LMB of Cambodia, Lao PDR, and Thailand are not connected to the ocean.

In addition to the LMB regional LULC map, the country-wise LULC 2020 maps are presented in Figures 30 to 34. Table 9 shows the percentage of the area for each classification.

Table 9. LULC 2020 classification area

Class	Area (sq.km)					Area (%)				
	Cambodia	Lao PDR	Thailand	Vietnam	LMB	Cambodia	Lao PDR	Thailand	Vietnam	LMB
Annual Crop	29,104	18,070	37,765	7,808	92,747	18.6%	8.7%	20.1%	11.7%	15.00%
Aquaculture	229	215	1,070	8,397	9,911	0.1%	0.1%	0.6%	12.5%	1.60%
Bamboo Forest	124	3,127	83	182	3,517	0.1%	1.5%	0.0%	0.3%	0.57%
Bare Soil	982	481	606	166	2,235	0.6%	0.2%	0.3%	0.2%	0.36%
Coniferous Forest	14	753	-	2,103	2,870	0.0%	0.4%	0.0%	3.1%	0.46%
Deciduous Forest	28,836	23,384	17,160	5,644	75,023	18.4%	11.3%	9.1%	8.4%	12.13%
Evergreen Forest	27,599	114,108	23,029	9,074	173,810	17.6%	55.1%	12.3%	13.6%	28.11%
Flooded Forest	9,075	12	2	1,105	10,195	5.8%	0.0%	0.0%	1.7%	1.65%
Forest Plantation	3	10	208	1	222	0.0%	0.0%	0.1%	0.0%	0.04%
Grassland	287	3,249	727	602	4,865	0.2%	1.6%	0.4%	0.9%	0.79%
Industrial Plantation	3,718	1,774	9,599	2,873	17,965	2.4%	0.9%	5.1%	4.3%	2.91%
Mangrove	-	-	-	915	915	0.0%	0.0%	0.0%	1.4%	0.15%
Marsh/Swamp Area	2,627	2,843	1,776	1,814	9,061	1.7%	1.4%	0.9%	2.7%	1.47%
Orchard	4,963	19,687	4,874	5,336	34,860	3.2%	9.5%	2.6%	8.0%	5.64%
Others	1,637	2,847	1,125	1,325	6,934	1.0%	1.4%	0.6%	2.0%	1.12%
Paddy Rice	38,712	10,847	73,807	15,119	138,485	24.8%	5.2%	39.3%	22.6%	22.39%
Shrubland	1,322	1,491	2,633	863	6,309	0.8%	0.7%	1.4%	1.3%	1.02%
Urban Area	1,961	932	10,107	1,193	14,194	1.3%	0.4%	5.4%	1.8%	2.30%
Water Body	5,203	3,396	3,287	2,412	14,298	3.3%	1.6%	1.7%	3.6%	2.31%
Grand Total	156,398	207,227	187,859	66,932	618,416	100.0%	100.0%	100.0%	100.0%	100.00%

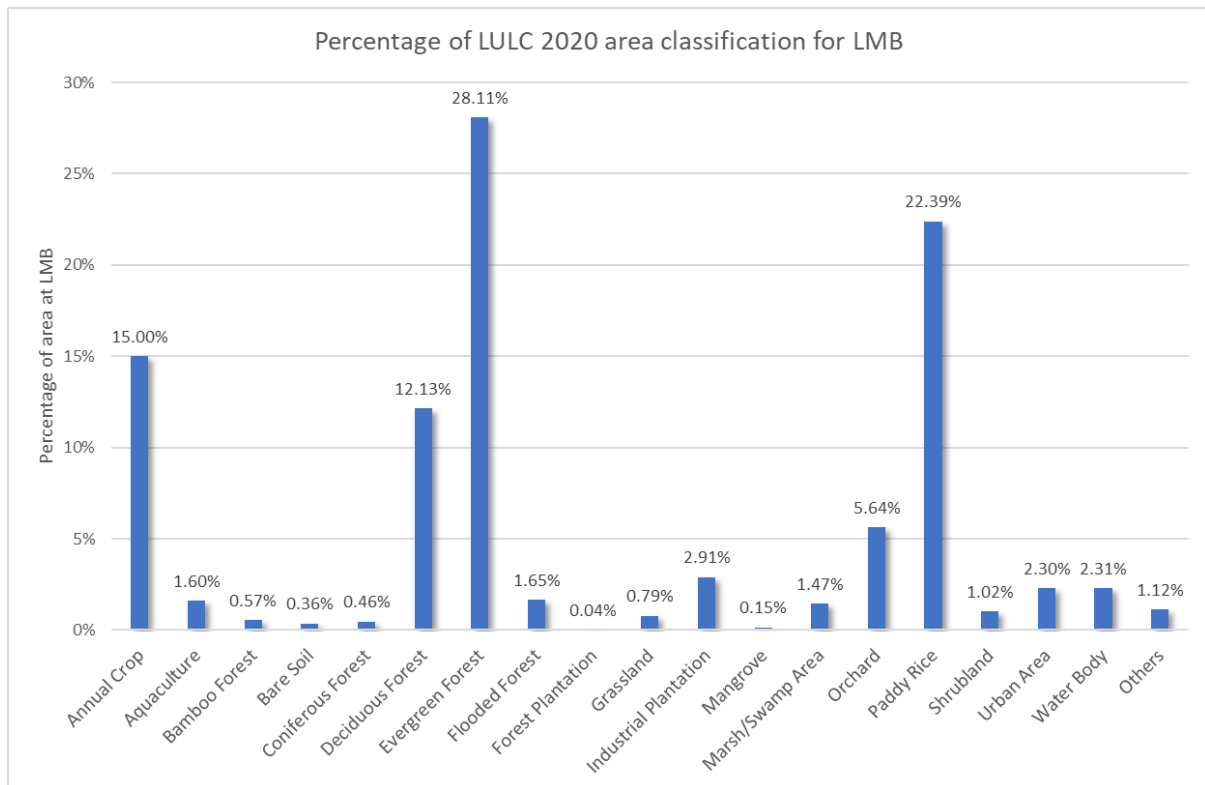


Figure 29. LULC 2020 area

Land Use Land Cover 2020

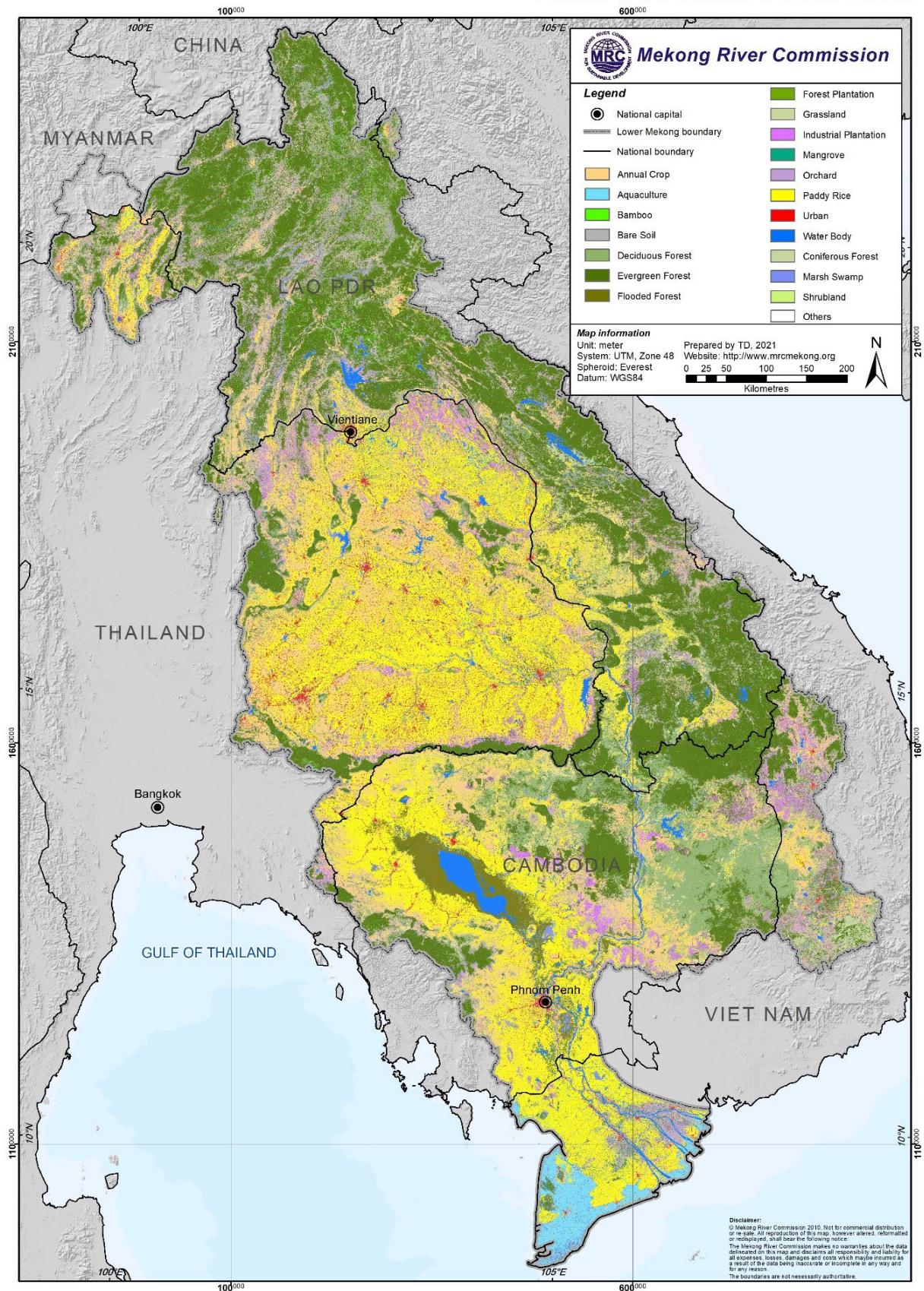


Figure 30. LULC map of 2020 in the Lower Mekong Basin

Land Use Land Cover 2020, Cambodia

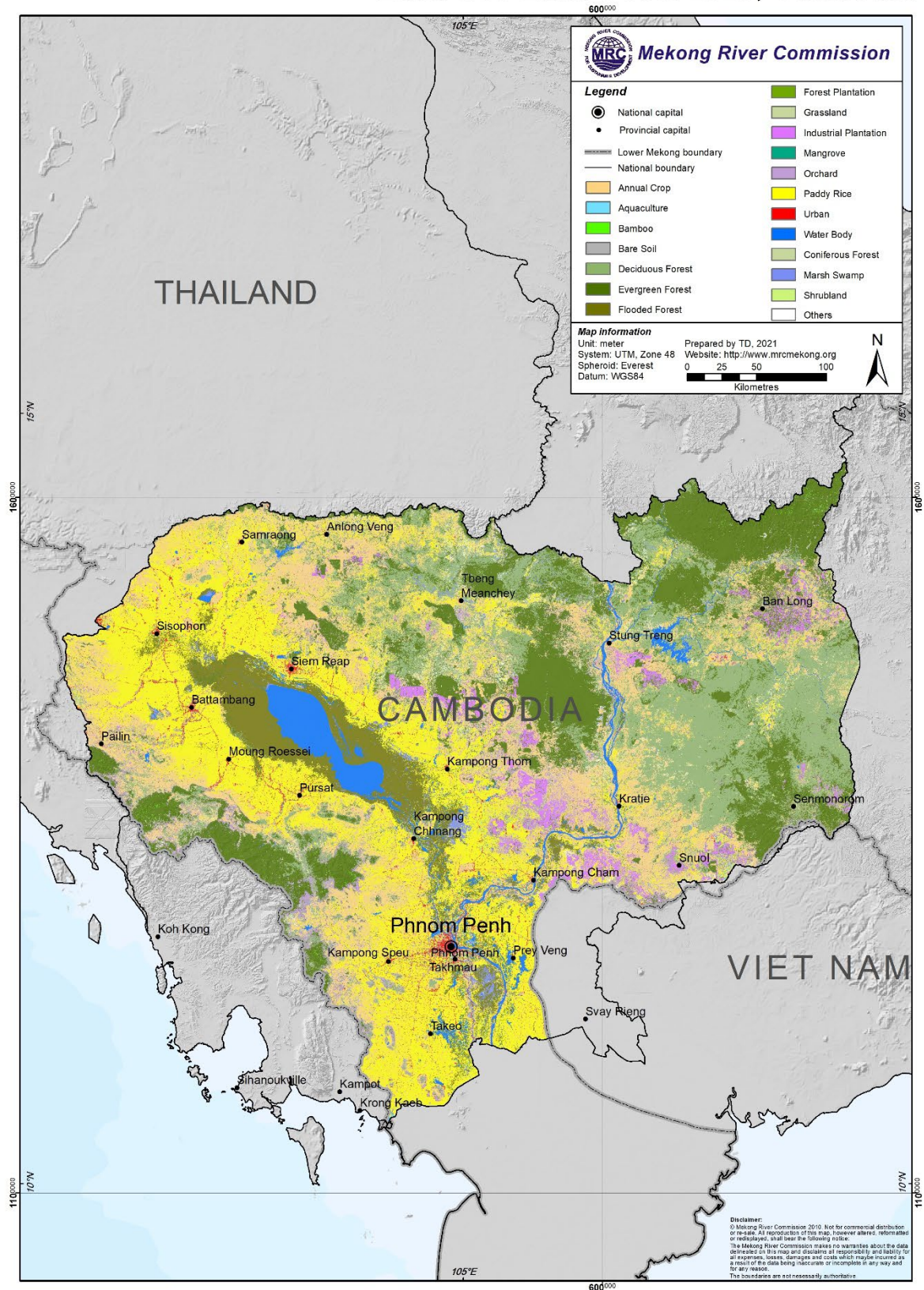


Figure 31. LULC of 2020 in Lower Mekong Basin for Cambodia

Land Use Land Cover 2020, Lao PDR

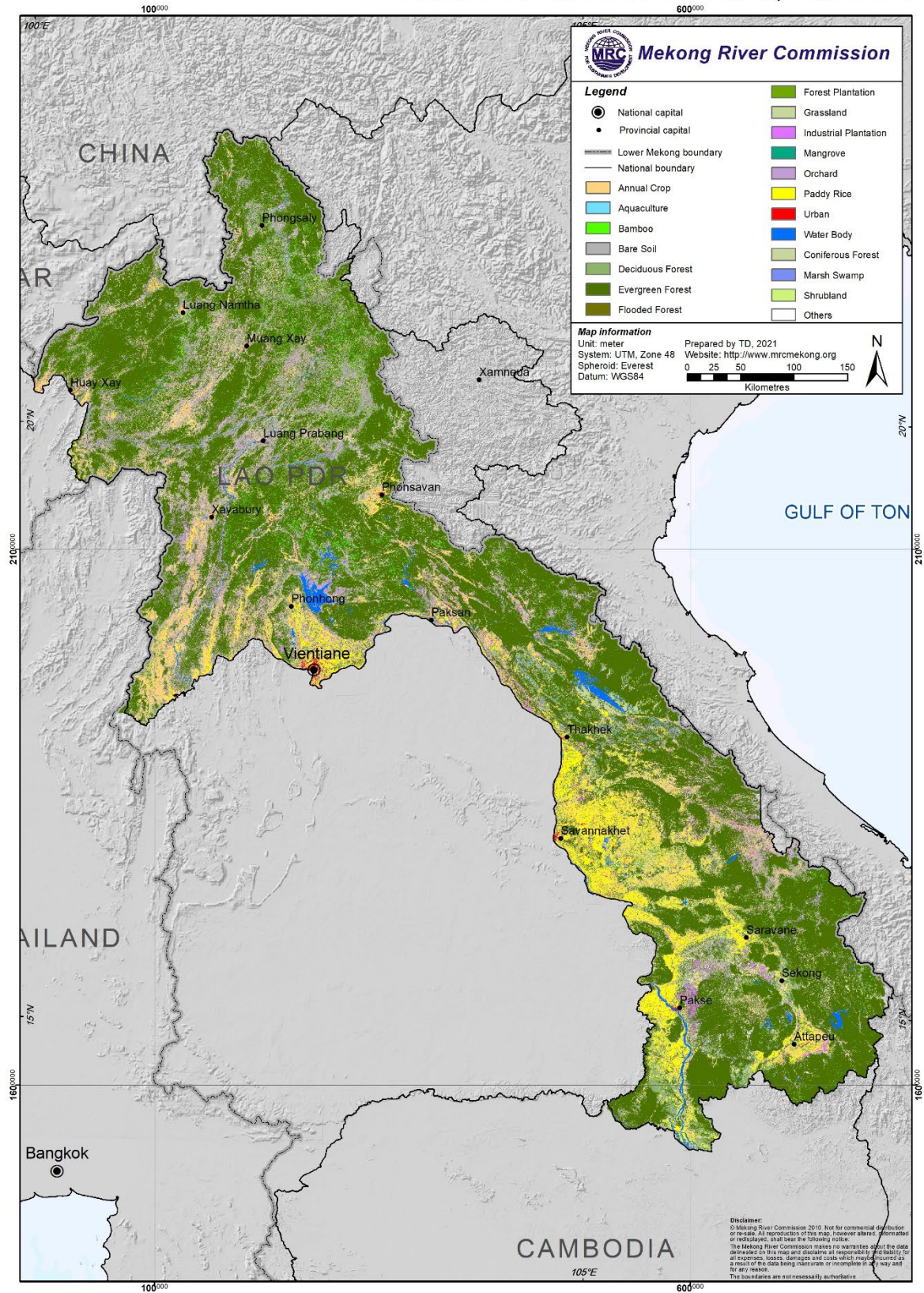


Figure 32. LULC map of 2020 in Lower Mekong Basin for Lao PDR

Land Use Land Cover 2020, Thailand

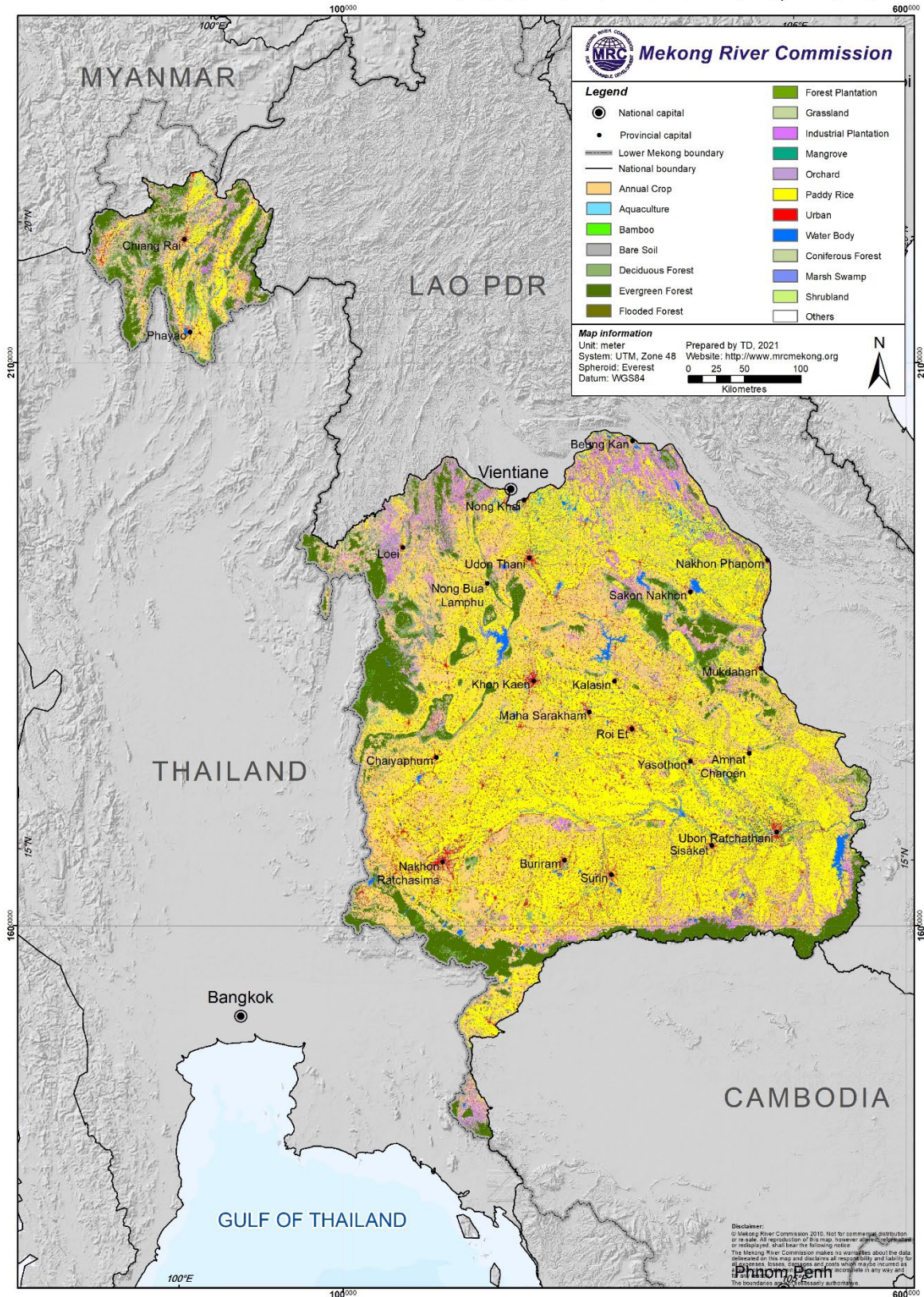


Figure 33. LULC of 2020 in Lower Mekong Basin for Thailand

Land Use Land Cover 2020, Vietnam

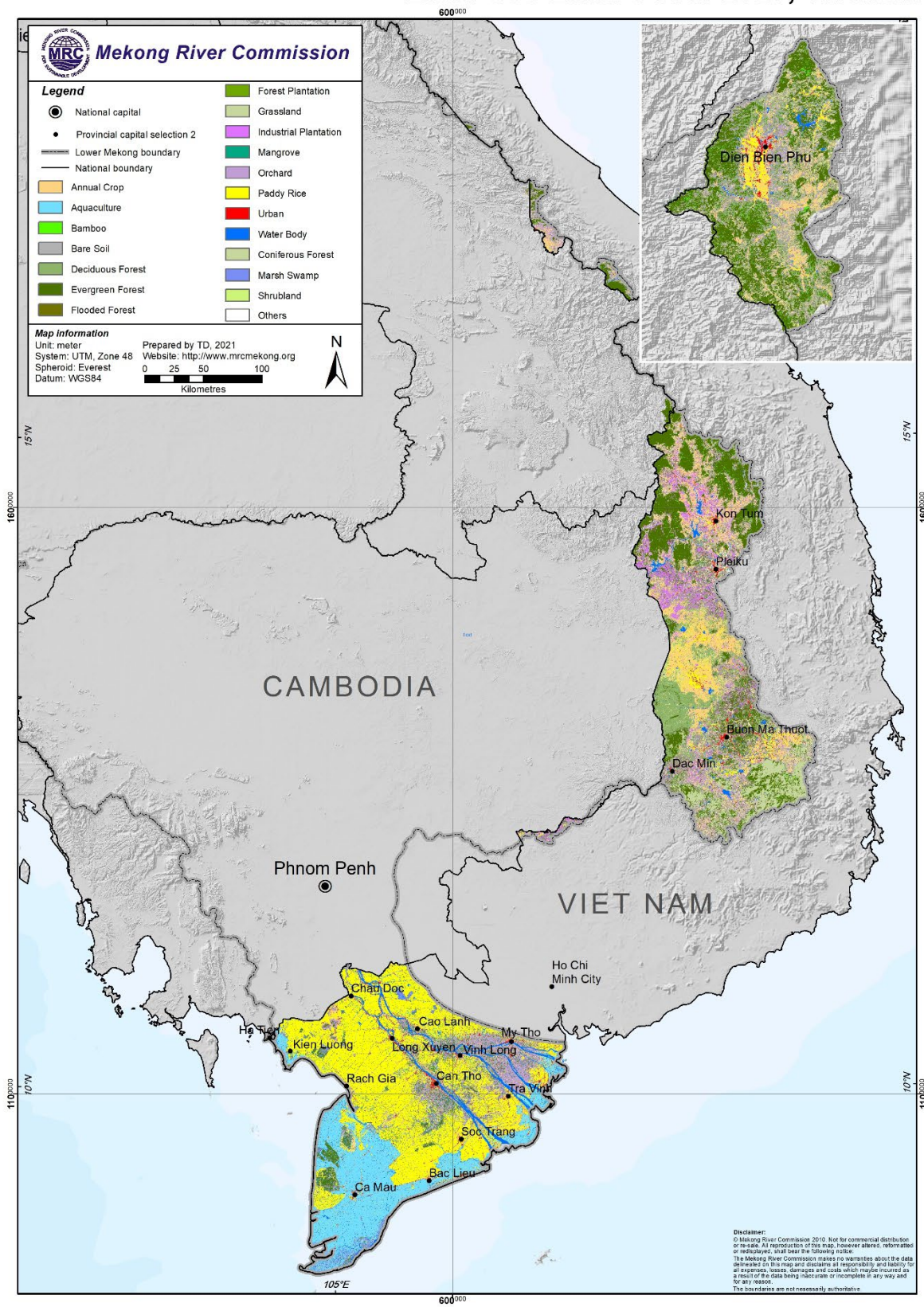


Figure 34. LULC map of 2020 in Lower Mekong Basin for Viet Nam

4.6 Land use change analysis

The land change analysis was developed to compare the static area of each land category based on the MRC data of 2003, 2010, and 2020, as shown in Table 10.

Table 10. Comparison of land changing in the LMB

Class Name	2003		2010		2020	
	Area (sq.km)	Percentage	Area (sq.km)	Percentage	Area (sq.km)	Percentage
Annual Crop	42,500	6.80%	52,461	8.39%	92,747	15.00%
Paddy Rice	154,995	24.81%	140,540	22.47%	138,485	22.39%
Shifting Cultivation	14,242	2.28%	9,724	1.56%	-	-
Orchard	3,663	0.59%	12,123	1.94%	34,860	5.64%
Flooded Forest	4,360	0.70%	4,886	0.78%	10,195	1.65%
Grassland	13,880	2.22%	8,637	1.38%	4,865	0.79%
Shrubland	20,988	3.36%	70,587	11.29%	6,309	1.02%
Urban Area	15,690	2.51%	15,780	2.52%	14,194	2.30%
Bare Soil	2,851	0.46%	3,843	0.61%	2,235	0.36%
Industrial Plantation	4,760	0.76%	25,343	4.05%	17,965	2.91%
Deciduous Forest	133,024	21.30%	180,436	28.85%	75,023	12.13%
Evergreen Forest	186,798	29.91%	65,177	10.42%	173,810	28.11%
Forest Plantation	480	0.08%	1,498	0.24%	222	0.04%
Bamboo Forest	9,167	1.47%	5,699	0.91%	3,517	0.57%
Coniferous Forest	232	0.04%	3,900	0.62%	2,870	0.46%
Mangrove	1,839	0.29%	1,303	0.21%	915	0.15%
Marsh/Swamp Area	913	0.15%	1,866	0.30%	9,061	1.47%
Aquaculture	2,101	0.34%	6,886	1.10%	9,911	1.60%
Water Body	12,135	1.94%	14,667	2.35%	14,298	2.31%
Others	-	-	-	-	6,934	1.12%

Table 10 represents the static area of each land cover category for each study year starting from 2003, 2010, and 2020.

In general, the trend of increasing areas for each land class in 2020 is similar to the trend of 2003 but it is slightly different from 2010. The comparison found that the area of annual crop significantly increased, from less than 10% in both 2003 and 2010, to 15% in 2020, which is almost 40,000 km².

Orchard area in 2020 significantly increased from 0.59% in 2003 to 1.94% in 2010 and then to 5.64% in 2020.

It can be observed that the urban area in 2020 slightly changed, from 2.51% in 2003 and 2.52% in 2010, and 2.30% in 2020. This may be due to the higher resolution of the satellite imagery, which was able to detect much more and matches the actual details of the urban area.

Shifting cultivation is not classified for 2020 due to the limitation of the time series of the satellite image for the study. In addition, some area might be mixed between one category and another, which cannot be classified clearly. The category of "others" developed for this reason.

The overall distribution of the area comparison for LULC changes in 2003, 2010, and 2020 is shown in Figure 35.

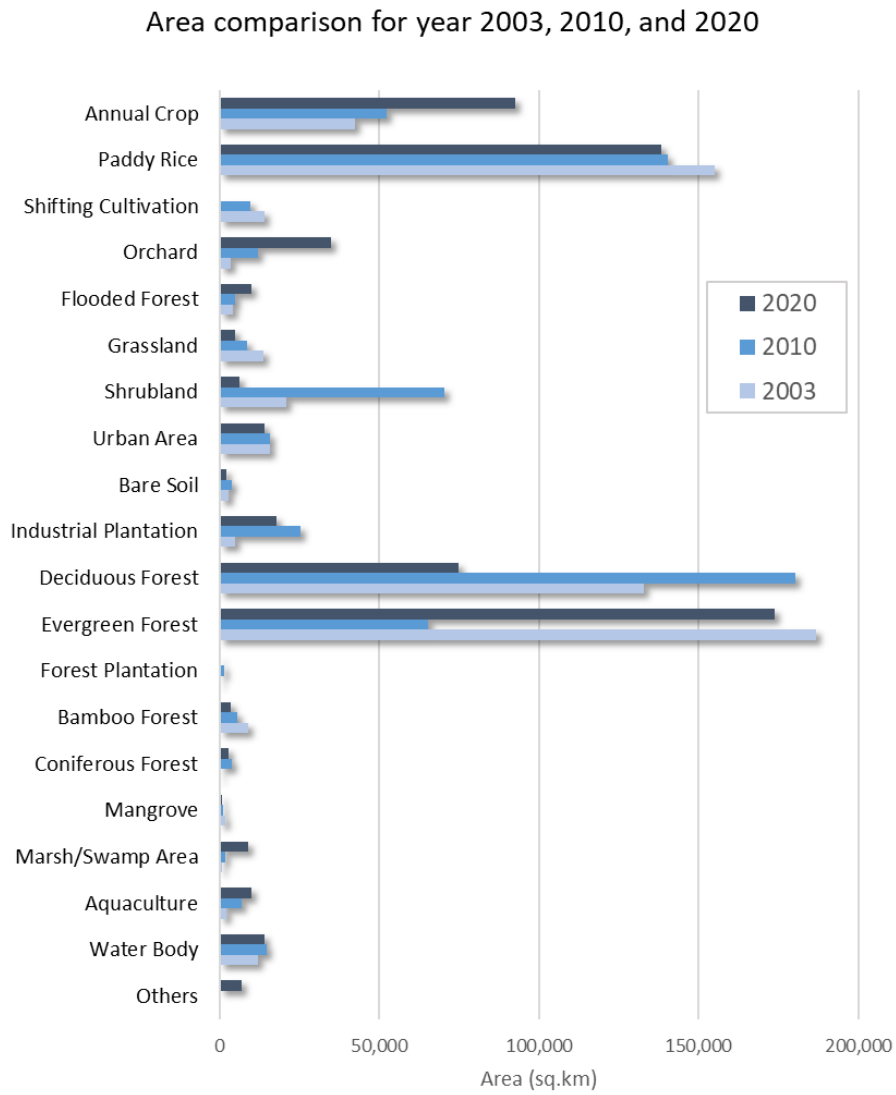


Figure 35. Area comparison for LULC change in 2003, 2010, and 2020

5 Conclusions and recommendations

5.1 Conclusions and recommendations

This study presented the process of using machine learning methods to develop the updated, high-resolution land cover map of the LMB region. The method was used to fuse optical and SAR imagery, a combination of Sentinel-2, Landsat-8 and Sentinel-1, and Planet imagery, which provide sufficient spectral and spatial information to map the detailed land use land cover class that refers to the MRC LULC 2010 classification.

The LULC 2020 is the result of the above-mentioned processes and was created through collaboration between the remote sensing expert of the MRC with experts from Member Countries and SERVIR-Mekong.

Uncertainty about the time of data collection compared to the period when the satellite imagery was taken could affect the results in some of the LULC classification types. For example, some of the area might be paddy field during the data collection period, but during the period when the satellite imagery was taken, the area might have changed to aquaculture.

Reference training data for the model obtained from the field survey data collection were critical in carefully deciding on information and the location of the field data collection. The details of the information of the different types of LULC must be clear before designing the questionnaire data collection form. This would avoid the confusion of field data collection. Field data for land cover is often difficult to obtain, but there is an optional platform that could help by using higher-resolution satellite imagery, such as Collect Earth Online, offer exciting opportunities to collect new data.

The new technology of cloud computing platform for remote sensing data analysis used in this study saves a great deal of time for the analysis. It does not require any GIS/RS software nor huge storage for the data.

To improve the LULC in the future, time series of the satellite data (1–2 years) would help analyse on complex classification (e.g. shifting cultivation, annual crops) to observe the trend of the land change during the year of analysis. The visual interpretation from the local LULC expert could also contribute to the validation and correction of the results.

5.2 Lesson learned


Lesson learned and the limitation of the work are valuable information to take into consideration for further improving the LULC mapping in the future and to acquire experience on which activities should be considered and which should be avoided. The following information should be considered:

- Due to the COVID-19 pandemic, project was delayed from the start. The field data collection was affected due to the 'Work From Home (WFH)' modality. The line agency was not able to go to the field and collect the data.
- During the work, the MRC lacked a remote sensing specialist for months. Most of the LULC work cannot be carried out during this period. This was one of the reasons for the delay.
- The questionnaire for data collection needs to be agreed on among the Member Countries with regard to the type of data and land classification in order to avoid the confusion that could arise during the analysis of the LULC process.
- Classification of the LULC can be updated according to the needs of the Member Countries in the future.
- There is a trade-off between the resolution and the size of the homogenous zone of the LULC; The higher resolution, the smaller the homogenous zone. Hence, high-resolution data (10–15 m) would allow to carry out a more detailed classification of the land. Homogeneous areas would be difficult to distinguish because land classification varies within a small area. However, if the lower resolution had been used (e.g. MODIS 250 m), it would have been easier to observe the homogenous areas due to the larger pixel size. The size of pixels on the high resolution LULC could be a challenge for the user with low internet connection bandwidth.
-
- The methodology and satellite resolution of developing LULC 2020 are different from the methodology used for 2003 and 2010. Some of the land classification might not have the same results as in 2003 and 2010.
- With the new approach of the methodology for updating the LULC 2020 by using the cloud computing platform, the time for the development and analysis can be significantly shortened due to the performance of the cloud computing.

Annex

Annex 1: Field Data Collection Form

FIELD DESCRIPTION CARD: Land Use and Land Cover Mapping 2020 Products in LMB



A. GENERAL INFORMATION

RELEVÉE N°

AREA NAME

LOCATION

OBSERVER

DATE

TIME

RELEVÉE SIZE (in m² or ha)

FIELD SAMPLE COORDINATES

North	East

ACCESSIBILITY

Very Good

Good

Medium

Bad

COORDINATES

North	East	LAT/LONG	UTM	GPS	Topo Map
<input style="width: 100%;" type="text"/>	<input style="width: 100%;" type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input style="width: 100%;" type="text"/>	<input style="width: 100%;" type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input style="width: 100%;" type="text"/>	<input style="width: 100%;" type="text"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

On the spot

Indicate Relative Position of Coordinates



Observing the spot from a distance

Distance from viewpoint to observed point (m)

The bearing of the observed point (°)

FIELD PHOTOGRAPHS



Film Roll N°	<input style="width: 100%;" type="text"/>
Photo Shot No	Position
N°	
<input style="width: 100%;" type="text"/>	<input style="width: 100%;" type="text"/>
<input style="width: 100%;" type="text"/>	<input style="width: 100%;" type="text"/>
<input style="width: 100%;" type="text"/>	<input style="width: 100%;" type="text"/>
<input style="width: 100%;" type="text"/>	<input style="width: 100%;" type="text"/>

Relative Position of photograph

GENERAL LANDFORM

Slope

- Flat to Gently Sloping Terrain (0 - 7 %)
- Gently Sloping to Moderately Sloping (8 - 30 %)
- Sloping to Moderately Steep, Undulating to Rolling terrain (14 - 20 %)
- Steep to Very steep, Rolling to Hilly Terrain (21 - 55 %)
- Extremely Steep Terrain, Steeply Dissected Hilly and Mountainous Terrain (56 - 140 %)

B. GENERAL LULC INFORMATION

LULC Type

- General LULC Type
Relevee Site

A Vegetated Non-Vegetated

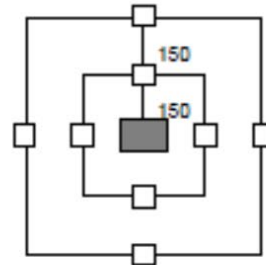
B Terrestrial Aquatic or Regularly Flooded Land
(Including Water Body Areas)

- Specific LULC Type

	Single Major LULC Aspect	Two Mixed Major LULC Aspects	
		Most Important	Second
- Cultivated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- Natural/Semi-Natural	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- Built-up	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- Bare	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- Artificial Water Inland	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
- Inland Water	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

AREA LULC HOMOGENITY (Applicable if on spot)

LULC Homogeneous for more than 300 m
around the sample area: Yes
 No



LULC SEASONAL ASPECTS

	Natural / Semi-Natural Vegetation				Cultivated Fields			
	dry	green	flowering	fruits	ploughed	initial stage	full maturity stage	harvested
TREE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
SHRUB	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HERB	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

C. SPECIFIC LULC INFORMATION

NATURAL AND SEMI-NATURAL VEGETATION

	Level	Cover	Height	Leaf Type			Leaf Phenology	
				Broad	Needle	Aphyllous	Evergreen	Deciduous
WOODY	Trees	1		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		2		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
		3		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Shrubs	1			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	2			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
HERBACEOU	Graminoids							
		Forbs						

Cover Estimation of vegetation Visual Instrumental Other

CULTIVATED TERRESTRIAL AREA AND MANAGED LAND

- Life Form of **MAIN CROP**

	Leaf Type		Leaf Phenology		Fruit Trees	Plantation
	Broad	Needle	Evergreen	Deciduous		
<input type="checkbox"/> Trees	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Shrubs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/> Herbaceous						
<input type="checkbox"/> Graminoids						
<input type="checkbox"/> Other						

- Crop Name:

- Life Form of **SECOND CROP**

	Leaf Type		Leaf Phenology		Fruit Trees	Plantation
	Broad	Needle	Evergreen	Deciduous		
<input type="checkbox"/> Trees	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Shrubs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/> Herbaceous						
<input type="checkbox"/> Graminoids						
<input type="checkbox"/> Other						

- Crop Name:

- Average Field Size (m² or ha)

- Field Distribution

<input type="checkbox"/> Bordering Fields
<input type="checkbox"/> Distance between fields < average field size
<input type="checkbox"/> = 1 to 3 X average field size
<input type="checkbox"/> = 3 to 9 X average field size
<input type="checkbox"/> > 9 X average field size

- Cultivation period

<input type="checkbox"/> main crop, during two or more different periods within same year
<input type="checkbox"/> second crop in same period as main crop
<input type="checkbox"/> second crop in different period as main crop
<input type="checkbox"/> second crop starts during active period main crop

- Cultivation Time Factor

Time lap between two consecutive active periods	<input type="checkbox"/> =<1 year
	<input type="checkbox"/> 1 to 4 years
	<input type="checkbox"/> > 4 years

- Water Supply / Irrigation

<input type="checkbox"/> Not Irrigated	<input type="checkbox"/> Postfloodind
	<input type="checkbox"/> Surface
	<input type="checkbox"/> Sprinkler
<input type="checkbox"/> Supplementary Irrigation	<input type="checkbox"/> Drip
	<input type="checkbox"/> Other <input type="text"/>

- Life Form **MANAGED LAND**: Urban Vegetated Area covered by trees is::

<input type="checkbox"/> > 40 %
<input type="checkbox"/> between 20 % and 40 %
<input type="checkbox"/> < 20 %

BARE AREAS

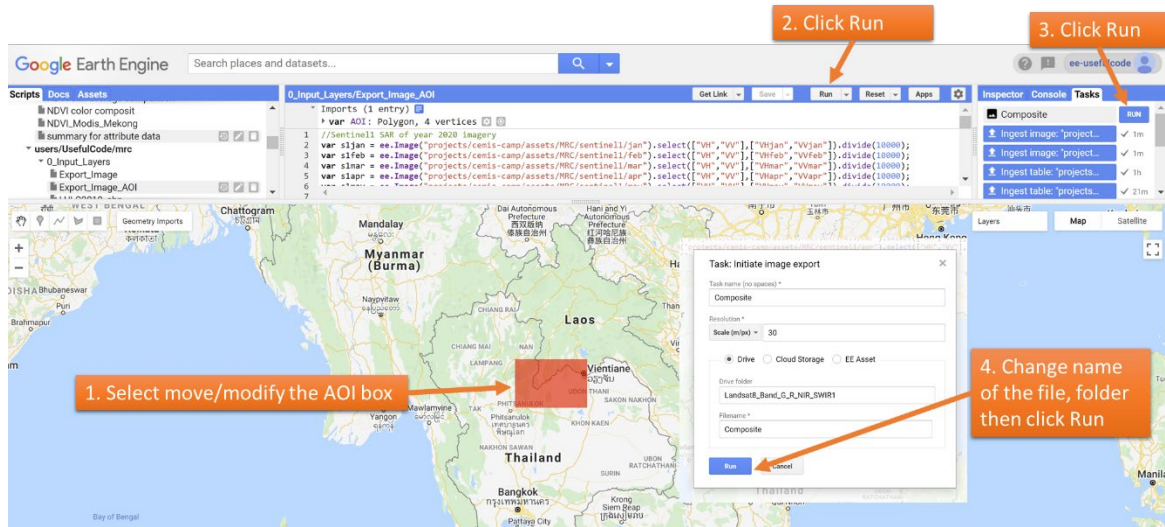
<input type="checkbox"/> Consolidated	<input type="checkbox"/> Bare Rock	
	<input type="checkbox"/> Gravel, Stones and Boulders	
	<input type="checkbox"/> Hardpans	
<input type="checkbox"/> Unconsolidated	<input type="checkbox"/> Bare Soil	<input type="checkbox"/> Stony (5 - 40%)
	<input type="checkbox"/> Loose and shifting sand	<input type="checkbox"/> Very Stony (40 - 80%)
<input type="checkbox"/> Dunes	<input type="checkbox"/> Barchans	<input type="checkbox"/> Saturated
	<input type="checkbox"/> Parabolic	<input type="checkbox"/> Unsaturated
	<input type="checkbox"/> Longitudinal	

Annex 2: GEE Code to Generate a Composite Satellite Image and Download

Below is the GEE code to download the satellite images by area of interest (AOI)

<https://code.earthengine.google.com/b304ba9a0a996b1351db7b19100cb25c>

The instructions to download are shown below.



Reference for the satellite image composite depository in GEE

Products	GEE code
Landsat 8 - Image composite	https://code.earthengine.google.com/8130765ccae38c56fb29de088292655f
Sentinel- 2 Image composite	https://code.earthengine.google.com/402877f25da39dac19e3645451efc1b7

Annex 3: Example of Field Data Collection Images



N	Coordinate X_N1 489226	Coordinate Y_N1 1790346			E
	Coordinate X_E1 490076	Coordinate Y_E1 1790196			
W	Coordinate X_W1 489776	Coordinate Y_W1 1790206			S
	Coordinate X_S1 489918	Coordinate Y_S1 1790046			
	Coordinate X_W2 489776	Coordinate Y_W2 1790496			
	Coordinate X_E2 490226	Coordinate Y_E2 1790206			
	Coordinate X_W2 489226	Coordinate Y_W2 1790206			
	Coordinate X_S2 489926	Coordinate Y_S2 1789996			

Annex 4: Access to Training Datasets

<i>ID</i>	LULC types	Number of points each primitive	Number of reference data for other class	GEE Code to reference data
1	Annual crop	11,002	11,676	https://code.earthengine.google.com/f10377d1164e14a16ea96aa1cc378d36
2	Paddy rice	8,577	9,463	https://code.earthengine.google.com/099212f086274423a025d2abe31d9e5e
3	Orchard	6,975	6,132	https://code.earthengine.google.com/96f67d8776e8015c38696e2a443b56d1
4	Flooded forest	1,261	1,370	https://code.earthengine.google.com/fd1a079be1cb4b31a756925e68d41a2a
5	Grassland	2,039	2,259	https://code.earthengine.google.com/fb33b6ae04986a5c2ec2db5db8929912
6	Shrubland			https://code.earthengine.google.com/6ca9a37f9f8cbc21f4c1e270a82233a3
7	Urban area	9,174	9,119	https://code.earthengine.google.com/1ce4a8c803771fe8a64109b3a89556c7
8	Bare soil	1,332	1,220	https://code.earthengine.google.com/cf95858c18b39516081193680521c82f
9	Industrial plantation	8,260	9,332	https://code.earthengine.google.com/102f3784752d4755e9b896cf919e944e
11	Evergreen forest	6,421	6736	https://code.earthengine.google.com/25c1acc7835b1744e0ec0cecb41cc802
12	Bamboo forest	11,138	11,137	https://code.earthengine.google.com/39b2fa2b7f1476663fd0afc6a5e4f40f
13	Marshes/swamp	907	2,230	https://code.earthengine.google.com/df3fb9e0a9245f81d9a0b46713af3516
14	Aquaculture	1,987	2,351	https://code.earthengine.google.com/9483ab64b70a3e178625db6455f421b3
15	Water body	5,672	5758	https://code.earthengine.google.com/d33917c77ed8fc92c42d905646392a87
16	Forest plantation	4,165	4,176	https://code.earthengine.google.com/bcb9c4f615efaed259672f49df60b20b
17	Coniferous forest			https://code.earthengine.google.com/0d6248287f242b7c4e6f1a427c9852a9
18	Mangrove forest			https://code.earthengine.google.com/5563fdceb38461f369d3fd36c14d98d3
19	Deciduous forest			https://code.earthengine.google.com/03449d01470f01cde5eaae274642f08a

Annex 5: List of GEE Codes of Primitive Models for the LULC Map of 2020

LCC	Land cover types	Description	Primitives	GEE Code
1	Annual crop	Mandatory herbaceous growth forms; cultivated and managed vegetation	Crop	https://code.earthengine.google.com/fb46da0e264208835223763f8e691287
2	Paddy field	Mandatory Graminae; cultivated and managed vegetation with field size:0.2–2.0 hectares; rice species	Paddy	https://code.earthengine.google.com/7b4f12f05749e96d4544ea0f05588352
3	Shifting cultivation	Temporary sequence between herbaceous growth forms of cultivated and managed vegetation and woody growth forms of natural/semi-natural vegetation; sequency length: 3 to 7 years	Not include in the map	
4	Orchard	Mandatory trees; cultivated and managed vegetation of orchard and other plantation; field size: 1–3 hectares	Tree plantation/crop	https://code.earthengine.google.com/b7a71663faeb1064e72718d5b1807f76
5	Flooded forest	Multi-stratum of mandatory trees/shrubs/herbaceous; natural or semi-natural vegetation; cover 10–100%.	Flooded forest	https://code.earthengine.google.com/f6f842c969ab1a7aa0d460da8451f906
6	Grassland	Mandatory herbaceous growth forms; natural or semi-natural vegetation; cover 10–100%	Grass	https://code.earthengine.google.com/030b293bb375931e7ab2b73c59d017dd
7	Shrubland	Mandatory shrubs; natural or semi-natural vegetation; cover:v10–80%	Shrubland	https://code.earthengine.google.com/730d4169d95de44ca5ffefb783c578b8?
8	Urban area	Mandatory built-up surface	Urban	https://code.earthengine.google.com/d311fc02d1be104385ae7a7e6884297b
9	Bare land	Mandatory bare soil; optional coarse mineral fragments (stone:v1–40%)	Barren	https://code.earthengine.google.com/b5a50a48598cefe06fd033015ab8196b
10	Industrial plantation	Mandatory of woody growth form; cultivated and managed vegetation of orchard and other plantation; species of industrial crops	Industrial plantation	https://code.earthengine.google.com/71a4f91dfa26bb74a887c0eef8d8efce
11	Deciduous forest	Mandatory Trees; cover 10–100%; natural or semi-natural vegetation; deciduous and broadleaved	Deciduous	https://code.earthengine.google.com/708879923cc6fb3a18805538fe53ae8a

LCC	Land cover types	Description	Primitives	GEE Code
12	Evergreen forest	Mandatory Trees; cover 10–100%; natural or Seminatatural vegetation; evergreen	Evergreen	https://code.earthengine.google.com/9695b9fd0e325f9750306d7ca57764fd
13	Bamboo forest	Mandatory Woody growth forms; cover: 10–100%; height: 4–15 m; natural or semi-natural vegetation; species of bamboo	Bamboo	https://code.earthengine.google.com/cf930b45c1e1ccb64432c91760146245
14	Marshes/swamp area	Mandatory herbaceous growth forms	Marsh Swamp	https://code.earthengine.google.com/3be257e64e81a182b6c9f9c44be02c66
15	Aquaculture	Mandatory artificial water body above surface; general aquaculture	Aquaculture	https://code.earthengine.google.com/ce5341d8bedce0672e2c079ae1c14dda
16	Water body	Mandatory periodic variations water body above surface; fresh water	Water	https://code.earthengine.google.com/59db9321e2df25b5de53b5a6f6de4c05

Assemblage Code: <https://code.earthengine.google.com/8409d077124a8fecc535017d5ca85764>

References

- Binder, K. H. (1993). Monte Carlo simulation in statistical physics. *Computers in Physics*, 7 (2), *Monte Carlo Simulation in Statistical physics. Computers in Physics*, 7 (2), 156–157. <https://doi.org/10.1007/978-3-642-03163-2>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen, B. X. (2017). A mangrove forest map of China in 2015: Analysis of time series Landsat 7/8 and Sentinel-1A imagery in Google Earth Engine cloud computing platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 131, 104–120. <https://doi.org/10.1016/j.isprsjprs.2017.07.011>
- Colby, J. D. (1991). *Topographic normalization in rugged terrain. Photogrammetric Engineering and Remote Sensing*.
- Hansen, M. C. (2011). Continuous fields of land cover for the conterminous United States using Landsat data: First results from the Web-Enabled Landsat Data (WELD) project. *Remote Sensing Letters*, 2(4), 279–288.
- Hansen, M. C. (2016). Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and (Placeholder1) 8 data. *Remote Sensing of Environment*, 185, 221–232.
- Housman, I. S. (n.d.). A quantitative evaluation of cloud and cloud shadow³¹² masking algorithms available in Google Earth Engine.
- Housman, I. W. (2018). An evaluation of forest health insect and disease survey data and satellite-based remote sensing forest change detection methods: Case studies in the United States. *Remote Sensing*, 10(8), 1184.
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote sensing of environment*, 25(3), 295–309.
- Kityuttachai, K. H. (2016). Land Cover Map of the Lower Mekong Basin, MRC Technical Paper No. 59. Phnom Penh, Cambodia: Information and Knowledge Management Programme, Mekong River Commission.
- Lee, J. S.-2. (2008). Improved sigma filter for speckle filtering of SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1), 202-213.
- Liaw, A. & Weiner, M. (2002). Classification and regression by random Forest. *R news*, 2 (3), 18–22.
- Markert, K. N. (2018). On the merging of optical and SAR satellite imagery for surface water mapping applications. 9, 275–277.

- McFeeters, S. (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.
- Pekel, J. F. (2016). High-resolution mapping of global surface water and its long-term changes. 540(7633), 418–422.
- Poortinga, A. T. (2019). Mapping plantations in Myanmar by fusing Landsat-8, Sentinel-2 and Sentinel-1 data along with systematic error quantification. *Remote Sensing*, 11(7), 83.
- Riaño, D. C.-T.-1. (2003). Riaño, D., Chuvieco, E., Salas, J., & Aguado, I. Assessment of different topographic corrections in Landsat-TM data for mapping vegetation types (2003). *IEEE Transactions on geoscience and remote sensing*, 41(5), 1056–1061.
- Rouse, J. H. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA special publication*, 351, 309.
- Roy, D. P. (2008). Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sensing of Environment*.
- Roy, D. P. (2016a). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote sensing of Environment*, 185, 57–70.
- Roy, D. P.-D. (2016b). A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*. 176, 255–271.
- Roy, D. P. (n.d.). Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sensing Environment*. 199,25–38.
- Saah, D. J. (2019). Collect Earth: An online tool for systematic reference data collection in land cover and use applications. *Environmental Modelling and Software*, 118(May9), 166–171.
- Salomonson, V., & Appel, I. (2004). Estimating fractional snow cover from MODIS using the normalized difference snow index. *Remote Sensing of Environment*, 89 (3), 351–360.
- Shepherd, J. D. (2003). Correcting satellite imagery for the variance of reflectance and illumination with topography. *International Journal of Remote Sensing*. 24(17), 3503–3514.
- Smith, J. A. (1980). The Lambertian assumption and Landsat data. *Photogrammetric Engineering and Remote Sensing*, 46(9), 1183–1189.

- Soenen, S. A. (2005). A modified sun-canopy-sensor topographic correction in forested terrain. *IEEE Transactions on geoscience and remote sensing*, 43(9), 2148–2159.
- Tadono, T. I. (2014). Precise global DEM generation by ALOS PRISM. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2(4), 71.
- Takaku, J. T., Tadono, T., & Tsutsui, K. (2014). Generation of High Resolution Global DSM from ALOS PRISM. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2(4).
- Team, R. (2018). Core. R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Teillet, P. M. (1982). On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, 8(2), 84–106.
- Vermote, E. J. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185, 46–56.
- Xu, H. (2008). A new index for delineating built-up land features in satellite imagery. *International Journal of Remote Sensing*, 29 (14), 4269–4276.
- Zhai, H. Z. (2018). Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144, 235–253.
- Zhu, Z. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote sensing of environment*, 118, 83–94.



Mekong River Commission Secretariat

P. O. Box 6101, 184 Fa Ngoum Road, Unit 18 Ban Sithane Neua,
Sikhottabong District, Vientiane 01000, Lao PDR
Tel: +856 21 263 263. Fax: +856 21 263 264
www.mrcmekong.org

© Mekong River Commission 2023