EO Clinic project:

Ecosystem-Based Management in River Basins in the Philippines

Work Order Report

Support requested by:
Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ)

e-geos Sertit planetek hellas

Reference: EOC0016_WOR_v01
Date: 2021 July 20

TABLE OF CONTENTS

ABOUT THIS DOCUMENT

This publication was prepared in the framework of the EO Clinic (Earth Observation Clinic, see below), in partnership between ESA (European Space Agency), the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and team of service providers contracted by ESA: e-GEOS S.p.A. (Italy) as Prime with support from SERTIT (France) and PLANETEK HELLAS (Greece).

This Work Order Report (WOR) describes the context of the team activities on Ecosystem-Based Management in River Basins in the Philippines, the geoinformation requirements of the activities and finally, the EO products and services delivered by the EO Clinic service providers in support of those activities.

ABOUT THE EO CLINIC

The EO Clinic (Earth Observation Clinic) is an ESA (European Space Agency) initiative to create a rapid-response mechanism for small-scale and exploratory uses of satellite EO information in support of a wide range of International Development projects and activities. The EO Clinic consists of "on-call" technically pre-qualified teams of EO service suppliers and satellite remote sensing experts in ESA member states. These teams are ready to demonstrate the utility of satellite data for the development sector, using their wide range of geospatial data skills and experience with a large variety of satellite data types.

The support teams are ready to meet the short delivery timescales often required by the development sector, targeting a maximum of 3 months from request to solution.

The EO Clinic is also an opportunity to explore more innovative EO products related to developing or improving methodologies for deriving socio-economic and environmental parameters and indicators.

The EO Clinic was launched in March 2019 and is open to support requests by key development banks and agencies during the 3 years project duration.

AUTHORS

The present document was prepared and coordinated by Valeria Donzelli (project manager, e-GEOS) with support from the following contributors: Dimitrios Bliziotis, Vasiliki Avgikou and Dionysis Grigoriadis (Planetek Hellas), Henri Giraud and Robin Faivre (ICube-SERTIT).

ACKNOWLEDGEMENTS

The following colleagues provided valuable inputs, insights and evaluation feedback on the work performed: Engr Mari Trix L. Estomata, (GIS Specialist, GIZ), and Zoltan Bartalis (ESA Coordinator and Technical Officer).

For further information	Please contact Valeria Donzelli, project manager, e-GEOS S.p.A. (valeria.donzelli@e-geos.it) with copy to Zoltan.Bartalis@esa.int if you have questions or comments with respect to content or if you wish to obtain permission for using the material in this report. Visit the ESA EO Clinic: https://eo4society.esa.int/eo_clinic.

1 DEVELOPMENT CONTEXT AND BACKGROUND

Deforestation in the Philippines is one of the major environmental issue. The Philippines' forest cover has declined from 17.8 million hectares or about 60% of the land area in 1934 to about 7.17 million hectares or 24% in 2011 (PFS, 2011). From a position as one of the top ten deforestation countries contributing to global greenhouse gas emissions of 17-20 percent from global forest loss in 2000 (FAO, 2006), the country has since recovered with modest forest cover increase and is now in the list of countries with positive forest growth (FAO, 2010). The socioeconomic and ecological consequences of forestland degradation include widespread poverty, accelerated soil erosion and massive flooding of low-lying areas.

The current work represents a feasibility analysis of large-scale land monitoring services based on automatic processing of optical and SAR satellite time series data. The analysis is oriented to provide valid support for the development of a system for the analysis of the forest ecosystem evolution. The services can help to design an effective monitoring of the forest changes and to set a proper plan of nature conservation measures.

2 PROPOSED WORK LOGIC FOR EO-BASED SOLUTIONS

The proposed products is meant as a support to the GIZ project "Ecosystem-based management and application of ecosystem values in two mentioned river basins in the Philippines (E2RB)", therefore to the Department of Environment and Natural Resources (DENR) and local communities in the Philippines. The aim is to provide more frequent and reliable information about land cover and forest dynamics with respect to the dataset available at national level.

The work includes two inter-related services and is structured in 6 tasks or Work Packages (WP):

- WP1 takes care of all the management activities and lasts for the whole project's duration.
- WP2 takes care of acquisition of data required to realize services. It has been initiated immediately at the start of the project.
- WP3 takes care of resuming the state of art related to Tree cover and related changes, which is described in details in next paragraph 3.1.1.
- WP4 generates Service 1: Long-Term Forest Dynamics, which is described in details in next paragraph 3.1.2.
- WP5 generates Service 2: Forest Loss Rate, which is described in details in next paragraph 3.1.3.
- WP6 provides a proposal on training and capacity building to local partners in using the derived datasets.

3 DELIVERED EO-BASED PRODUCTS AND SERVICES

3.1 Work logic and results

The services delivered intended to provide a comprehensive overview of the recent history and trends concerning land cover changes with a focus on the evolution of the forest towards agricultural areas within the Ilog-Hilabangan River Basin (IHRB) in the Visayas Region and the Tagum-Libuganon River Basin (TLRB) in Mindanao over the 2000 to 2020 period.

3.1.1 State of art

A state-of-the-art analysis was conducted in the initial phase to get an overview of the different on-going methods and developments within this particular field of interest.

Regarding the forestry dynamics mapping topic (Service 1), an initial focus was put on the references provided by the Department of Geographical Sciences of the University of Maryland, and by the Joint Research Center. A specific interest was dedicated to the papers dealing with optical/SAR imagery contributions, as the cloud cover over the Philippines is quite regular, widely limiting the possibilities of optical imagery usage.

In addition, a further analysis was realised on the second topic on forestry loss rate (Service 2).

3.1.1.1 Introduction

Since the 40's, the Food and Agriculture Organization of the United Nations (FAO) raised the need for forestry assessment at a global scale through survey and mapping initiatives.

At the first session of the Conference of FAO in the autumn of 1945, the need for up-to-date information on the forest resources of the world was recognized. In May 1946 the Forestry and Forest Products Division was founded and work was initiated on FAO's first worldwide assessment of forests. The sixth session of the FAO Conference in 1951 recommended that the Organization maintain a permanent capability to provide information on the state of forest resources worldwide on a continuing basis. Since that time, various other regional and global surveys have been conducted every five to ten years. Each has taken a somewhat different form.[1]

Over the years, while mainly based on local field surveys, the FAO global assessment was then progressively supported by the integration of EO data archives.

Since 1990, FAO Global Forest Resources Assessment (FRA) complements the information collected through the country reporting process with global and regional analysis of the world's forest resources using remote sensing. With better access to a growing archive of satellite imagery and availability of new tools to facilitate image processing and interpretation, remote sensing is becoming an important tool for the assessment of status and changes in tree cover and land use. FAO has conducted Remote Sensing Surveys (RSS) as part of the FRA 1990, FRA 2000, FRA 2010, FRA 2015 and now FRA 2020 in close collaboration with FAO Members and other partners. The objectives of the RSS are to build country capacities to use remote sensing for forest monitoring as well as to generate independent, robust and consistent estimates of forest area and its changes over time at global, regional and biome levels.[2]

Through the last decade, thanks to the public availability of Landsat archives and to the generalisation of large-scale processing platforms, new initiatives for forestry dynamics mapping have emerged.

In the scope of Service 1, an initial focus has been put on the references provided by the Department of Geographical Sciences of the University of Maryland, and by the Joint Research Center. A specific interest has

[1] http://www.fao.org/forest-resources-assessment/past-assessments/en/
[2] http://www.fao.org/forest-resources-assessment/remote-sensing/en/

been dedicated to the papers dealing with optical/SAR imagery contributions, as the cloud cover over the Philippines is quite regular, widely limiting the possibilities of optical imagery usage. In addition, a further analysis has been realised on the second topic linked to forestry loss rate (Service 2).

3.1.1.2 EO-based forest dynamics mapping

Global Forest Change (UMD)

Since 2013, the Global Land Analysis and Discovery (GLAD) laboratory at the University of Maryland, in partnership with Global Forest Watch (GFW), proposes the Global Forest Change dataset. It corresponds to an annually updated global-scale forest loss data, derived using Landsat time-series imagery. These data are a relative indicator of spatiotemporal trends in forest loss dynamics globally.

Currently, the Global Forest Change dataset covers the period ranging from 2000 to 2020. In terms of thematic content, the GFC dataset consists in six files presented here below.

- Tree canopy cover for year 2000:

The tree cover in the year 2000 corresponds to canopy closure for all vegetation taller than 5m in height. Values are encoded as a percentage per output grid cell, in the range 0–100.

- Global forest cover gain 2000–2012:

This layer corresponds to new forest areas appeared between 2000 and 2012. Unlike the forest cover loss data, there is no indication of the year. The result is simply encoded as either “1” for areas changing from non-forest to forest, or “0” otherwise.

- Year of gross forest cover loss event:

The content of this layer is defined as a stand-replacement disturbance, or a change from a forest to non-forest state, during the period 2000–2020. Practically, all the clear-cuts detected over a given year between 2000 and 2019 have been assigned a specific class. For sake of simplicity, the pixel value of the classes corresponds to the year when the clear-cuts have been detected (e.g. clear-cuts detected in 2013 will result with a pixel value equal to 13). The “0” values correspond to areas where no clear-cut has been detected over the period 2000-2020.

- Data mask:

The data mask contains three values representing areas of no data (0), mapped land surface (1), and permanent water bodies (2).

- Circa year 2000 Landsat 7 cloud-free image composite:

A reference multispectral imagery from the first available year, typically 2000. If no cloud-free observations were available for year 2000, imagery was taken from the closest year with cloud-free data, within the range 1999–2012.

- Circa year 2020 Landsat cloud-free image composite:

A reference multispectral imagery from the last available year, typically 2020. If no cloud-free observations were available for year 2020, imagery was taken from the closest year with cloud-free data.

The Global Forest Change data are processed on the Google Earth Engine platform. Results can be visualized through a dedicated portal[1] and are available to download in 10x10 degree tiles[2]. The GFD dataset is provided as a complete set of granules covering the range 180W–180E and 80N–60S (with meaningful data over land

---

[1] https://glad.earthengine.app/view/global-forest-change#dl=1;old=off;bl=off;lon=20;lat=10;zoom=3;
[2] https://storage.googleapis.com/earthenginepartners-hansen/GFC-2020-v1.8/download.html

surfaces). All files have a spatial resolution of 1 arc-second per pixel, which corresponds approximately to 30 meters at the equator. Figure 1 provides an overview of the GFC tiling and extent.

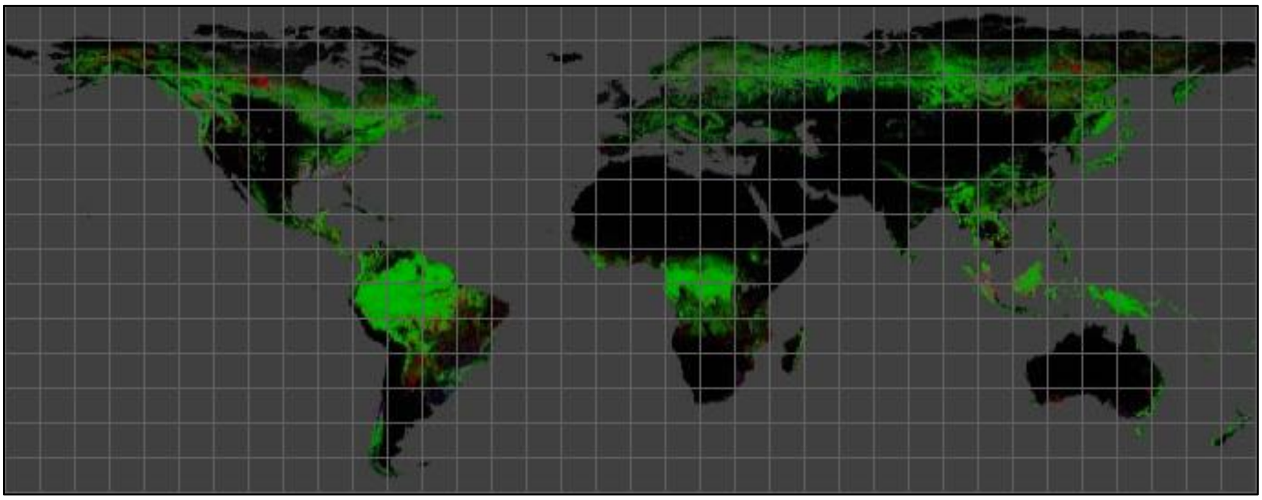


Figure 1: Overview of the GFC tiles and overall extent

The current version 1.8 is an update of gross forest cover loss that includes new 2020 loss-year and multispectral imagery layers. According to the authors, the overall method has been modified in numerous ways, and leads to a different and improved detection of global forest loss. However, the years preceding 2011 have not yet been reprocessed in this manner, resulting to some inconsistencies. This new version has to be seen as part of a transition to a future version 2.0 that will be more consistent over the entire 2000-onward period.

Tropical Moist Forest (JRC)

In 2021, the European Commission's Joint Research Centre (JRC) released a new dataset on forest cover change in tropical moist forests (TMF) using 39 years of Landsat time series[1]. The TMF dataset covers the period 1990-2020 and consists in two map products (transition and annual changes), with associated metrics characterizing the timing (year of deforestation, year of degradation, duration), the number of annual disruption observations and the intensity (total number of disruption observations over the full observation period) of the disturbances. The Tropical Moist Forest data are also processed on the Google Earth Engine platform.

Regarding the exact thematic content, the authors provide the following map product description[2].

- Transition map:

The transition map shows the spatial distribution of the moist forest at the end of the year 2020. It depicts the sequential dynamics of changes by providing transition stages from the first year of the monitoring period to the end of the year 2020. Two maps are proposed and described for the transition map: (i) a first version entitled "Transition Map - Main Classes" with the main transition classes, and (ii) a detailed version entitled "Transition Map-Subtypes" with sub-classes(period of disturbance, age of regrowth, several types of forest, several types of degradation and deforestation, change types within the mangroves and tree plantations).

- Undisturbed and degraded tropical moist forest:

The Undisturbed and degraded tropical moist forest is a simplification of the Transition Map - Main Classes and shows the spatial distribution of undisturbed and degraded tropical moist forests remaining at the end of the year 2020. Forests include mangroves and bamboo-dominated forest types

- Annual change collection:

[1] https://doi.org/10.1126/sciadv.abe1603
[2] https://forobs.jrc.ec.europa.eu/TMF/download/TMF_DataUsersGuide.pdf

The annual change collection depicts the extent and status of the TMF (degraded, deforested, regrowing) for each year between 1990 and 2020). The timeline allows seeing how the TMF is changing over the past 3 decades.

Some metrics of timing and intensity of the disturbances are provided as additional layers.

- Degradation year:

The degradation year is the year when the forest cover has been degraded for the first time. It concerns all the degraded forest classes of the transition map including the mangroves and the recent degradation (2020).

- Deforestation year:

The deforestation year is the year when the forest cover has been deforested for the first time (followed or not by a regrowth).It concerns all the deforested classes of the transition map including the mangroves that have been deforested, the conversion into tree plantation, the conversion into water and the recent deforestation (2018-2020).

- Duration (only available on GoogleEarth Engine):

The duration corresponds to the number of days between the first and last disruptions detected for all the areas classified as TMF change in the transition map.

- Number of annual disruption observations (only available on GoogleEarth Engine):

This dataset provides the number of disruption observations on an annual basis. A disruption observation is defined as an absence of tree foliage cover within a Landsat pixel for a single-date observation.

- Intensity (only available on GoogleEarth Engine):

The intensity of the disturbance documents the total number of disruptions detected over the full observation period (from the first year of the monitoring period to 2020) for all the areas classified as TMF change in the transition map.

The Tropical Moist Forest results can be visualized through a dedicated portal[1], and are also available to download in 10°x10° tiles covering tropical areas[2]. All files have a spatial resolution of 1 arc-second per pixel, which corresponds approximately to 30 meters at the equator. Figure 2 provides an overview of the TMF tiling and extent.

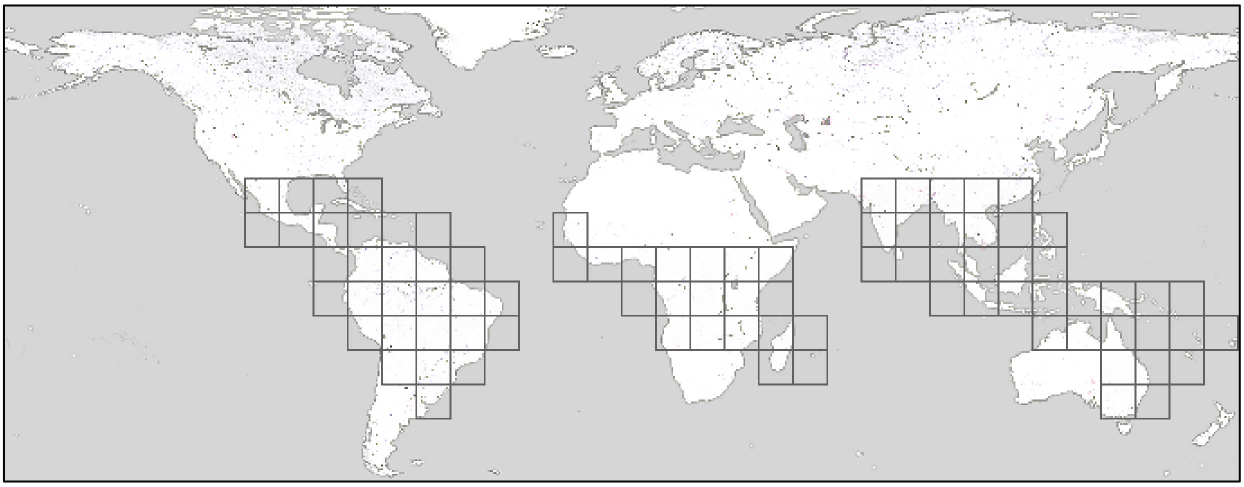


Figure 2: Overview of the TMF tiles and overall extent

[1] https://forobs.jrc.ec.europa.eu/TMF/
[2] https://forobs.jrc.ec.europa.eu/TMF/download/

For what concern Service 1, the annual change collection could be of high interest for the generation of annual forest cover maps. This dataset corresponds to a collection of 31 maps depicting for each year between 1990 and 2020 the spatial extent of undisturbed forest and changes (deforestation, degradation and regrowth). Each of the annual changes map is an individual file containing the following 6 classes:

1. Undisturbed tropical moist forest,
2. Degraded tropical moist forest,
3. Deforested land,
4. Forest regrowth,
5. Permanent and seasonal water,
6. Other land cover.

Deforestation refers to a change in land cover (from forest to non-forested land) when degradation refers to a temporary disturbance in a forest remaining forested such as selective logging, fires and unusual weather events (hurricanes, droughts, blowdown).

3.1.1.3 On the use of optical/SAR imagery for forestry mapping

In tropical regions, the use of optical sensors is limited by high cloud coverage throughout the year. As an alternative, Synthetic Aperture Radar (SAR) products can be used, alone or in combination with optical images, to monitor tropical areas. Numerous studies on the combined use of optical and SAR data have been conducted over the years. An exhaustive inventory would be difficult and out of interest; however, we propose to focus on two recent studies performed over tropical areas similar to the present one.

The first study deals with the test of several combinations of optical and SAR data to identify the four dominant vegetation types that are prevalent in the Brazilian Cerrado[1]. The Cerrado biome is considered as being among the most extensive and diverse ecosystems in the tropics and is a hotspot in the context of biodiversity. It is also one of the most threatened ecosystems in South America, with over 40% of the biome converted to agriculture and the remainder highly fragmented. The Cerrado biome is the second largest complex vegetation present in Brazil and occupies about 200 million hectares. In this way, Brazil needs to improve the monitoring system of deforestation and land use change.

The aim of this study is evaluate the use of optical and radar remote sensing for mapping the different types of vegetation in the transitional area between the Cerrado and Amazon biomes. To do so, the authors extracted features from both sources of data such as intensity, grey level co-occurrence matrix, coherence, and polarimetric decompositions using Sentinel-2, Sentinel-1, ALOS-PALSAR 2 dual/full polarimetric, and TanDEM-X images during the dry and rainy season of 2017. In order to normalize the analysis of these features and to reduce their temporal dimensionality, principal component analysis was applied for each seasonal temporal stack, and subsequently applied the Random Forest algorithm to evaluate the classification of vegetation types.

A supervised classification was first performed on two different Sentinel-2 and ALOS2 combinations, to produce a first forest/non-forest mask. The resulting thematic accuracy was excellent for both of them (nearly 99%), and in each case, the highest contribution was provided by the near and shortwave infrared spectral bands from Sentinel-2 data. In that case, SAR data can be seen as a support to the information provided by optical data.

A forest type mapping was then conducted in the areas masked as forests in the previous step. To analyse the synergy of optical and radar data for mapping Cerrado vegetation types, all possible combinations between optical and radar sensors were tested in two different scenarios, dry season, dry and rainy seasons. In addition, the authors used the sensors separately and analysed the SAR classifications.

[1] https://doi.org/10.3390/rs11101161

During the dry season, the overall accuracy ranged from 48 to 83%, and during the dry and rainy seasons, it ranged from 41 up to 82%. The classification using Sentinel-2 images during the dry season resulted in the highest overall accuracy values, followed by the classification that used images from all sensors during the dry and rainy season. Optical images during the dry season were sufficient to map the different types of vegetation in our study area. When considering both dry and rainy seasons, the combination of optical and radar sensor data usually improved the vegetation classification.

The second study deals with the use of SAR and optical time series for forest disturbance mapping (meaning both deforestation and forest degradation) in tropical areas[1] where frequent cloud cover and fast regrowth often prevent forest disturbance monitoring with optical data only. The authors intended to overcome these limitations by combining dense time series of optical (Sentinel-2 and Landsat 8) and SAR data (Sentinel-1) over test sites located in Peru and Gabon, to increase forest disturbance detection accuracies in the humid tropics.

SAR and optical data are highly complementary, since they detect different disturbed forest areas. Therefore, the authors aim at demonstrating that higher accuracy values can be obtained by merging detections from both sensor types. Their approach begins also with the generation of a forest/non-forest mask, which serves as the starting point for the forest disturbance mapping. Separate forest disturbance maps are then calculated from the SAR and optical time series. The final forest disturbance maps combine the forest disturbance results from SAR and optical time series by a union process. The forest masks and all the forest disturbance maps (SAR, optical, combined) are validated with a set of sample plots that were visually interpreted from VHR and HR imagery.

The forest/non-forest mask is first produced from several classifications performed on optical data for which the acquisition date spans from March 2015 to March 2016. A final binary mask is obtained by applying a weighted majority approach to merge the initial classification. Their resulting thematic accuracy is respectively of 93% and 98% for Peru and Gabon. As for disturbance, the accuracies of the individual maps from optical and SAR time series are compared with the accuracies of the combined map.

The authors then evaluate the detection accuracies by disturbance patch size and by an area-based sampling approach. The results show that the individual optical and SAR-based forest disturbance detections are highly complementary, and their combination improves all accuracy measures. The overall accuracies increase by about 3% in both areas, producer accuracies of the disturbed forest class increase by up to 25% in Peru when compared to only using one sensor type. The assessment by disturbance patch size shows that the number of detections of very small disturbances (<0.2 ha) can almost be doubled by using both datasets.

The authors acknowledge that further improvements should focus on developing more advanced methods to derive forest disturbances from SAR and optical data. In addition, further studies are needed in different parts of the tropics and under different forest conditions to better understand the potential and limitations of sensor type combinations for tropical forest disturbance detections.

3.1.1.4 Methodologies on forestry loss rate

The forest ecosystem in the Philippines, like in many countries worldwide, is severly affected by human activities. This is particularly preoccupying as forests are essential, not only for the environment, but also for the social and human well-being. Among the problems that forest have to face, deforestation linked to agriculture is probably the main threat, due to uncontrolled and excessive exploitation caused by increasing population and unsustainable management of natural resources.

---

[1] https://doi.org/10.3390/rs12040727

Besides the provision of several EO-based Land cover and Forest cover maps over both areas of interest, from archive (up to 2000) and recent satellite imagery to get a reliable and objective overview of landscape evolution within the IHRB and TLRB areas, another key objective consists in identifying the most critical zones within both river basins, in terms of deforestation and forestry loss rate. This component, integrated within Service 2, is indeed crucial to support the national and local stakeholders to have a better knowledge of the situation and hence prioritize their actions and resources.

The process is widely based on the paper by Nowosad and Stepinski (2019)[1] who developed an innovative methodology to simulate long-term evolution of landscapes. The difference with other existing approaches is that it simulates the probability distribution function of long-term trajectories for a single landscape based on observations of short-term transitions (1992-2015), for a large number of different landscapes. The most likely evolutionary scenario for each landscape is then calculated using the Monte Carlo method.

This methodology is particularly adapted for specific types of land changes such as deforestation, desertification or urbanization. Moreover, the major advantage is that the model is empirical, making it unnecessary to account for all the different individual processes responsible for the land changes. On the other hand, the method also has some shortcomings, the main one being the fact that it is based on the assumption that the land change processes remain stationary during the whole transition period (1992-2015). The implication of this assumption is that the model shows how the landscape would evolve if processes driving the change and their intensities are stable through time. Another second assumption is that the model is built on worldwide statistics, so that the probabilities are calculated from the largest possible statistics of transitions.

The results show that the fastest forest-to-agriculture transit (FAT) scenarios occur through the sequence of highly aggregated forest/agriculture mosaics with a decreasing share of the forest, and that once the forest share drops below 50% the remainder of the transit is rapid. Therefore, this suggests that conservation policies could focus on preserving the forest landscapes before its share drops below 50% at a mesoscale (~100 km^2).

3.1.1.5 Perspectives

In a near future, initiatives such as Global Forest Change and Tropical Moist Forest will take benefit from the next Landsat 9 launch in September 2021. This new satellite will be placed in an orbit that it is eight days out of phase with Landsat 8 to increase temporal coverage of observations. Landsat 9 will continue the Landsat series legacy with the existing 44-year data record.

Since 2014 and 2015, new constellation of Sentinel-1 (SAR) and Sentinel-2 (optical) satellites in the frame of the Copernicus program. Sentinel-2 satellites ensure a kind of synergy with Landsat sensors by offering similar spectral characteristics, and even offer higher both temporal and spatial resolutions; and Sentinel-1 allows to better taking benefit from the optical/SAR capabilities.

Besides this, the continuous development of both High-Performance Computing (HPC) and cloud computing platforms and their increasing capacities offered by Space Agencies, Universities, DIAS[2], and private companies (e.g. Google or Amazon, among others) paves the way to new forest mapping initiatives.

Moreover, the increasing research on new machine learning and deep learning algorithms will allow to better exploit the richness of EO data. Due to the development of new technologies, new ancillary data might be also of interest for improving the forest dynamics mapping.

In conclusion, whatever the sensors, computing, and algorithms capabilities are, the need for reliable ground truth for both training and validation is of paramount importance.

[1] J. Nowosad, T. F. Stepinski (2019) Stochastic, Empirically Informed Model of Landscape Dynamics and Its Application to Deforestation Scenarios. Geophysical Research Letters 46 (23): 13845-13852.
[2] https://www.copernicus.eu/en/access-data/dias

3.1.2 Service 1: Long-Term Forest Dynamics

3.1.2.1 Specifications

The Service 1 includes a set of tasks and products as follows:

1. **Land cover maps** for the following three epochs:
    a. 2020 land cover maps based on S2+S1
    b. 2015(16) land cover maps based on S2+S1
    c. 2000 land cover map based on Landsat
2. **Forest cover maps** for the following time ranges:
    a. 2001-2015 forest cover maps
    b. 2016-2020 forest cover maps based on S2+S1.

For each of these products in the following are reported the necessary input data, the methodology and the resulting outputs.

Land cover maps

Three land cover maps have been realized for the epochs: 2020, 2015(16), 2000 both for the Ilog-Hilabangan River Basin (IHRB) and the Tagum-Libuganon River Basin (TLRB). Furthermore, two land cover change maps (2000-2016; 2016-2020 for both the basins) have been realized with particular focus on the changes from forest versus artificial and agricultural classes.

1. Input data
    – Shapefile containing the exact basin contours provided by GIZ
    – Landsat 5/7 time series for epochs 2000 (± 1 year for cloud filling)
    – Sentinel-2 L1C time series for epoch 2016 from the Copernicus Open Access Hub (www.scihub.copernicus.eu)
    – Sentinel-2 L2A time series for epoch 2020 from the Sentinel-2 Global Mosaic service (https://s2gm.sentinel-hub.com/)
    – Sentinel-1 data time series for epochs 2016 and 2020

The land cover maps for earlier epoch 2000 are based on Landasat5 and Landsat7 merged dataset. The 2000 map was based only on the optical component due to the unavailability of SAR data in the selected epoch. Moreover, due to the high cloud coverage in the Landsat images series of 2000, also one image from 2001 has been used as input for the classification (see **Errore. L'origine riferimento non è stata trovata.**).

The LC maps for the recent years 2015(16) and 2020 are derived by Sentinel-2/Sentinel-1. In particular, for the year 2020, the Sentinel-2 Global Mosaic service (https://s2gm.sentinel-hub.com/) has been used to collect monthly mosaic compositions of S/2 images. The Sentinel-2 Global Mosaic (S2GM) service is a component of the Copernicus Global Land Service providing composites from time-series of Sentinel-2 surface reflectance observations. S2GM comprises best representative pixels in three spatial resolutions and from different compositing periods ranging from one day to one year on a global scale.

For the year 2015(16), composites of the Sentinel-2 images with the less cloud coverage have been applied, as the Sentinel-2 Global Mosaic service is not available for this epoch.

In the following table, the data available for each mapping epoch is shown with the corresponding cloud coverage and the final data selection.

		Acquisition Date	% Valid Pixels over AoI
1999	L7	10/8/1999	36.5
	L7	11/25/1999	17.3
2000	L5	2/21/2000	57.5
	L7	4/17/2000	55.9
	L5	4/25/2000	45.3
	L5	7/14/2000	65.7
	L7	8/7/2000	81.7
	L7	9/24/2000	65.1
2001	L7	4/4/2001	55.3
	L7	5/22/2001	95.0
	L7	9/11/2001	73.5
	L7	11/14/2001	57.5

Table 1: List of available data and identification of the selected input data for TLRB 2000 LC production

Sentinel 2							Sentinel1	
Month	DAY	Tiles Available		Tiles Corrupted		% ValidPixels over AoI	Ascending	Descending
		NYJ	NZJ	NYJ	NZJ			
1								X
2								X
3	3	X	X			71.1		X
4	2	X	X			86.9		X
5								X
6	11	X	X			53.3		X
7	21		X					X
8	10	x	x			1.3		X
9	19	X	X	X	X			X
10	9	X	X			77.8		X
11	28	X	X			66.1		X
12	8		X					X

Table 2. List of available data and identification of the selected input data for TLRB 2016 LC production

Sentinel2		Sentinel1	
Composite Data	% Valid Pixels	Ascending	Descending
2020_01	87.2	x	x
2020_02	72.7	X	X
2020_03	90	x	x
2020_04	97.9	X	X
2020_05	94.5	x	x
2020_06	92.9	X	X
2020_07	92.4	x	x
2020_08	94.6	X	X
2020_09	89.3	x	x
2020_10	95.4	X	X
2020_11	71	x	x
2020_12	92.7	X	X

Table 3: List of available data and identification of the selected input data for TLRB 2020 LC production

		Acquisition date	% Valid Pixels over AoI
1999	L7	10/6/1999	29.1
2000	L5	4/23/2000	66.3
	L7	5/1/2000	69.3
	L5	5/25/2000	49.6
	L5	6/26/2000	53.1
	L7	9/22/2000	44.8
	L7	11/25/2000	39.8
2001	L7	5/4/2001	66.3
	L7	8/8/2001	26.1
	L7	11/12/2001	94.9

Table 4: List of available data and identification of the selected input data for IHRB 2000 LC production

Sentinel 2												Sentinel1	
	Month	DAY	Tiles Available				Tiles Corrupted				% Valid Pixels over AoI	Asc.	Desc.
			PVL	PVM	PWL	PWM	PVL	PVM	PWL	PWM			
2015	11												X
	12												X
2016	1	16	x	x	x	x				x			X
	2												X
	3												X
	4	15	x	x	x	x					62.7		X
	5												X
	6												X
	7												X
	8	3	x	x	x	x					High Cloud Cover		X
	9												X
	10	22	x	x	x	x					63.6		X
	11	11	x	x	x	x					96.1		
	12												X

Table 5: List of available data and identification of the selected input data for IHRB 2016 LC production

Sentinel2		Sentinel1	
Composite Data	% Valid Pixels	Ascending	Descending
2020_01	95.2	X	X
2020_02	96.1	X	X
2020_03	98.7	X	X
2020_04	95	X	X
2020_05	92.1	X	X
2020_06	90	X	X
2020_07	60	X	X
2020_08	14.3	X	X
2020_09	85.1	X	X
2020_10	80.1	X	X
2020_11	94.4	X	X
2020_12	90.3	X	X

Table 6. List of available data and identification of the selected input data for IHRB 2020 LC production.

2. <u>Methodology</u>:
   - <u>Pre-processing</u>

     A set of pre-processing operations, aimed to correct radiometric and geometric distortions of data and to harmonize the input multisensory and multiscale data was carried out with the aim to create a single time series image as input for each classification mapping epoch. SAR and optical data were integrated for the generation of a multilayer data stack used as input for the classification phase based on AI methodologies. All images were combined into one multistack dataset for each production year that contains time sorted optical and SAR images.

Since the different nature of the input data used for each epoch, the pre-processing is described in details for each one.

The general workflow followed for the pre-processing of **optical data** was:

1. Perform the atmospheric correction of each optical image in the time series.

For Landsat data, surface reflectance images have been generated using the LEDAPS algorithm, developed by NASA, while for Sentinel-2 data, the Sen2Cor plugin available in SNAP was adopted for the LC 2016. This phase was not carried out for the LC 2020 since the Sentinel-2 Global Mosaic service provides data atmospherically corrected and mosaiced over the requested AoI and time windows. For the present work, the time window was settled on a monthly scale. For this specific epoch therefore the pre-processing of optical data started from the cloud masking and interpolation.

2. Cloud masking and filling.

It was an important issue and also a relevant component to manage for both the selected AoIs since the generally high cloud presence in the Philippines. A pixel-based interpolation process was implemented using the time series data to fill no-data cloud covered values using a simple linear interpolation since it is able to balances the final obtained accuracy with the computational time.

3. Computation of vegetation indices

Some vegetation indices were added to multistack input data to highlight different behaviours between pixels and separate different thematic classes allowing a more effective extraction of features. Some vegetation indices were calculated and added to the multistack: NDVI, NDWI, MSAVI2 were estimated from Sentinel2 and Landasat7 data, while from the Landsat 5, due to the absence of a band into the SWIR region, only NDVI, and MSAVI2 were calculated.

With specific reference to the **SAR data pre-processing**, starting from the GRDH data from both ascending and descending orbits, it was done performing the following steps:

- ellipsoid correction (to project the image on the ground on UTM projection)
- conversion in sigma0
- cut over the area of interest
- suppress speckle effects preserving relevant features (edges, point target, artificial areas, water, etc.).
- create the monthly SAR mosaic over the AoI
- co-registrate it with the optical data and resampling to the latter's resolution.

All the SAR pre-elaborations were performed using the software SNAP. Notice that when both ascending and descending SAR data were available, their average value was used to perform the LC classification to reduce the size of the multistack used for the classification.

– Training

Different sets of training samples have been collected for each epoch applying a stratification scheme based on the land cover classes distribution of the Global Land Cover ESA CCI. For each basin about 4% of the basin area has been used as training for the AI classifiers for each epoch. The training polygons have been delineated and classified on screen exploiting the available reference data (e.g., Google Earth, OpenStreetMap, Spatial Database of Planted Trees (SDPT) 2015 dataset2, Global Forest Change (GFC) 2000-2019 dataset3). In some cases, a preliminary automatic image segmentation of the best optical image available for each epoch over the basin has been useful to help the polygons selection on screen.

LC Code	LC Class	Area Class [km^2]
1	Artificial	2.9
2	Forest	51.5
3	Tree crops	59.3
4	Permanently herbaceous	3.4
5	Periodically herbaceous	10.0
6	Water	2.1
Total Training [km^2]		129.16
Area AoI [km^2]		3205.25
Total Area Training / Area AoI		4.0%

Table 7. Training sample composition used for TLRB 2000 LC production.

LC Code	LC Class	Area Class [km^2]
1	Artificial	2.9
2	Forest	51.5
3	Tree crops	59.2
4	Permanently herbaceous	3.4
5	Periodically herbaceous	3.4
6	Water	2.1
Total Training [km^2]		122.44
Area AoI [km^2]		3205.25
Total Area Training / Area AoI		3.8%

Table 8. Training sample composition used for TLRB 2016 LC production.

LC Code	LC Class	Area Class [km^2]
1	Artificial	2.2
2	Forest	51.5
3	Tree crops	59.2
4	Permanently herbaceous	3.4
5	Periodically herbaceous	10.8
6	Water	2.1
Total Training [km^2]		129.19
Area AoI [km^2]		3205.25
Total Area Training / Area AoI		4.0%

Table 9. Training sample composition used for TLRB 2020 LC production.

LC Code	LC Class	Area Class [km^2]
1	Artificial	0.4
2	Forest	29.9
4	Permanently herbaceous	6.0
5	Periodically herbaceous	43.4
6	Water	0.6
Total Training [km^2]		80.4
Area AoI [km^2]		2173.2
Total Area Training / Area AoI		3.7%

Table 10. Training sample composition used for IHRB 2000 LC production.

LC Code	LC Class	Area Class [km^2]
1	Artificial	0.3
2	Forest	29.3
4	Permanently herbaceous	3.5
5	Periodically herbaceous	46.3
6	Water	0.6
Total Training [km^2]		80.1
Area AoI [km^2]		2173.2
Total Area Training / Area AoI		3.7%

Table 11. Training sample composition used for IHRB 2016 LC production.

LC Code	LC Class	Area Class [km^2]
1	Artificial	0.4
2	Forest	29.3
4	Permanently herbaceous	6.1
5	Periodically herbaceous	43.3
6	Water	0.6
Total Training [km^2]		79.7
Area AoI [km^2]		2173.2
Total Area Training / Area AoI		3.7%

Table 12. Training sample composition used for IHRB 2020 LC production.

– AI algorithms

Among the AI algorithms available, which allows to reach high accuracy in classification of time series data the *Light Gradient Boosting Method* (LGBM https://lightgbm.readthedocs.io/en/latest/) has been applied. The idea behind this standard machine learning algorithm is very simple: it combines the predictions of multiple decision trees by adding them together. Moreover, the LightGBM package, offers an intuitive, fast, distributed, and a high-performance gradient boosting framework based on decision tree algorithms. For tuning the parameters of the classifier, various approaches such as a grid search or a random search can be pursued and evaluated on the test dataset. To perform this multiclass classification was settled as metric the *'multi logloss'* and a number of nodes ("num_leaves") equal to 31 and decision trees equals to 1,000. Only the 15% of the training sample was used for the validation while the remaining 85 % was used for the model's training.

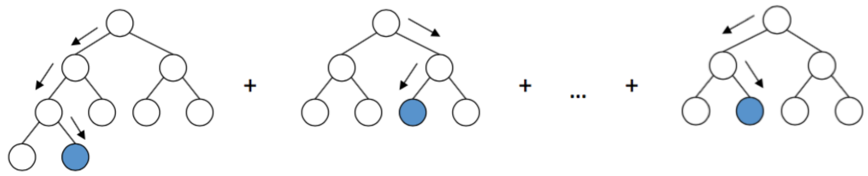


Figure 3. Schematic of the decision trees in the LightGBM framework. (Source: http://arogozhnikov.github.io/2016/06/24/gradient_boosting_explained.html)

3. <u>Output data:</u>
   - List of the output maps for TLRB and IHRB basins
     - Land cover map 2020
     - Land cover map 2016
     - Land cover map 2000

       The land cover maps were combined to extract the changes from forest areas towards artificial and agriculture:
     - Land cover change map 2000-2016 (forest changes)
     - Land cover change map 2016-2020 (forest changes)
   - Format
     - Raster: TIFF (Tagged Image File Format)
     - Vector: ESRI Shapefile

     Both are provided in the Coordinate Reference System: WGS 84 / UTM zone 51N (EPSG: 32651)
   - Minimum Mapping Unit (MMU)
     - Land cover maps: 0.09 ha MMU (for LC 2000) / 0.10 ha MMU (for LC 2016 and LC 2020)
     - Land cover changes: 1 ha MMU

   The LC maps for the recent years (2016 and 2020) derived by Sentinel data are characterized by a spatial resolution higher than the LC maps derived by Landsat datasets. In order to make the maps spatially coherent, the 2000 outputs were resampled from the native Landsat resolution of 30 meters to the Sentinel resolution. As a result, 10 meters LC maps were retrieved for all the three epochs considered.

   Since the size of one Landsat pixel (30 m x 30 m = 900 sq m = 0.09 ha) is near the minimum mapping unit (0.1 ha) no post-processing was carried out to eliminate isolated pixels obtained in the LC map. Conversely, for the S2-based LC maps, since the size of one pixel (10 m x 10 m = 0.01 ha) is 10 times smaller than the minimum mapping unit, a post-processing was carried out by identifying the contiguous pixel groups smaller than the MMU and replacing them with the biggest contiguous LC class.

4. <u>Nomenclature</u>

   The following list of classes, described in Table 13, are included into the land cover maps. As shown in the table, the nomenclature is slightly different in the IHRB and TLRB basins.

LC Code	Color code (RGB)	Description	TLRB 2000; 2016; 2020	IHRB 2000; 2016; 2020
1	225,225,225	Artificial (settlements)	X	X
2	56,168,0	Forest	X	X
3	255,170,0	Tree crops (coconut palm, banana, other permanent crops etc.)	X	
4	163,255,115	Permanent herbaceous (grasslands)	X	X
5	255,255,0	Periodically herbaceous (arable land, annual crops)	X	X
6	0,197,255	Water	X	X

Table 13: LC nomenclature

For TLRB basin the separation between forest and tree crops was implemented not only in the 2016 and 2020 classifications (as planned in the proposal phase), but even in the 2000 mapping based on Landsat data. The reason is in the landscape pattern of TLRB basin, characterized by large tree crops fields having a size compatible with the scale of satellite input data.

On the other hand, in IHRB basin the tree crops are marginal and the fields are small and fragmented with low tree cover density. In this context, the application of Sentinel and Landsat data to separate tree crops becomes difficult and subjected to misclassification with others classes (e.g., forest, perennial herbaceous), as demonstrated by some testing activity performed in this area. For this reason, the classifications for the three epochs in IHRB includes the same nomenclature of TLRB except for the tree crops class.

5. Thematic Accuracy: overall thematic accuracy ≥ 80%

File	Naming	Description
Land cover map raster layers	Land_Cover_TLRB_20xx.tif Land_Cover_IHRB_20xx.tif	• GeoTIFF file format • Unsigned integer pixel type • 8-bit pixel depth • UTM51N coordinate system • GSD 10m
Land cover map change layers	Land_Cover_change_TLRB_20xx.tiff Land_Cover_change_IHRB_20xx.tiff	
Forest map layer files	Land_Cover.lyr Land_Cover_change.lyr	ArcGis layer file containing the forest cover map symbology

Table 14: Description of the land cover maps (2000, 2015(16)-2020) deliverables

Forest cover maps

This service is aimed to demonstrate the feasibility of a frequent mapping of the forest cover based on optical and/or radar data. For the time range 2001-2015, annual forest cover maps have been produced based on available forest product at global scale (i.e. Tropical Moist Forest). For the period 2016-2020, the feasibility of a «twice per year» forest cover has been tested. After the initial evaluation of the available input dataset (detailed in the following paragraphs) annual forest cover maps have been realised based on S2+S1 data for the years 2016-2020. Furthermore, the classification of an entire times series over the year allows to follow the phenology of the crops over a longer observation window getting more details useful to improve the final accuracy of the product.

2001-2015 annual forest cover maps

Initially, a combination of the various datasets mentioned previously (Global Forest Change 2000-2020, Tropical Moist Forest 1990-2020) was intended to provide the annual forest cover maps over both AOIs, over the period 2001-2015. The content of the layers contained in these datasets is different; therefore, a particular attention was given when trying to mix these layers. It appeared first that TMF dataset provides annual changes for both gain and loss aspects, while the gain for GFC is integrated over time.

Moreover, the current version of the GFC dataset is 1.8, and is in transition toward a major 2.0 release. This means, among several aspects, that the current dataset do not provide any consistency over the whole 2000-2020 period, due to processing discrepancies for data acquired before 2011 and those from 2011 onward.

Therefore, only the Tropical Moist Forest annual change collection was considered. In addition, a conservative approach was followed by considering only deforestation and regrowth, not degradation, to avoid commission errors that may be encountered when using the GFC dataset. The 2000 land cover map is also used to make sure that for 2001:

- deforested areas are not outside the 2000 land cover forest class,
- new forest areas are outside the 2000 land cover forest class.

To produce each annual forest cover maps, the process is sequential, starting from the 2000 land cover map and adding for each particular year the deforestation/regrowth information resulting from the information content provided by the appropriate classes of the TMF dataset. Before to update the year $n+1$, the year n was first processed to remove deforested areas (value set to “0”) and to convert the regrowth areas as forest (value set to “1”). Table 15 provides some details about the coded values used for each annual forest cover map.

Value	Color code (RGB)	Description
NoData	N/A	Initial cloud mask from the 2000 land cover map
0	255,255,255	Non-forest areas
1	110,170,0	Undisturbed forest
2	255,0,0	Deforested land
3	0,0,255	Forest regrowth

Table 15: Description of the annual forest cover maps (2001-2015) values

Both geometric and thematic consistency between the 2000 land cover map and all the annual forest cover maps was ensured. Given that the TMF is derived from Landsat time series the spatial resolution is around 30 meters. Therefore, the size of one pixel (30m x 30m = 900 sq.m^{-1} = 0.09 ha) is near the minimum mapping unit (0.1 ha) and hence no post-processing was carried out to eliminate isolated pixels. Table 16 describes the technical details for annual forest cover maps.

File	Naming	Description

Forest map raster layers	Forest_Cover_TLRB_20xx.tif Forest_Cover_IHRB_20xx.tif	• GeoTIFF file format • Unsigned integer pixel type • 8-bit pixel depth • UTM51N coordinate system • GSD 30m
Forest map layer files	Forest_Cover_TLRB_20xx.lyr Forest_Cover_IHRB_20xx.lyr	ArcGis layer file containing the forest cover map symbology

Table 16: Description of the annual forest cover maps (2001-2015) deliverables

2016-2020 forest cover maps based on S-2 +S1

Input data:

- Shapefile containing the exact basin contours provided by GIZ
- Sentinel-2 L2A time series retrieved from the Sentinel-2 Global Mosaic service (https://s2gm.sentinel-hub.com/) for epochs 2017, 2018, 2019
- Sentinel-1 data time series for epochs 2017, 2018, 2019
- LC maps retrieved for the epoch 2016 and 2020.

For the 2016 and 2020 epochs, the LC maps produced in the first phase of the project have been used to derive the Forest map products for the same corresponding epochs for both the basins. For this reason, for these epochs a recoding of the LC classes, available in the LC maps, has been performed in order to obtain the output Forest maps: the classes "Artificial", "Tree crops" (in LC where the class is applicable), "Permanent herbaceous", "Periodically herbaceous", "Water" are assigned to "No-Forest" while the LC class "Forest" maintains the same assignment.

The following tables with the the data available for each mapping epoch show how the cloud coverage issue reduced the final data selection to a limited number of scenes.

Sentinel2			Sentinel1	
Composite Data	% Valid Pixels	% Valid Pixels	Ascending	Descending
2017_01	0			X
2017_02	0			X
2017_03	4.5			X
2017_04	24.9			X
2017_05	82.5	86		X
2017_06	0.8			X
2017_07	21.4			X
2017_08	96.5	96.5	X	X
2017_09	0		X	X
2017_10	45.4		X	X
2017_11	91.9	96.7	X	X
2017_12	85.5		X	X

Table 17. List of available data and identification of the selected input data for TLRB 2017 Forest cover map production.

Sentinel2		Sentinel1	
Composite Data	% Valid Pixels	Ascending	Descending
2018_01	91.5		X
2018_02	98.6	X	X
2018_03	80.2	X	X
2018_04	96.9	X	X
2018_05	94.7	X	X
2018_06	30.5	X	X
2018_07	95.8		X
2018_08	94.6	X	X
2018_09	97.6	X	X
2018_10	98.9	X	X
2018_11	89.4	X	X
2018_12	94.7	X	X

Table 18. List of available data and identification of the selected input data for TLRB 2018 Forest cover map production.

Sentinel2		Sentinel1	
Composite Data	% Valid Pixels	Ascending	Descending
2019_01	16.6	X	X
2019_02	90.3	X	X
2019_03	97.1	X	X
2019_04	96.8		X
2019_05	91.3		X
2019_06	96.4	X	X
2019_07	92.5	X	X
2019_08	99.1	X	X
2019_09	95.4	X	X
2019_10	98.7	X	
2019_11	97	X	X
2019_12	95.1	X	

Table 19. List of available data and identification of the selected input data for TLRB 2019 Forest cover map production.

Sentinel2		Sentinel1	
Composite Data	% Valid Pixels	Ascending	Descending
2017_01	1.2		X
2017_02	0		X
2017_03	43.4		X
2017_04	90.6		X
2017_05	3.3		X
2017_06	67.9	X	X
2017_07	8	X	X
2017_08	78.6	X	X
2017_09	3.6	X	X
2017_10	43.3	X	X
2017_11	53.2	X	X
2017_12	92.9	X	X

Table 20. List of available data and identification of the selected input data for IHRB 2017 Forest cover map production.

Sentinel2		Sentinel1	
Composite Data	% Valid Pixels	Ascending	Descending
2018_01	64.2	X	X
2018_02	92.3	X	X
2018_03	95.3	X	X
2018_04	93.5	X	X
2018_05	87.4	X	X
2018_06	92.8	X	
2018_07	69.4	X	X
2018_08	85.4	X	X
2018_09	50	X	X
2018_10	97.7	X	X
2018_11	96.9	X	X
2018_12	96.6	X	X

Table 21. List of available data and identification of the selected input data for IHRB 2018 Forest cover map production.

Sentinel2		Sentinel1	
Composite Data	% Valid Pixels	Ascending	Descending
2019_01	63.9	X	X
2019_02	96	X	X
2019_03	98.4	X	X
2019_04	94.8	X	X
2019_05	96	X	X
2019_06	84.1	X	X
2019_07	52.5	X	X
2019_08	45.9	X	X
2019_09	98.5	X	X
2019_10	94.2	X	X
2019_11	86.8	X	X
2019_12	92.7	X	X

Table 22. List of available data and identification of the selected input data for IHRB 2019 Forest cover map production.

Methodology:

– Pre-processing

The pre-processing of the data, since the coincidence of the type of input data used, follows entirely the methodology previously described for the 2020 LC map. For the Sentinel/2 mosaics time series, a set of vegetation indices have been produced and added to the multispectral information in order to derive the multistack layer to be used as input for the automatic classifier.

– Training

The training collection for the forest maps production follows the same strategy adopted for the LC map production. Even in this case, the samples cover about 4% of the mapping area for each epoch.

Forest map Code	Forest map Class	Area Class [km^2]
1	Forest	51.5
2	No Forest	78.4
Total Training [km^2]		129.9
Area AoI [km^2]		3205.0
Total Area Training / Area AoI		4.1%

Table 23. Training sample composition used for TLRB 2017 Forest Map production.

Forest map Code	Forest map Class	Area Class [km^2]
1	Forest	29.3
2	No Forest	47.5
Total Training [km^2]		76.8
Area AoI [km^2]		2173.2
Total Area Training / Area AoI		3.5%

Table 24. Training sample composition used for IHRB 2017 Forest Map production

– AI algorithms

As for the LC maps, even for the forest cover maps the Light Gradient Boosting Method (LGBM https://lightgbm.readthedocs.io/en/latest/) was selected, but in this case its parameters were settled to perform a binary classification (instead of a multiclass classification) using as metric the *'binary cross entropy'*.

6. Output data:
   - List of the output maps
     - Forest cover map 2016
     - Forest cover map 2017
     - Forest cover map 2018
     - Forest cover map 2019
     - Forest cover map 2020
   - Format
     - Raster: TIFF (Tagged Image File Format)

   Both the basins are provided in the Coordinate Reference System: WGS 84 / UTM zone 51N (EPSG: 32651)

– MMU: 0.10 hectares

Since the size of one Sentinel 2 pixel (10 m x 10 m = 0.01 ha) is 10 times smaller than the minimum mapping unit, a post-processing was carried out by identifying the contiguous pixel groups smaller than the MMU and replacing with the biggest contiguous LC class.

– Nomenclature
The output forest maps will include the following information content.

Value	Color code (RGB)	Description
0	N/A	No data
1	110,170,0	Forest
2	255,255,255	No-forest

Table 25: Description of the annual forest cover maps (2016-2020) values

File	Naming	Description
Forest map raster layers	Forest_Cover_TLRB_20xx.tif Forest_Cover_IHRB_20xx.tif	• GeoTIFF file format • Unsigned integer pixel type • 8-bit pixel depth • UTM51N coordinate system • GSD 30m
Forest map layer files	Forest_Cover.lyr	ArcGis layer file containing the forest cover map symbology

Table 26: Description of the annual forest cover maps (2016-2020) deliverables

3.1.2.2 Quality Control and Validation

- **Land cover maps**
  1. The validation of the land cover maps have been performed by sample points randomly stratified over the land cover class areas. The stratified random technique, also called proportional or quota random sampling, involves dividing the classification map into homogeneous subgroups (the individual land cover classes) and then taking a simple random sample in each subgroup. The sampling is proportionate producing sample sizes that are directly related to the size of the classes (i.e., the larger the class, the more samples will be drawn from it).
  2. The land cover class assignment of the validation points has been done using the available reference data not directly used in the automatic classification process (e.g. Google Earth, OpenStreetMap, Spatial Database of Planted Trees (SDPT) 2015 dataset2, Global Forest Change (GFC) 2000-2019 dataset3)
  3. Results
     The following tables represent the output confusion matrices resulting from the comparison between the validation points and the classification maps for each epoch both in Ilog-Hilabangan River Basin (IHRB) in the Visayas Region and the Tagum-Libuganon River Basin (TLRB).

Overall Accuracy = (332/369) 89.9729%
Kappa Coefficient = 0.8428

Ground Truth (Pixels)

Class	Artificial	Forest	Tree Crops	Permanent her	Periodically
Unclassified	0	0	0	0	0
Artificial	5	1	0	2	0
Forest	0	191	7	1	1
Tree Crops	0	6	54	0	0
Permanent her	0	6	4	33	2
Periodically	0	2	2	2	43
Water	0	0	0	0	0
Total	5	206	67	38	46

Figure 4: Confusion matrix of LC 2000 in TLRB basin

Overall Accuracy = (357/422) 84.5972%
Kappa Coefficient = 0.7597

Ground Truth (Pixels)

Class	Artificial	Forest	Tree Crops	Permanent her	Periodically
Unclassified	0	0	0	0	0
Artificial	5	2	1	0	0
Forest	0	190	6	1	0
Tree Crops	0	38	96	3	3
Permanent her	0	4	2	37	3
Periodically	0	1	0	0	24
Water	0	0	0	0	0
Total	5	235	105	41	30

Figure 5: Confusion matrix of LC 2016 in TLRB basin

Overall Accuracy = (410/461) 88.9371%
Kappa Coefficient = 0.8412

Ground Truth (Pixels)

Class	Artificial	Forest	Tree Crops	Permanent her	Periodically
Unclassified	0	0	0	0	0
Artificial	12	0	1	0	0
Forest	0	178	4	7	0
Tree Crops	0	21	124	0	1
Permanent her	0	4	4	60	6
Periodically	0	0	0	0	30
Water	0	0	1	0	0
Total	12	203	134	67	37

Figure 6: Confusion matrix of LC 2020 in TLRB basin

Overall Accuracy = (318/385) 82.5974%
Kappa Coefficient = 0.7263

Ground Truth (Pixels)

Class	Forest	Permanent her	Periodically	Water	Total
Unclassified	0	0	0	0	0
Forest	121	5	9	1	136
Permanent her	8	41	7	0	56
Periodically	10	25	150	2	187
Water	0	0	0	6	6
Total	139	71	166	9	385

Figure 7: Confusion matrix of LC 2000 in IHRB basin

Overall Accuracy = (380/410) 92.6829%
Kappa Coefficient = 0.8922

	Ground Truth (Pixels)				
Class	Artificial	Forest	Permanent her	Periodically	Water
Unclassified	0	0	0	0	0
Artificial	4	0	2	0	1
Forest	0	141	2	2	0
Permanent her	0	3	102	6	1
Periodically	0	2	11	130	0
Water	0	0	0	0	3
Total	4	146	117	138	5

Figure 8: Confusion matrix of LC 2016 in IHRB basin

Overall Accuracy = (368/402) 91.5423%
Kappa Coefficient = 0.8677

	Ground Truth (Pixels)				
Class	Artificial	Forest	Permanent her	Periodically	Water
Unclassified	0	0	0	0	0
Artificial	4	0	0	0	1
Forest	0	129	7	0	0
Permanent her	0	3	51	1	0
Periodically	0	2	18	178	2
Water	0	0	0	0	6
Total	4	134	76	179	9

Figure 9: Confusion matrix of LC 2020 in IHRB basin

- **Forest cover maps**
  1. The validation of the forest cover maps have been performed by sample points randomly stratified over the forest/no forest areas.
  2. The class assignment of the forest/ no forest validation points has been done using the available reference data not directly used in the automatic classification process (e.g. Google Earth, OpenStreetMap, Spatial Database of Planted Trees (SDPT) 2015 dataset2, Global Forest Change (GFC) 2000-2019 dataset3)
  3. Results
     The following tables represent the output confusion matrices resulting from the comparison between the validation points and the Forest cover maps for each epoch both in Ilog-Hilabangan River Basin (IHRB) in the Visayas Region and the Tagum-Libuganon River Basin (TLRB).

Overall Accuracy = (356/403) 88.3375%
Kappa Coefficient = 0.7373

	Ground Truth (Pixels)		
Class	Forest	No Forest	Total
Unclassified	0	0	0
Forest	246	32	278
No Forest	15	110	125
Total	261	142	403

Figure 10: Confusion matrix of Forest map 2017 in TLRB basin

Overall Accuracy = (382/416) 91.8269%
Kappa Coefficient = 0.8268

Class	Ground Truth (Pixels) Forest	No Forest	Total
Unclassified	0	0	0
Forest	241	25	266
No Forest	9	141	150
Total	250	166	416

Figure 11: Confusion matrix of Forest map 2018 in TLRB basin

Overall Accuracy = (376/400) 94.0000%
Kappa Coefficient = 0.8743

Class	Ground Truth (Pixels) Forest	No Forest	Total
Unclassified	0	0	0
Forest	231	19	250
No Forest	5	145	150
Total	236	164	400

Figure 12: Confusion matrix of Forest map 2019 in TLRB basin

Overall Accuracy = (378/407) 92.8747%
Kappa Coefficient = 0.8384

Class	Ground Truth (Pixels) Forest	No Forest	Total
Unclassified	0	0	0
Forest	118	2	120
No Forest	27	260	287
Total	145	262	407

Figure 13: Confusion matrix of Forest map 2017 in IHRB basin

Overall Accuracy = (397/404) 98.2673%
Kappa Coefficient = 0.9569

Class	Ground Truth (Pixels) Forest	No Forest	Total
Unclassified	0	0	0
Forest	109	2	111
No Forest	5	288	293
Total	114	290	404

Figure 14: Confusion matrix of Forest map 2018 in IHRB basin

```
Overall Accuracy = (368/403)  91.3151%
Kappa Coefficient = 0.7906

                      Ground Truth (Pixels)
    Class               Forest    No Forest        Total
 Unclassified                0            0            0
       Forest              100            6          106
    No Forest               29          268          297
        Total              129          274          403
```

Figure 15: Confusion matrix of Forest map 2019 in IHRB basin

- **Annual forest cover maps (2001-2015)**

The annual forest cover maps products for the period 2001-2015 were inspected before delivery according to the Quality Control list presented in Table 27.

QC elements	Deliverables	QC measure	Quality target
Completeness	Raster layers	Full visual check in GIS software	100% AOI coverage
	Delivery folders	Full inspection of the delivery folder	Missing items: 0
Readability	Raster layers	Full visual check in GIS software	100% readability
Compliance	Raster layers	Full visual inspection in GIS software	100% format compliance
Thematic accuracy	Raster layers	Thematic accuracy for input data	Overall accuracy >80%

Table 27: QC list for annual forest cover maps (2001-2015)

As for the thematic accuracy, a validation process was carried out in order to ensure that the overall accuracy for 2000 land cover maps reaches at least 80% as required by the user. Regarding Tropical Moist Forest data, the authors presented an overall accuracy of 91.4% for the 1990-2019 dataset.

3.1.2.3 Usage, Limitations and Constraints

Some examples of the land cover products are presented in this section.

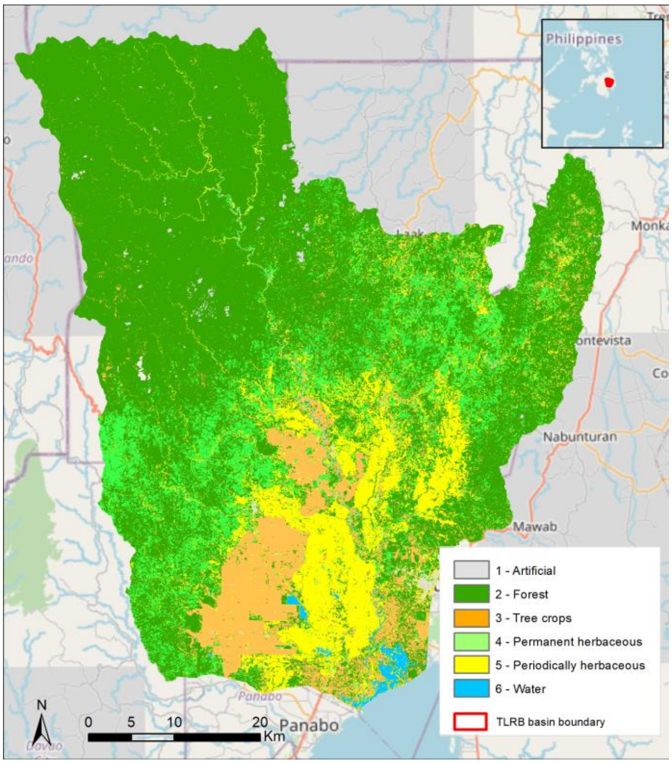


Figure 16: LC 2000 in TLRB basin

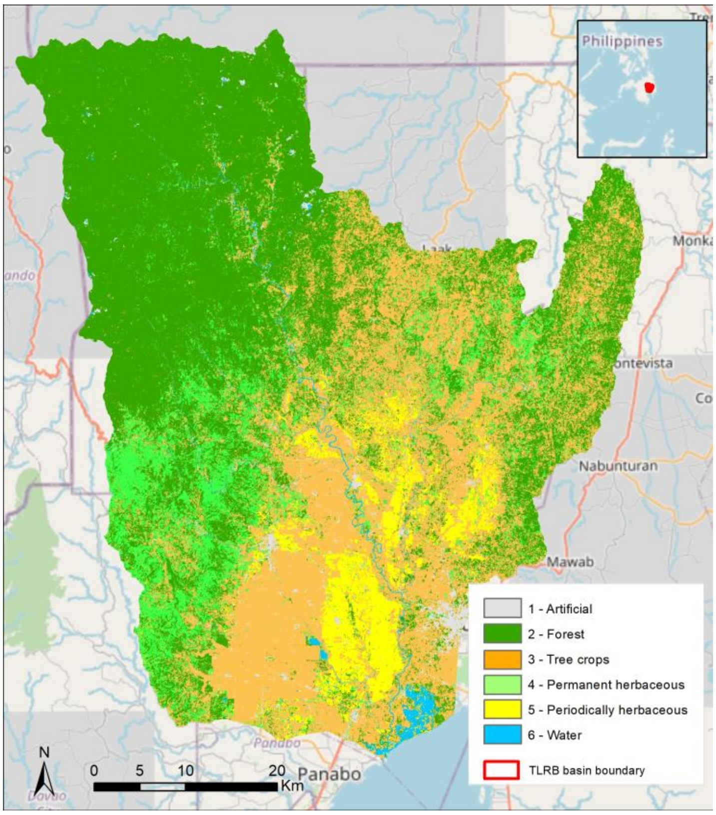


Figure 17: LC 2016 in TLRB basin

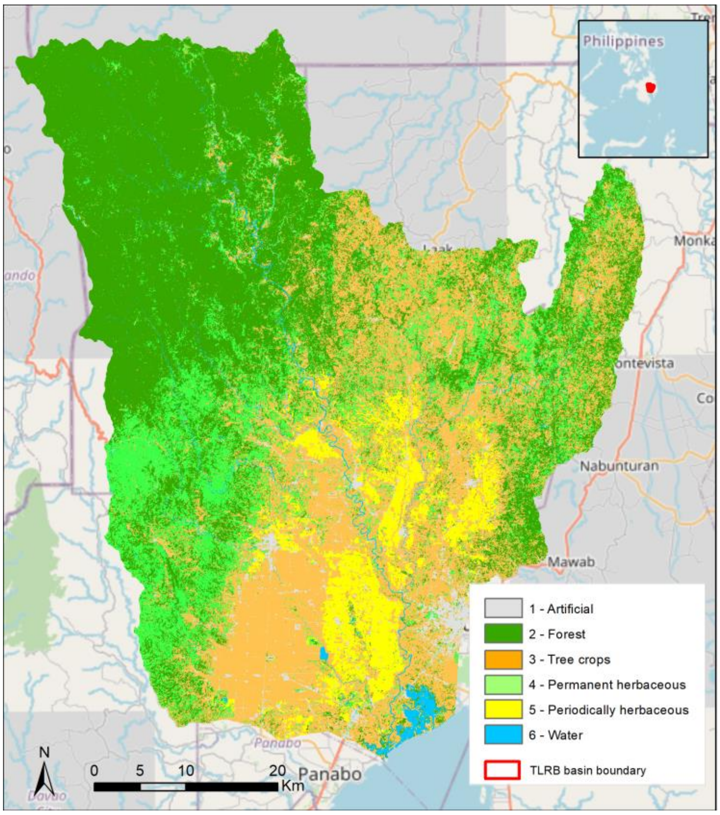


Figure 18: LC 2020 in TLRB basin

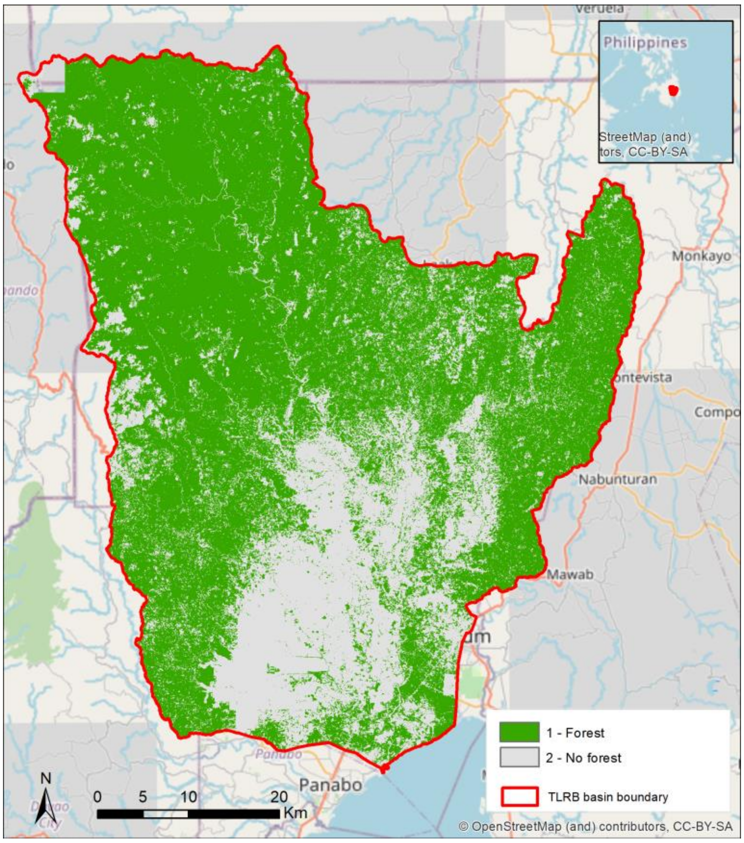


Figure 19: Forest map 2017 in TLRB basin

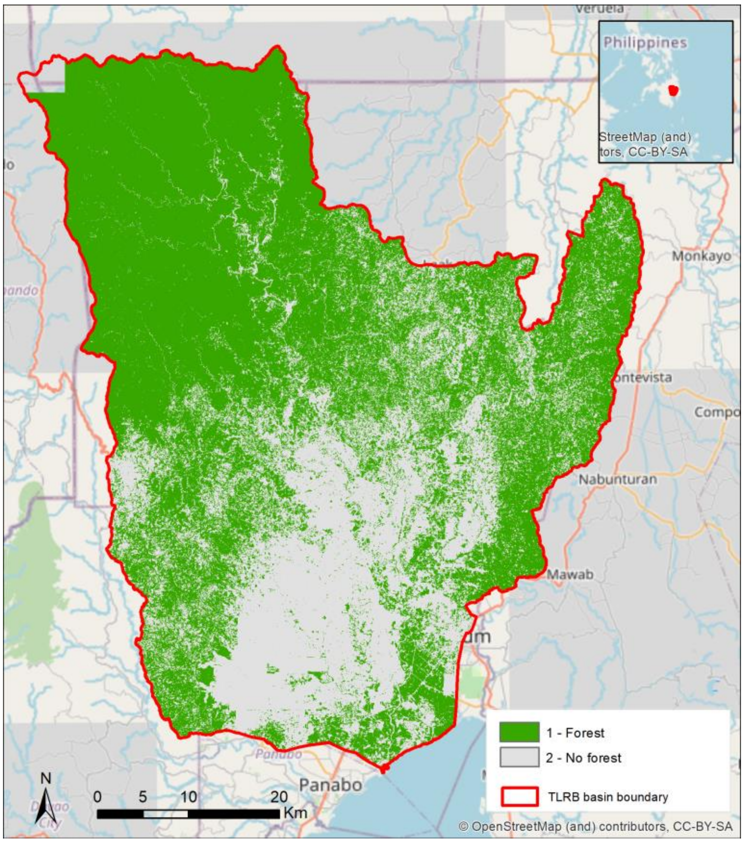


Figure 20: Forest map 2018 in TLRB basin

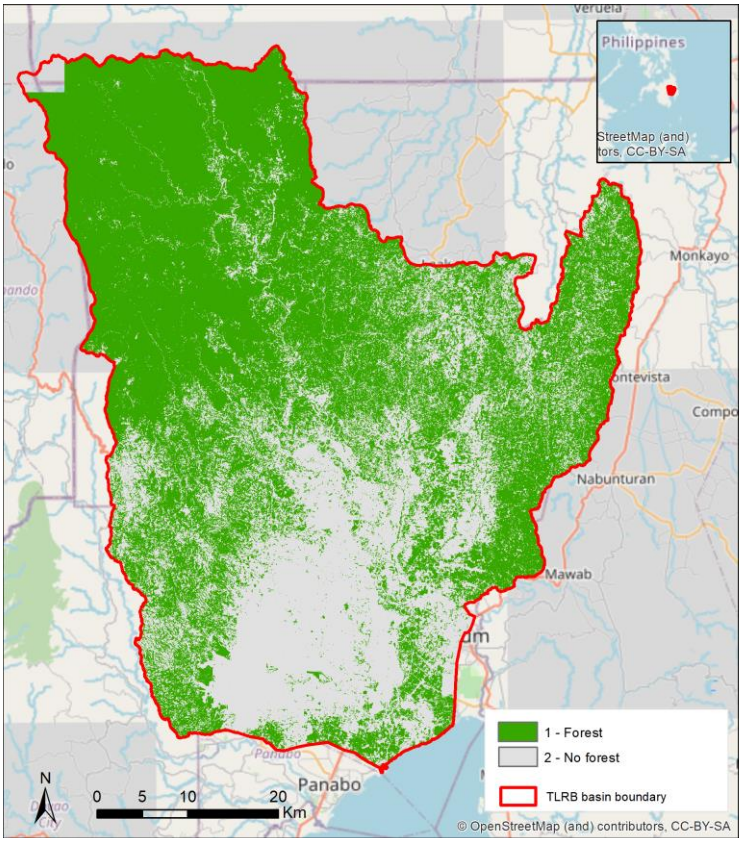


Figure 21: Forest map 2019 in TLRB basin

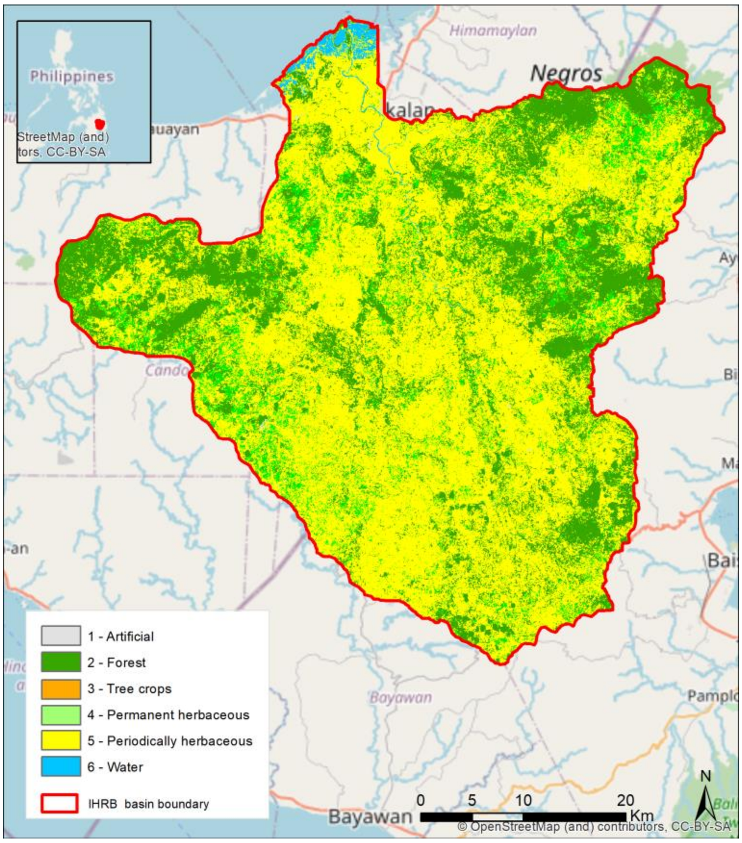


Figure 22: LC 2000 in IHRB basin

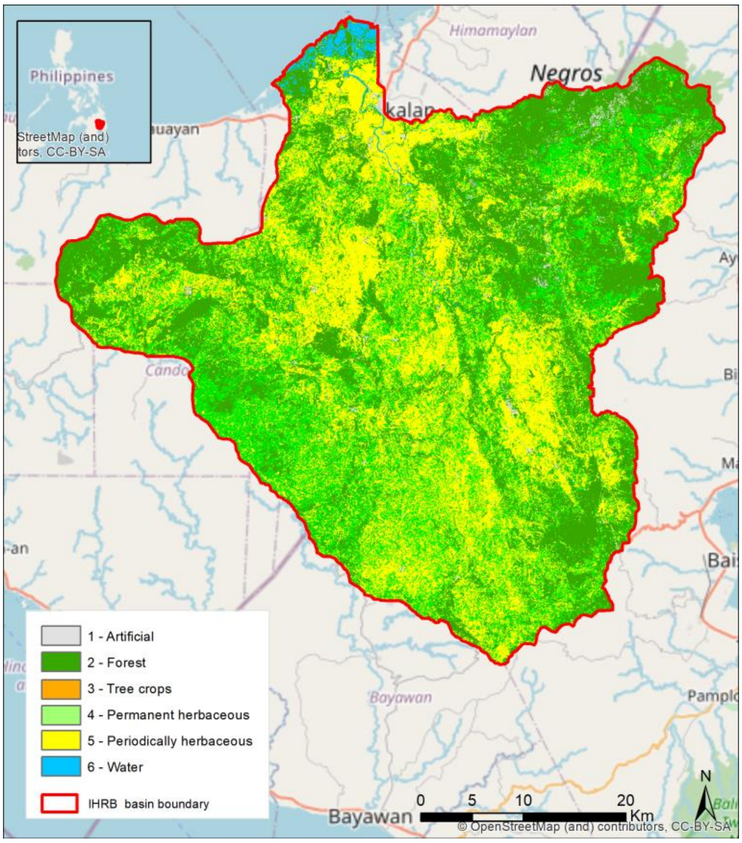


Figure 23: LC 2016 in IHRB basin

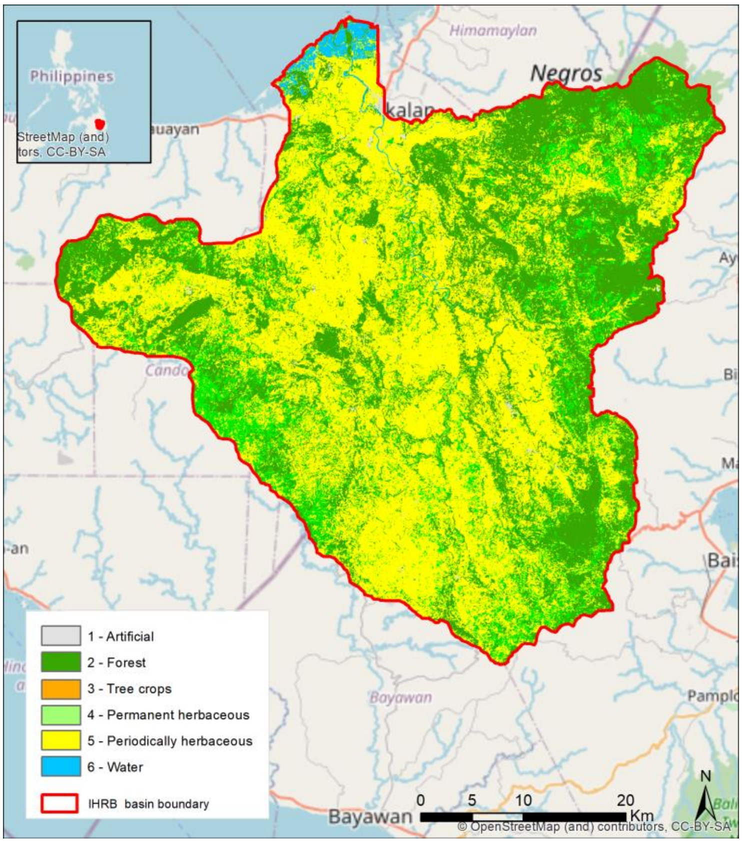


Figure 24: LC 2020 in IHRB basin

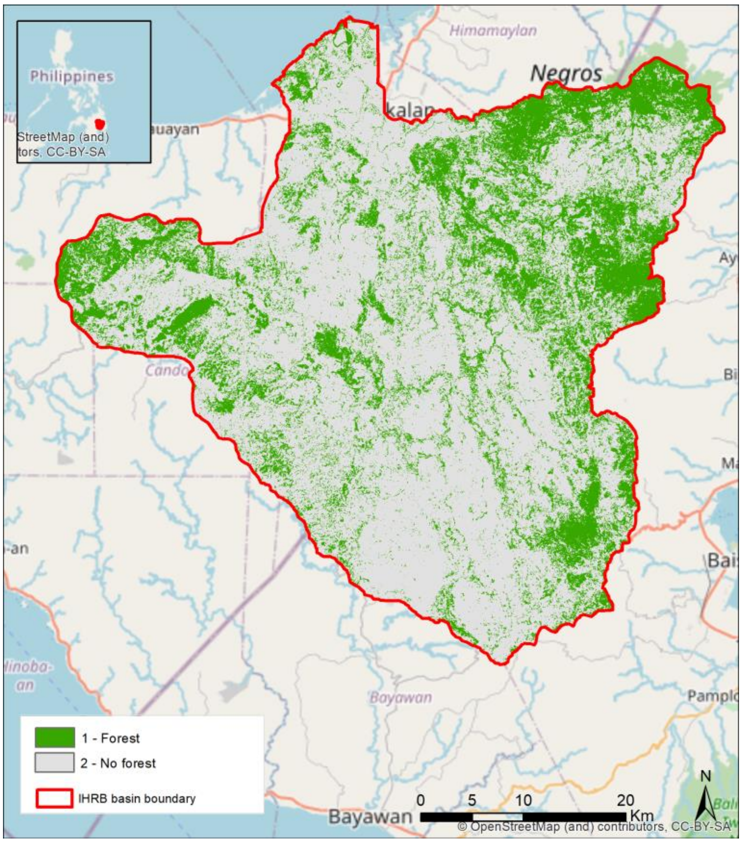


Figure 25: Forest map 2017 in IHRB basin

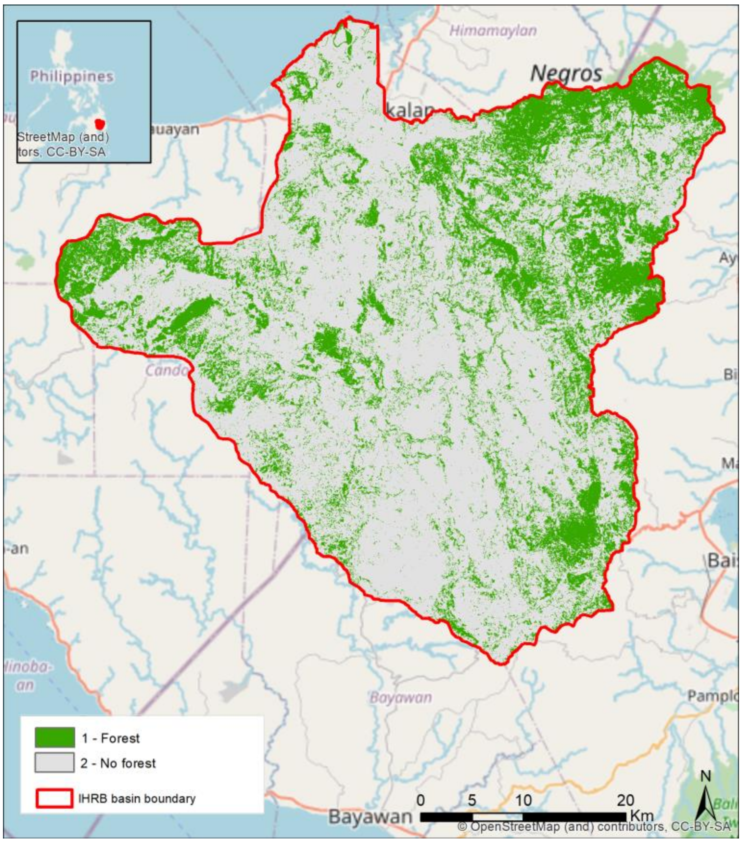


Figure 26: Forest map 2018 in IHRB basin

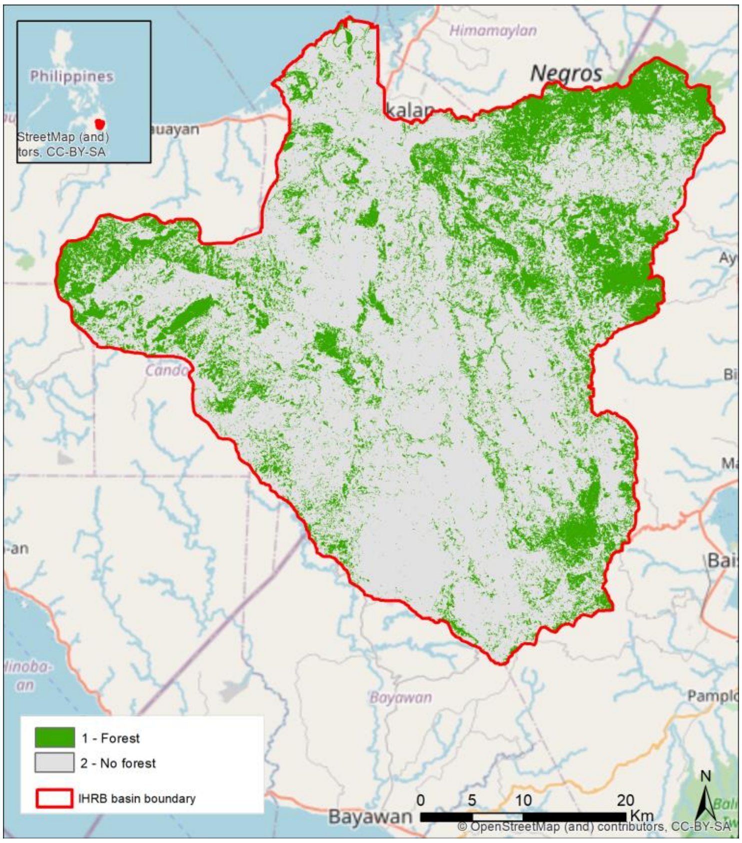


Figure 27: Forest map 2019 in IHRB basin

Annual forest cover maps (2001-2015)

Thematic content and accuracy for the annual forest cover maps are dependent on both the 2000 land cover maps and the Tropical Moist Forest dataset accuracy. Moreover, some missing areas (i.e. NoData) due to cloud coverage were already present in the initial Land Cover map for the Tagum-Libuganon river basin. Hence, these areas were kept during the annual forest cover maps processing.

3.1.3 Service 2: Forest Loss Rate Methodology

3.1.3.1 Specifications

The Service 2 was focused on the understanding a long-term evolution of landscapes which results in a profound change to the land use/cover over large spatial extents with a specific interest to the evolution from a forested to an agricultural area (Forest to Agriculture Transit - FAT), which leads to deforestation. To understand what are the most likely intermediate steps of such a transit and which evolutionary paths is more likely than others, we followed the Stochastic, Empirically Informed Model of Landscape Dynamics proposed by Nowosad and Stepinski (2019) to analyse the Deforestation Scenarios at global scale. The principal hypothesis under this model is that the evolution of the landscape from forested to agricultural happen starting from a landscape of homogeneous forest and ending with a landscape of homogeneous agricultural land. The model simulates the evolution of the landscape from forested to agricultural and yields different deforestation scenarios and their probabilities. A landscape currently covered entirely by forest will most likely evolve to an agricultural landscape along the maximum likelihood trajectory which reflects a prevailing series of circumstances; this is the most likely deforestation scenario. A different evolution of FAT along less likely trajectories can happen under rare circumstances; such occurrences are considered less likely deforestation scenarios.

The LC maps produced in Service 1 and created for the reference years (2000, 2016, and 2020) were the input for Service 2. Since, as previously described and motivated, the automatic classification applied to IHRB and TLRB basins was slightly different with respect to the tree crops separation, two different specific testing exercises were performed for the two basins. Specifically, for the IHRB only the "periodically herbaceous" class was included in the agricultural area while for the TLRB also "periodically herbaceous" and "tree crops" were considered as agricultural class. In the following paragraphs, the analysis of the complete time series 2000-2016-2020 of TLR basin is presented considering that the outcomes are more meaningful due to the separation of tree crops the tree covered areas.

- Input data:
    1. LC maps for the epochs 2000, 2016 and, 2020

- Methodology

  Following the work flow proposed by Nowosad and Stepinski (2019), the land change in a given areal unit was conceptualized as a modification of landscape pattern within this unit between two epochs of observations (t_0 and t_1). The entire workflow is reported in the following figures.

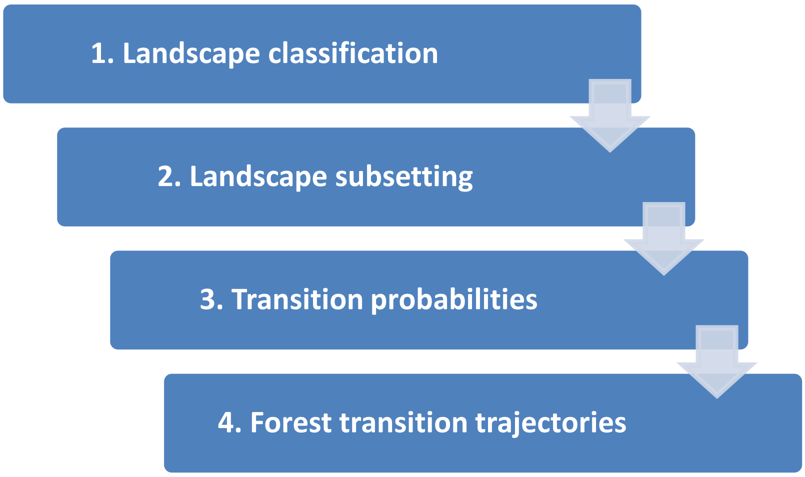


Figure 28. Workflow of Service 2

1. For the Landscape classification, the AoI was tessellated into 35,611 and 24,144 nonoverlapping square areal units of the size 0.3 km × 0.3 km (30 × 30 LC maps cells). A mosaic, formed by the selected land categories in a given areal unit at a given epoch, constitutes a landscape characterized by a specific configuration (a pattern geometry) and a thematic content (names of land cover classes present). The landscape classification was done by computing the 17 landscapes metrics listed in Nowosad and Stepinski (2018) using R software with the package *"landscapemetrics"*. In addition, also the percentages of area covered by each land cover category represented in the areal unit (i.e. PLANDs in the landscape metrics nomenclature) were estimated to characterize the composition of the patterns. This step was the highest computationally demanding part of the calculation for the Service 2.

	Name
ai	aggregation index
cohesion	patch cohesion index
contag	connectance
division	division index
iji	interspersion and juxtaposition index
pd	patch density
pladj	percentage of like adjacencies
lpi	largest patch index
msidi	modified simpson's diversity index
msiei	modified simpson's evenness index
shdi	shannon's diversity index
shei	shannon's evenness index
sidi	simpson's diversity index
siei	simspon's evenness index
pafrac	perimeter-area fractal dimension
contig_mn	contiguity index
contig_sd	contiguity index

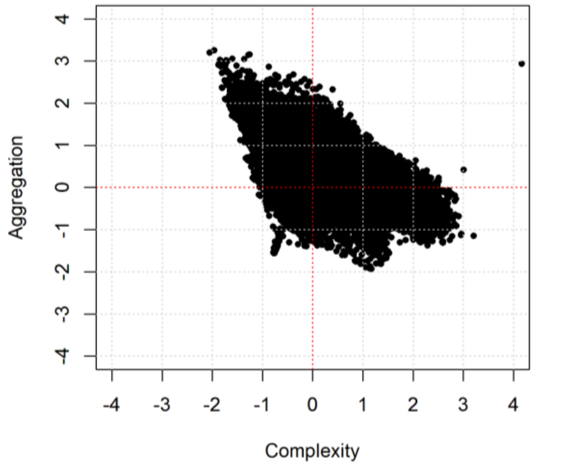
Table 28: Landscape metrics used to characterize the composition of the patterns.

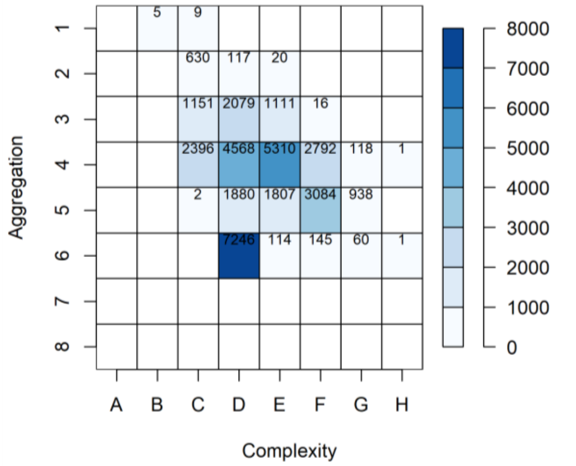
These 17 metrics could be briefly parametrized by only two metrics interpreted as "Complexity" (C) and "Aggregation" (A) following the Nowosad & Stepinski (2018) obtained by performing a PCA analysis (performed with R software) on these indexes: the first two rotated principal component define

the complexity and aggregation. The quantization of such two components in a bi-dimensional histogram describes the landscape types.
Indeed, the "C" component is positively correlated with all diversity metrics: it increases when the histogram of the LC categories composition inside a unit's cell is flat (all the LC categories are well represented inside the cell) and it decrease when the histogram is peaked (the unit is dominated by a single LC category).
The "A" component measures the aggregation or connectivity: it increases when cells of the same LC categories are more aggregated within the unit. Therefore, small values of A indicate a landscape with a large number of small patches.
The metrics "C" and "A" are principal components and have a theoretical range of values between -4 and 4 (Nowosad & Stepinski, 2018). It was possible to classify all landscapes, based on the values of C and A, by constructing an equispaced C-A grid, that divides the C/A space into 64 sections. In such grid, each section is characterized by ΔC = 1 and ΔA = 1.
This is a 2-D classification that takes into consideration values of C and A but not landscape's thematic content and allows to classify all the landscapes into up to 64 classes with respect to their configurations. Moreover 64 are the theoretically maximum possible, while the real number of classes is generally lower and depends on the specific characteristics of the considered landscape.
As output of this step, for each LC map, a corresponding C-A classification map (with a final resolution of 30 x 10 m = 300 meters) was obtained .

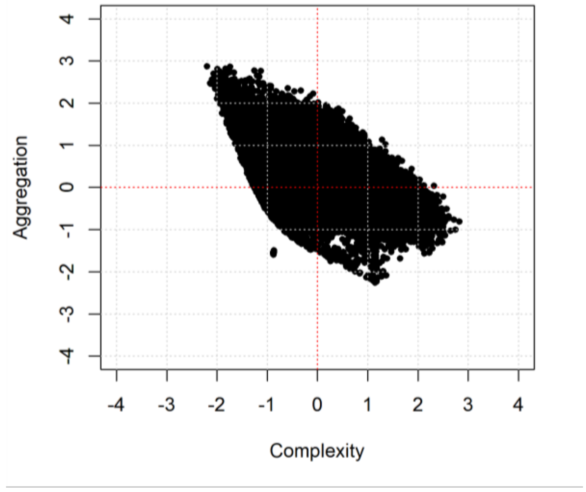
TLRB

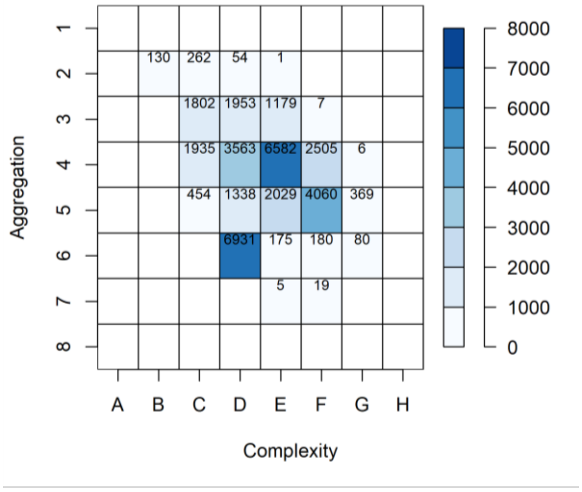
LC 2000



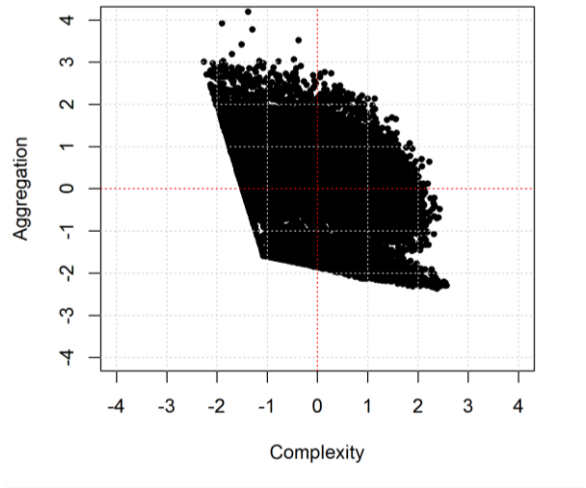


LC 2016





LC 2020



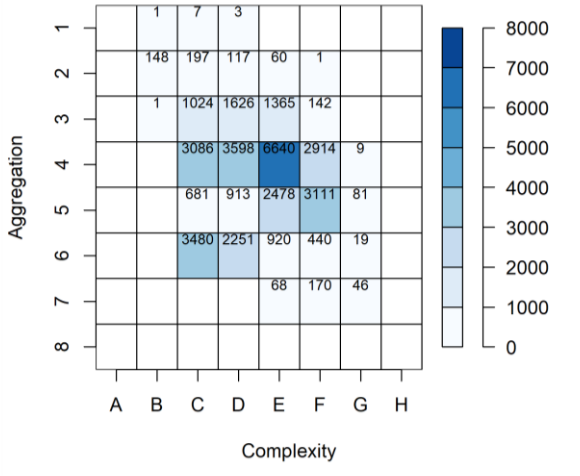
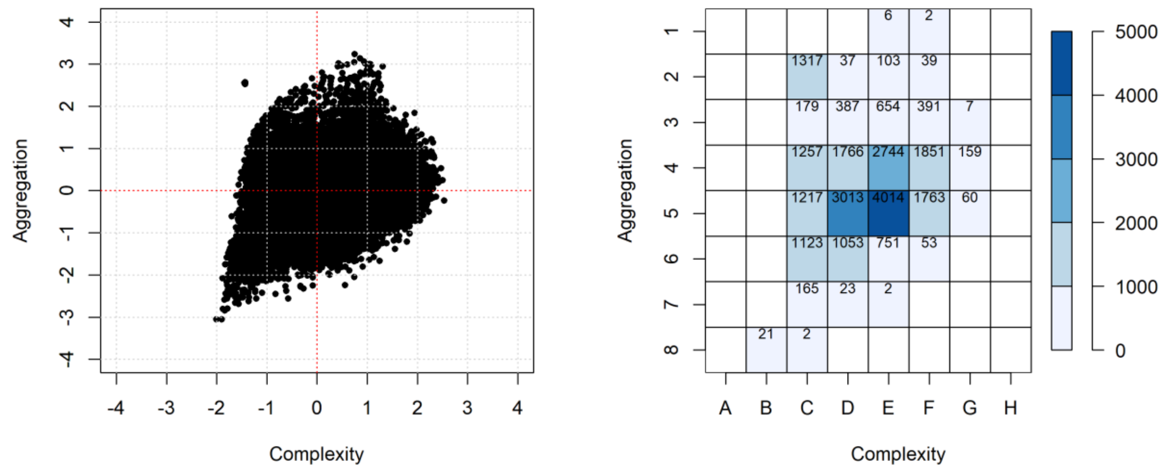
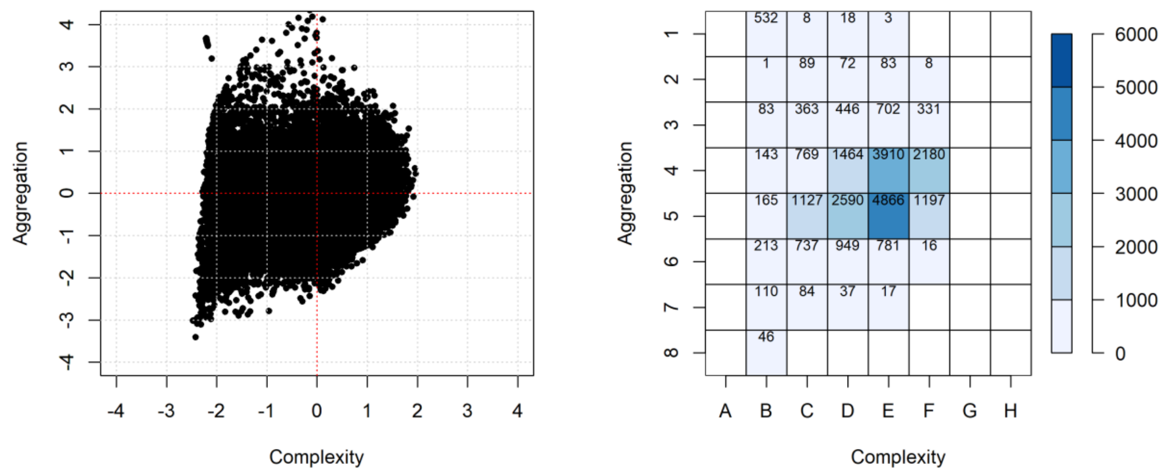


Figure 29. C/A classification for the TLRB

IHRB
LC 2000


Aggregation
Complexity
6
2
1317
37
103
39
179
387
654
391
7
1257
1766
2744
1851
159
1217
3013
4014
1763
60
1123
1053
751
53
165
23
2
21
2
5000
4000
3000
2000
1000
0

LC 2016


Aggregation
Complexity
532
8
18
3
1
89
72
83
8
83
363
446
702
331
143
769
1464
3910
2180
165
1127
2590
4866
1197
213
737
949
781
16
110
84
37
17
46
6000
5000
4000
3000
2000
1000
0

LC 2020

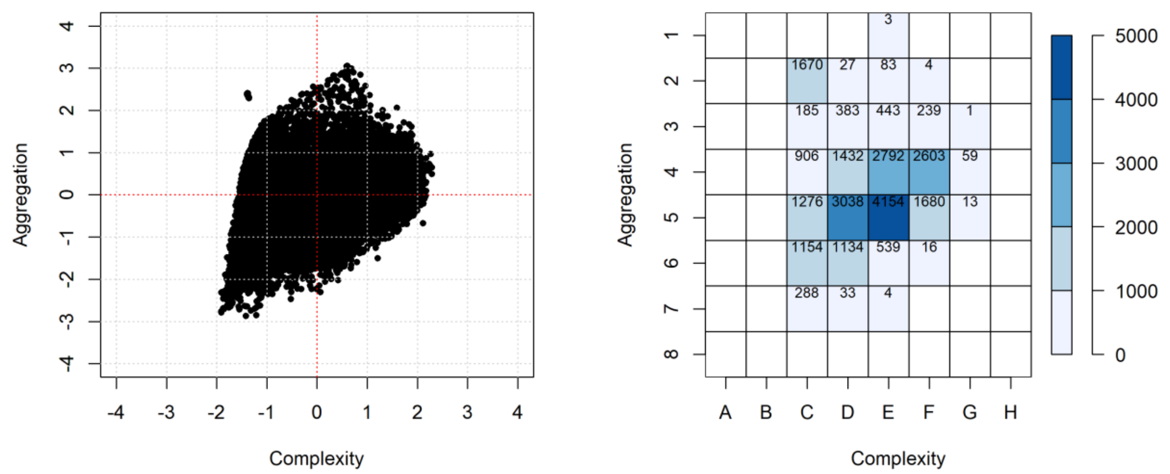


Figure 30. C/A classification for the IHRB

2. Landscape subsetting

The classification of landscapes based on C and A concerns their geometric configuration but ignores their thematic content (which is given only by landscape composition represented by the PLANDs[1] of each LC class). For this reason, it was necessary to focus only on a suitable subset of landscapes (i.e., LC categories composition) based on the process of deforestation.

In the scenario proposed by Nowosad and Stepinski (2018), where forest is replaced by agricultural land cover, a sequence of transitions between subsequent landscape types occurs. The starting point is a fully forested landscape and the ending point is a fully agricultural landscape; this sequence is referred to as a FAT trajectory. During our analysis, each FAT trajectory was divided into two phases:

1. Forest-dominant phase (FAT1).
   The model for FAT1 was built starting from a subset of the landscapes (i.e., aerial units) subject to the following conditions: forest is the primary land cover class and agriculture is the secondary land cover class in LC at t_0. Along the evolution of the landscape towards t_1, the transitions are towards a diminishing share of the forest.
2. Agriculture-dominant phase (FAT2).
   The model of FAT2 was built starting from a subset of the landscapes subject to the following conditions: agriculture is the primary land cover class and forest is the secondary land cover class in LC at t_0. Along the evolution of the landscape towards t_1, the transitions are towards an increasing share of agriculture.

TLRB	
FAT 1	**FAT 2**
2000 - 2016	

[1] The percentages of area covered by each land cover category represented in the areal unit.

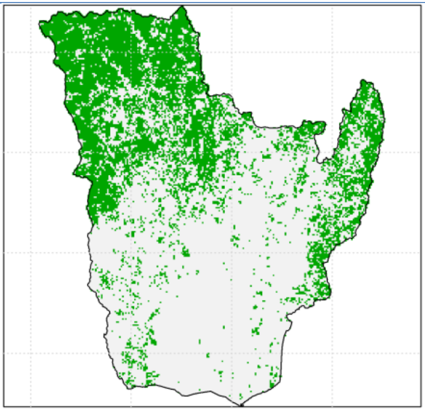
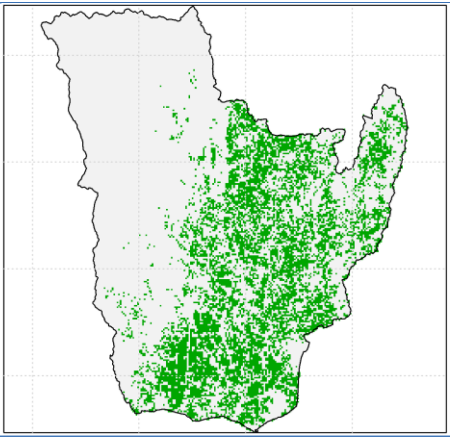
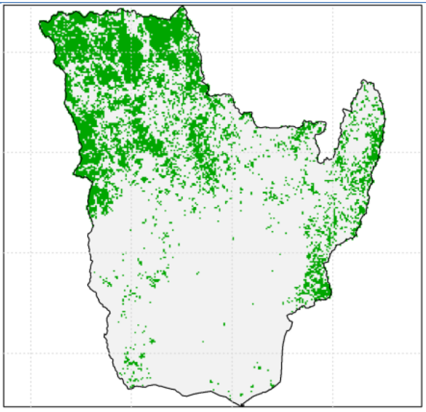
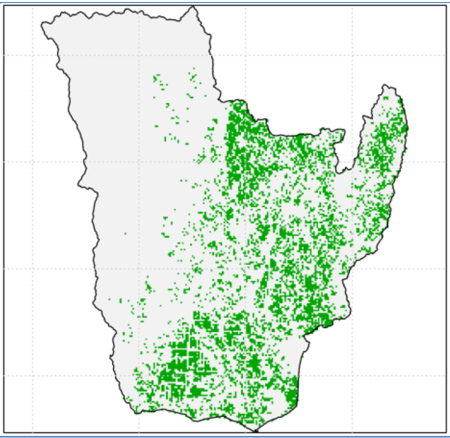
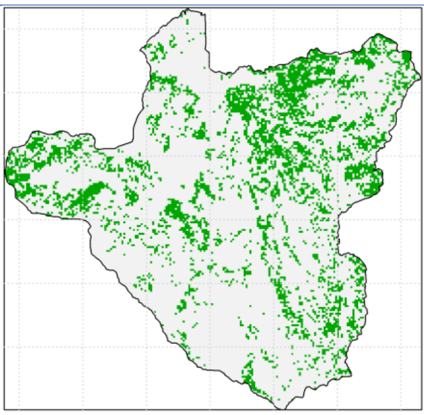
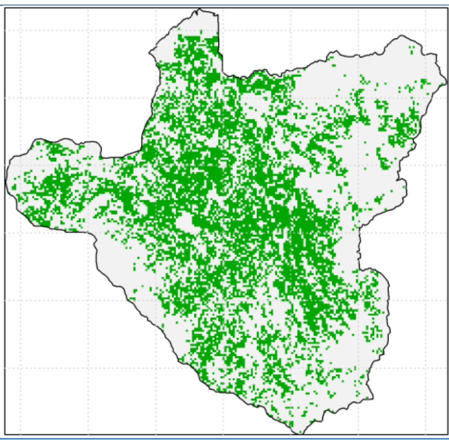
	
2016-2020	
	

Figure 31. Landscape subsetting (FAT1 and FAT2 landscape units in green) for the TLRB.

IHRB	
FAT 1	**FAT 2**
2000 - 2016	
	

2016-2020

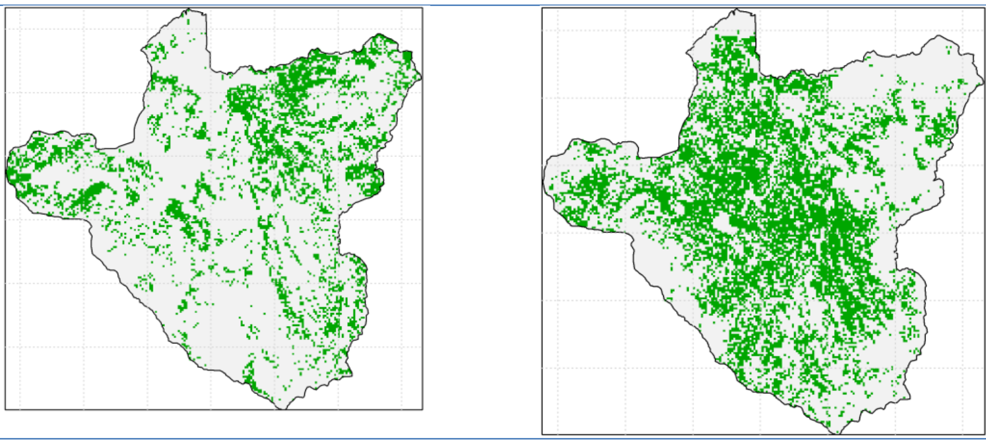


Figure 32. Landscape subsetting (FAT1 and FAT2 landscape units in green) for the IHRB.

3. Transition Probabilities
   A specific landscape (LC category composition) at t_o may change into different types of landscape at t_1. The frequencies of landscape types changes were calculated for each landscape type. This results in a probability distribution of changes that can occur for a given landscape type towards any other type of landscapes.

4. Forest transaction trajectories
   Instead of run thousands of simulated trajectories, following Nowosad (2021), we used the probability distribution of the transitions. In this way, we studied some defined trajectories choosing always the most probable transition from each stage to the next one. The direction was defined moving always towards an increase of C and a decreasing of A in case of FAT1, and the opposite for the case of FAT2.
   A stage-to-stage transition probability of a landscape during Δt was estimated. Such probability is the product of a probability (p_{ch}) that a landscape would transit to another type and a probability (p_{tr}) of the landscape transitioning to a specific different type. Thus, the product p_{ch} p_{tr} is a probability of a specific stage-to-stage transition occurring during a single Δt and the mean waiting time (in units of Δt) for such transition to occur is $1/p_{ch}p_{tr}$. This reflects the mean waiting time for such a transition to occur.
   The FAT trajectories for the TLRB are reported in Figure 33 and Figure 34.

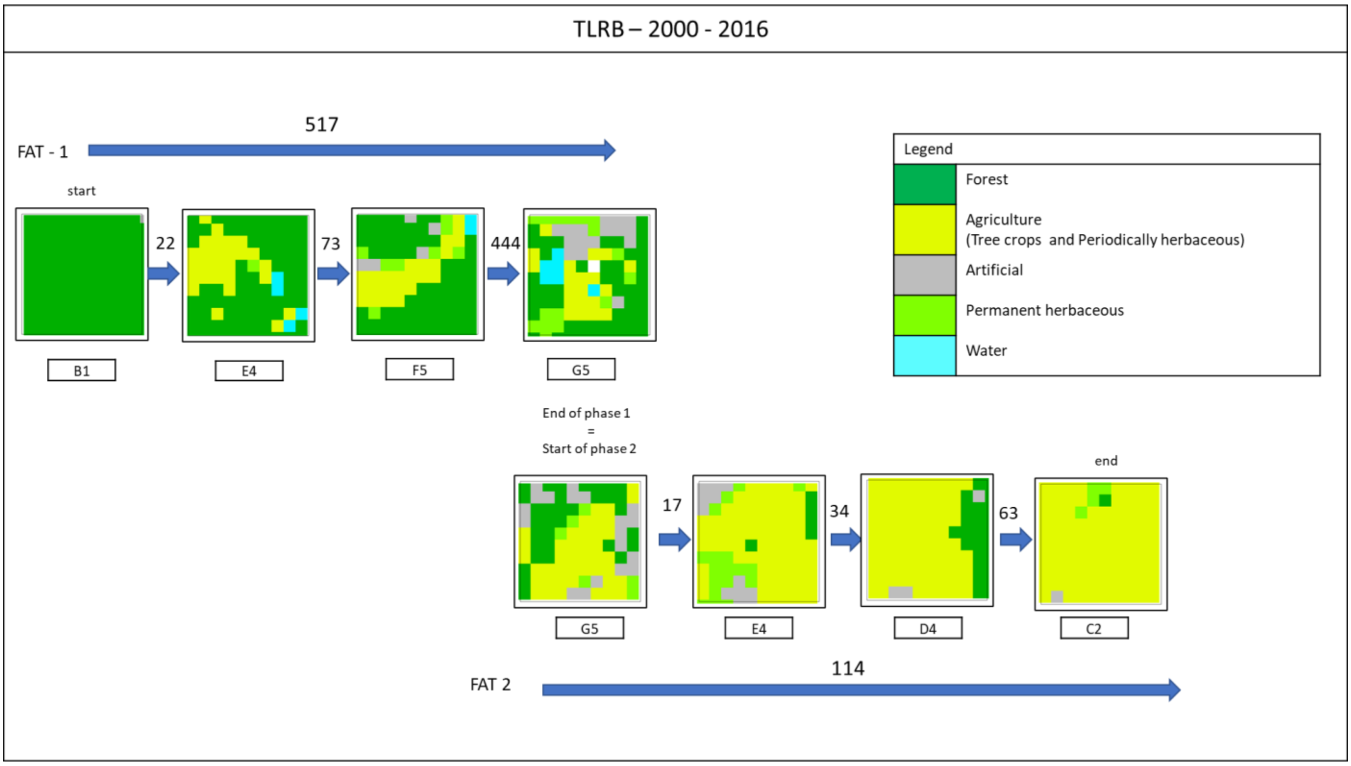


Figure 33. Summary of the forest-dominated phase (FAT1) and agriculture-dominated phase (FAT2) trajectories simulating an evolution from a homogeneous forested to a homogeneous agricultural landscape for the TLRB during the epochs 2000-2016.

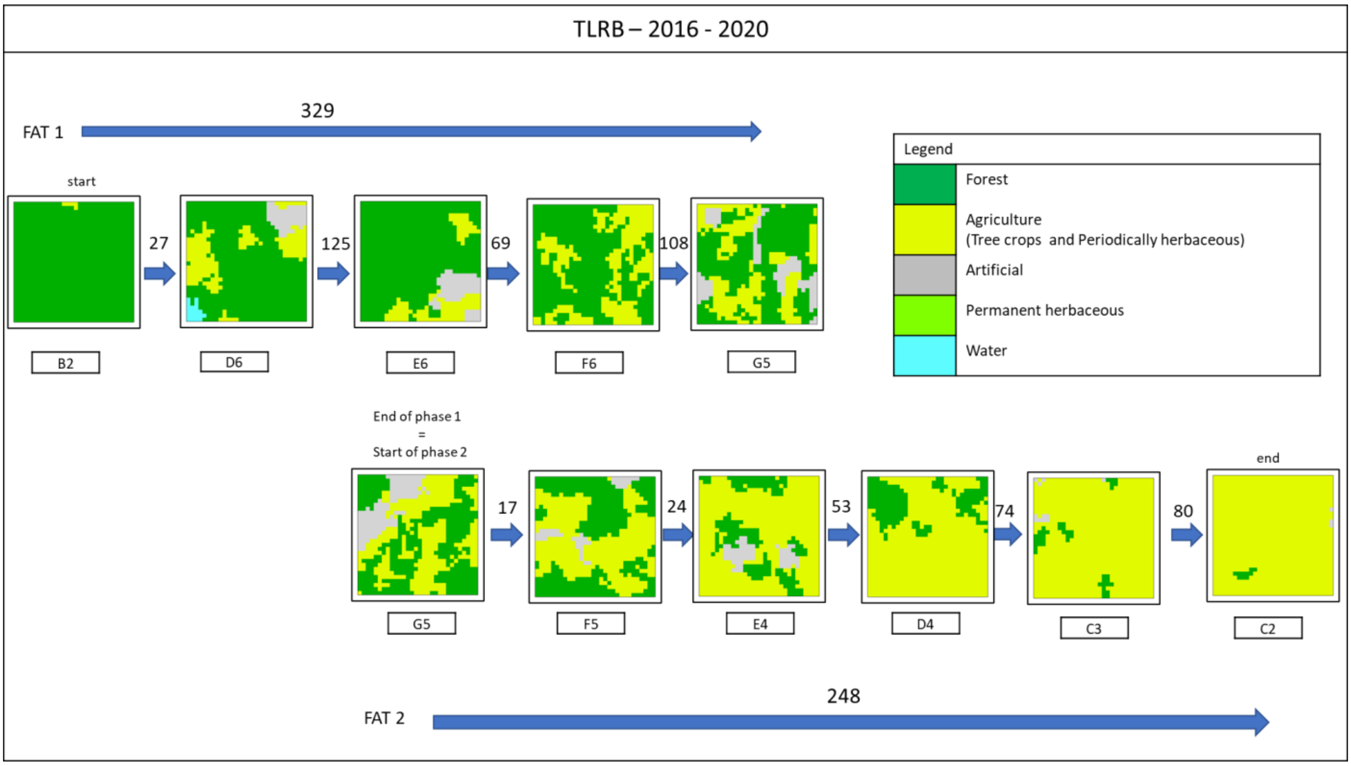


Figure 34:. Summary of the forest-dominated phase (FAT1) and agriculture-dominated phase (FAT2) trajectories simulating an evolution from a homogeneous forested to a homogeneous agricultural landscape for the TLRB during the epochs 2016-2020.

A FAT1 trajectory, which implies a diminishing share of the forest, starts from the upper left side of the C-A diagram characterized by high aggregation and a low complexity, typically of a uniform landscape. At each step of the trajectory, an increasing of complexity and/or a decreasing of aggregation occurs. The process continues until the opposite corner of the C-A diagram (characterized by high complexity and low aggregation) is reached. This condition represents a tipping point, because still the forest class was prevalent for the considered landscape. This point coincides with the end of the phase 1 and the starts of the phase 2. Phase 2 leads to a uniform agricultural landscape through states characterized by decreasing of complexity and/or an increasing of aggregation.

Numbers placed over the arrows in Figure3 and Figure 34 indicate the waiting time needed for the transition between two consecutive stages to occur. For example, a transition B1 → E4 Bd in the most likely FAT1 trajectory for the TLRB during the period 2000 -2016 is equal to 22 years. This value reflects the mean waiting time for such a transition to occur and depend on both p_{ch} and p_{tr}. The total mean waiting time for the completion of the FAT1 trajectory is 539 years while for the FAT2 is 114.

Analysing the mean waiting times in trajectories, reported synthetically in Table 29Table , it is possible to observe that Phase 1 of FAT trajectories takes longer to complete than Phase 2. This result was confirmed also by the study conducted at global scale by Nowosad and Stepinski (2019) which observed as FAT takes longer to lose the first ~50% of the forest but once reached this tipping point, the FAT for the remaining forest accelerates.

	FAT1	FAT2	FAT1 + FAT2
2000-2016	539	114	653
2016-2020	329	248	577

Table 29: Mean waiting times for FAT observed in the TLRB.

The results obtained for the TLRB suggest a moderate deceleration of FAT since the total waiting time passed from the 653 of the mid-term (2000-2016) to the 577 year of the short-term (2016-2020). This result was achieved for both FAT1 and FAT2.

3.1.3.2 Quality Control and Validation

The plausibility of the results of the Forest Loss Rate analysis has been checked by comparing them with the outcomes of the work published by Nowosad and Stepinski (2019), used as main reference for the model configuration.

3.1.3.3 Usage, Limitations and Constraints

The results obtained in Service 2 could be used to interpret the deforestation scenario at the river basin scale. Their use for conservation policy development should consider how the size of the grid used to define the landscape units can have a relevant impact on the final results. It should be carefully selected taking into account the size of AoI with the characteristics of the LC input map. In our case the selected 30x30 pixels grid was considered as enough representative for the two analyzed landscapes (i.e., forest and agricultural).

Finally, as suggest also by Nowosad and Stepinski (2019), it is possible to use the model developed in this Service also for land change scenarios others than FA (e.g., desertification or urbanization), by selecting an appropriate subset of the dataset.

3.1.4 Training and capacity building proposal

The proposed online training is composed by two sessions aimed to describe the results of the activities and the applied methodologies. The training will be held through an online training platform for an interactive videoconferencing: MS Teams platform is proposed. In particular, the following topics are proposed to be addressed.

Session 1 (3 hours)	
Service 1 Land cover mapping	○ Introduction to the input EO and other data availability ▪ EO input data availability over TLRB and IHRB basins ▪ Application and usability of available forest products at global scale ▪ Feasibility of forest monitoring (annual or twice per year) ○ AI algorithms for land cover and forest mapping ○ Mapping results obtained for TLRB and IHRB basins ○ Discussion and QA
	○ Training sample selection (sampling scheme, samples selection and interpretation) ○ Validation activity ○ Discussion and QA
Session 2 (2 hours):	
Service 2 Forest Loss Rate analysis	○ Introduction to the workflow applied for the Service 1 based on Nowosad and Stepinski (2019) model for forest dynamics analysis ○ The landscape metrics calculation ○ Building the forest transactions trajectories ○ Results obtained for TLRB and IHRB basins ○ Discussion and QA

Table 30: Proposed agenda for the online training and capacity building meeting

The schedule of the proposed online training will be set compatibly with the time zones of the people involved in the activity. A possible date will be planned following users availability.