

Warsaw University of Life Sciences WULS – SGGW  
in Warsaw  
Faculty of Forestry

Eberswalde University for Sustainable Development – HNEE  
University of Applied Sciences  
Faculty of Forest and Environment

Tobias Seydewitz  
Album number SGGW: 178311  
Album number HNEE: 15210024

# Kompleksowa analiza wylesiania w krajach tropikalnych - bezpośrednie czynniki wylesiania, emisje dwutlenku węgla i równowaga wartości usług ekosystemów

A comprehensive study on deforestation in the tropics - direct deforestation drivers, carbon emissions and ecosystem service value balance

Master's Thesis  
on the course of - Forestry

Thesis written under the supervision of  
Dr. Prajal Pradhan  
Potsdam Institute of Climate Impact Research  
Research Domain II - Climate Climate Impacts & Vulnerabilities

Potsdam, 2019



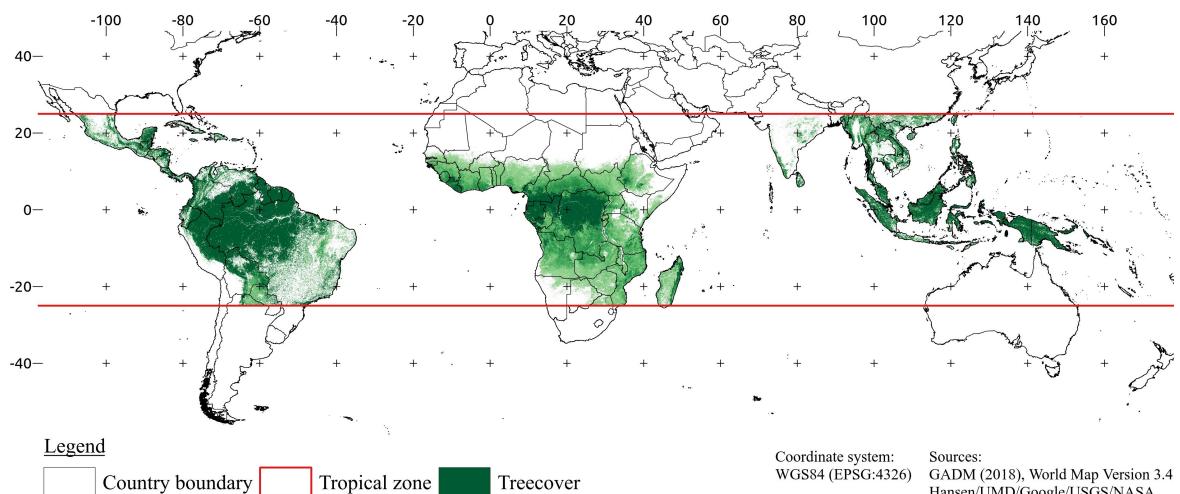
# Contents

1. Introduction	11
2. Data and methods	15
2.1. Data	15
2.1.1. Global Forest Change	16
2.1.2. GlobeLand30	17
2.1.3. Aboveground Woody Biomass	19
2.1.4. Intact Forest Landscapes	20
2.1.5. Global Soil Organic Carbon Map	21
2.1.6. Soil Organic Carbon change	22
2.1.7. Ecosystem Service Values	23
2.2. Methods	23
2.2.1. Preprocessing	23
2.2.2. Deforestation	26
2.2.2.1. Forest definition	26
2.2.2.2. Proximate deforestation drivers	28
2.2.2.3. Accuracy assessment	30
2.2.3. Emissions	33
2.2.4. Ecosystem service values	35
2.2.5. Binning analysis and visualization	36
3. Results	41
3.1. Deforestation	41
3.1.1. Forest definition	41
3.1.2. Patterns of tree cover and deforestation	47
3.1.3. Proximate deforestation drivers	54
3.1.4. Accuracy assessment	58
3.2. Emissions	60
3.3. Ecosystem service values	61
4. Discussion	66
4.1. Deforestation	66
4.1.1. Forest definition	66
4.1.2. Tree cover and deforestation	67
4.1.3. Proximate deforestation driver	67
4.1.4. Accuracy assessment	67
4.2. Emissions	68
4.3. Ecosystem service values	68

4.4. Binning analysis and visualization . . . . .	68
5. Conclusion	69
Bibliography	70
List of Figures	I
List of Tables	II
List of Abbreviations	III
A. Hexagonal country boundaries	V
B. Forest definition	VI
C. Proximate deforestation drivers	X

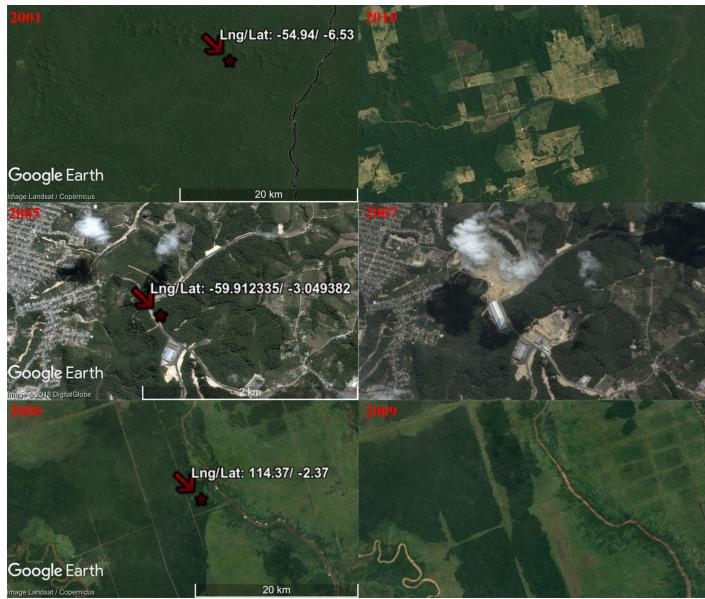
# 1. Introduction

Forest covers approximately 30% of the global land surface, while approximately 50% of the global forest area is covered by tropical forest [WWF 2016]. Tropical forests covering the global land surface between 23° north (the tropic of Cancer) and 23° south (the tropic of Capricorn) around the equator as figure 1.1 shows. This zone comprises several different forest types like tropical rain forest, tropical moist forest, tropical dry forest, and tropical mountain system forest. The distribution and extent of this forest types is affected by the following environmental properties: precipitation levels and seasonality, temperature profile, elevation, and soil characteristics. Rain forest account for approximately 60% of the tropical forest cover and is characterized by warm and wet climate, while the average annual temperature is above 20 degrees and the precipitation ranges between 2500 and 4500 mm  $y^{-1}$ . The canopy is dense and dominated by trees with a height between 30 and 45 m. The required climate conditions are typically found between 10° north and 10° south. Tropical forest is a crucial part of our earth system as important factor for climate, water cycle, soil health, biodiversity storage, and source of natural goods for human subsistence. These forests are a major carbon sink and stimulate precipitation. Further, these forests house the largest diversity of species and a large amount of endemic species. However, this unique ecosystem is endangered by forest loss and transitions to other Land Use/Land Cover (LULC) types [WWF 2016].



**Figure 1.1. Tropical zone and forest**

Since the 1990s approximately 3.1% of the global forest area is lost, while approximately 35% of the tree cover loss is within tropical forest cover [FAO 2016]. Between 1990 and 2015 all the top ten countries with the highest annual net loss of forest are located in the tropical zone [FAO 2010, 2016]. Forest loss or deforestation is defined as the removal of trees or forest stands from land surface which is subsequently converted to another LULC. Geist and Lambin [2001] introduced the conceptual framework of proximate, underlying, and other causes of tropical deforestation to classify the drivers of forest cover change. Direct deforestation driver or proximate deforestation driver are anthropogenic actions that change forest to other Land Cover (LC) types. Proximate causes of deforestation are grouped into three different categories: agricultural expansion, wood extraction, and infrastructure expansion. Examples for deforestation driven by agriculture are the expansion of pastures, cropland, and tree crops as shown in figure 1.2. Examples for wood extraction and infrastructure expansion would be the following categories: fuelwood extraction, charcoal production, transport infrastructure, and settlement expansion. Underlying causes are complex social, political, economic, technological, and cultural interactions, which underpin Proximate Deforestation Drivers (PDDs). The group of other causes for deforestation comprises land characteristics, biophysical drivers, and social trigger events. An example for the land characteristics is soil quality and topography of the landscape, which determines the shape of the deforestation. Scientific studies of PDDs at a global, continental or regional range are common in science (e.g. Curtis et al. [2018]; Hosonuma et al. [2012]; Sy et al. [2015]; Austin et al. [2019]; Boucher et al. [2011]; DeFries et al. [2010]; Zalles et al. [2018]; Carter et al. [2018]; Ickowitz et al. [2015]; Meyfroidt et al. [2013]; Caldas et al. [2013]). All studies suggest that agriculture expansion is the main cause for tropical deforestation. However, each of the beforehand mentioned studies tries to predict the PDDs by a varying methodology. Some studies using Food and Agriculture Organization of the United Nations (FAO) country-based data or case studies for certain areas and an empirical approach to predict PDDs (e.g. Hosonuma et al. [2012]; Sy et al. [2015]), while other studies uses sample-based methods combined with statistical models and visual interpretation of remotely sensed data from different sources (e.g. Hosonuma et al. [2012]; Sy et al. [2015]; Austin et al. [2019]; Curtis et al. [2018]). Additionally, some studies rely only on remote sensing of the environment to predict the proximate causes of deforestation (e.g. Caldas et al. [2013]). On the fact that these studies need a vast amount of expert knowledge and in some cases repetition of time consuming processes they are hard repeatable on an annual basis. Further, the studies that estimate the PDDs spatially explicit are only in low resolution available [Curtis et al. 2018; Caldas et al. 2013]. Additionally, all studies more or less present the proximate causes on deforestation already as aggregates of the three groups Geist and Lambin suggest. Therefore, it is not possible to derive regional or continental differences between the patterns of cropland and pasture expansion.



**Figure 1.2. Examples of proximate deforestation drivers:** From top to bottom and left to right, pasture expansion in Brazilian rain forest, urban expansion in Brazil, and tree crop expansion in Indonesia.

**Emissions:** Deforestation not only changes land cover and drives the loss of biodiversity it also releases emissions. Some numbers. Mention the sources of emissions. No the gaps that agb estimates are common but soil organic carbon emissions are not estimated till no.

Tropical forest ecosystems have a crucial impact on the well-being and subsistence of current and future generation out of humanity through the provision of regulatory, supporting, provisioning, and cultural services [Costanza et al. 1997]. On the fact that deforestation and LC changes lead to major changes in ecosystem services by altering the shape of forest biomes it is crucial to evaluate these impacts not only in terms of Greenhouse gas (GHG) emissions but also for key ecosystem services as water, regulation, biodiversity etc. For the quantification of these ecosystem services a economic process is applied to assess the monetary value of each service per ecosystem. These Ecosystem Service Values (ESVs) can be a strong tool to determine the impact of certain management practices on ecosystem structures. Till now several studies prepared estimates of the ESV loss by tropical deforestation by applying the global coefficients of Costanza et al. [Song 2018; Costanza et al. 2014]. Further, several studies tried to estimate the ESV changes by LULC change dynamics on global and regional scale [Costanza et al. 1997; Sannigrahi et al. 2018; Wang et al. 2006; Kreuter et al. 2001; Zhao et al. 2004]. Additionally, de Groot et al. [2012] prepared a study on global ESV dynamics by introducing alternative coefficients. To best of our knowledge no studies tried to determine the impact of PDDs on tropical forest cover in regards of ESV change dynamics. ESV change dynamics are defined by us as the ESV loss by tropical deforestation, the ESV gain by newly introduced LC on former forested areas, and the ESV net balance between both dynamics.

**Research questions:** Our goal is to develop an approach to determine spatially explicit the PDDs of tropical tree cover loss at a high spatial resolution. Our approach should meet the following criteria: high spatial resolution to consider for example small-holder deforestation; an easy accessible approach that can be reproduced without constraints; can be repeated by certain time steps in future. To achieve these goal we will combine the information from the most recent state of the art LC datasets Global Forest Change (GFC) and GlobeLand30 (GL30).

Our goal is to evaluate these dynamics by applying the most common in literature used ESV datasets of Costanza et al. and de Groot et al. at a global and continental scale between 2000 and 2010. Additionally we want to include the ESV for tropical forest by the recent study Siikamaki et al. [2015] to discuss differences between the three datasets. To compute the beforehand mentioned ESV dynamics we will derive LC change areas from our analysis on PDDs in the tropical zone.

## 2. Data and methods

In this chapter we describe our approach to answer the scientific questions stated during the introduction. The first section of this chapter introduces the datasets used during this study and the underlying causes for using them. For each dataset, we shortly describe by which approach it is derived and what are the fundamental meta-data properties. Additionally, if possible we try to give for each dataset an accuracy assessment, ideally provided by other research groups if available. Finally, we describe our idea behind using the data and how we acquired and filtered it. The second and last section of this chapter is focused on the applied methodology to prepare our analysis and results. For each processing step we give a short description of the methodical background and describe the core functionality of our processing algorithms. For implementing our processing algorithms and visualizing our results we selected individually the programming language or software which fulfills best the requirements. These approaches are encapsulated in a reusable software design to easily reproduce, alter or reuse our algorithms and findings.

### 2.1. Data

**Table 2.1. Datasets used in this study**

Data	Type	Source
Global Forest Change	spatial	Hansen et al. [2013]
GlobeLand30	spatial	Chen et al. [2015]
Aboveground Woody Biomass	spatial	Baccini et al. [2015]
Intact Forest Landscape	spatial	Potapov et al. [2017]
Global Soil Organic Carbon Content	spatial	FAO and ITPS [2018]
Global Administrative Areas	spatial	Hijmans et al. [2018]
Soil Organic Carbon Change	empirical	Don et al. [2010] Costanza et al. [2014]
Ecosystem Service Values	empirical	de Groot et al. [2012] Siikamaki et al. [2015]

The table 2.1 shows a comprehensive overview of the applied datasets for this study. Spatial datasets comprises vector as well raster data, while empirical data is extracted from the cited publications. The subsequent sections describe each dataset in further detail.

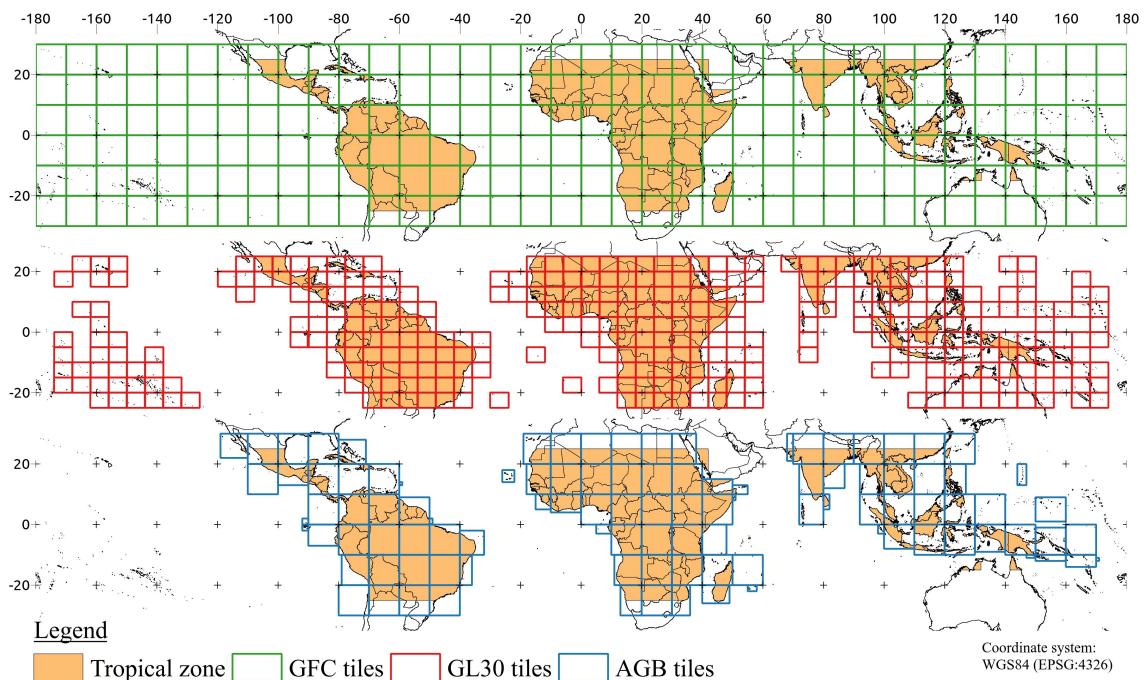
### 2.1.1. Global Forest Change

GFC 2000-2012 Version 1.0 is the first high-resolution dataset that provides a comprehensive view of the annual global forest cover change between 2000 and 2012 [Hansen et al. 2013; Li et al. 2017]. We will use this dataset to extract and determine the tropical deforestation and reforestation dynamics for our study time frame from 2000 till 2010. The initial GFC dataset released by Hansen et al. has been extended by recent releases, which encompass the annual forest cover changes between 2000-2013, 2000-2014, 2000-2015 and 2000-2016, respectively. All versions of this dataset are derived from growing-season imagery captured by the following remote sensing satellites: Landsat 7 ETM+, Quickbird, MODIS [Hansen et al. 2013]. On the satellite imagery, a time-series spectral metrics analysis is applied to gather the global forest extent in 2000 as well as the annual forest loss and the accumulated gain for the period 2001-2012. Hence, GFC comprises three independent data layers: tree cover, annual forest loss, and forest gain. Each of these layers is divided into 10x10 degree tiles by the Coordinate Reference System (CRS) World Geodetic System 1984 (WGS84) (EPSG:4326) with a spatial resolution of 1 arc-second per pixel or 30 m per pixel. Further, across the provided Geo-Tiff (GTiff) layers the pixel data is coded in unsigned 8-bit integers. Hansen et al. defined trees as all vegetation taller than 5 meters. For each pixel covered by trees, a canopy density ranging from 0 to 100% is computed. Forest loss is defined as a stand displacement disturbance leading from a forest state to a non-forest state (e.g. canopy density >50% to 0). Resulting from this definition, the underlying causes of forest loss range from anthropogenic impacts to natural causes. Tree cover gain is defined as the inverse of loss, and the canopy density must exceed 50% to get recognized.

Hansen et al. [2013] reports a tree cover loss accuracy assessment of 83% for the tropical region. The mapped tree cover gain achieves a producers accuracy of approximately 48%, while the users accuracy is approximately 81%. The large difference between users and producers accuracy highlights that tree cover gain is underestimated by the algorithm. For a part of the Riau province in Indonesia Arjasakusuma et al. [2018] reports for the gain layer a producers and users accuracy of  $64.5\%\pm7\%$  and  $75.9\%\pm8.1\%$ , respectively. By this independent validation the underestimate of tree cover gain is largely confirmed for this region. For Gabon Sannier et al. [2016] reports an accuracy assessment for the GFC tree cover map an overall accuracy of 95.87% and 96.6% for the canopy density thresholds of >30% and >70%, respectively. In subtropical areas the GFC dataset achieves an overall accuracy of 62%, 64%, 66% for the canopy density thresholds of >10%, >30%, and >50% [McRoberts et al. 2016].

This dataset is publicly available for download without any constraint. For a convenient bulk download, the dataset homepage provides a ".\*.txt files" comprising the Uniform Resource Locator (URL) of the tiles for each sub-dataset. The spatial location of an image

can be directly determined from the file name within the URL. Each file name has a common pattern shown by the following expression: "Hansen\_VERSION\_LAT[NS]\_LNG[WE]". LAT (latitude) and LNG (longitude) refer to the top left corner coordinates of a raster image, whereas these coordinates are only given in natural numbers. The orientation of the image on the hemisphere is determined by the four cardinal directions N (north), S (south), W (west) and E (east). For this project, we require all three sub-datasets, namely: Treecover2000, loss-year, and gain. The data acquisition is automatized with a Python script by using the Standard Library (stdlib) modules `urllib` and `re`. At first, the Python script downloads the provided "`*.txt`" files and creates a list data structure, where each URL is an element of this list. After, it cycles through the list and extracts the corner coordinates from the file name by means of a Regular Expression (REGEX). These corner coordinates and cardinal directions are converted to valid latitude and longitude coordinates between  $[-90, 90]$  and  $[-180, 180]$ , respectively. Now, an image is only downloaded if it is within the study extent between  $[-20, 30]$  latitude. The acquired image tiles in total 678 are shown in the top panel (green squares) in figure 2.1.



**Figure 2.1. Map of downloaded dataset tiles:** This map shows the acquired image tiles used for this study. From top to bottom Global Forest Change (GFC) dataset tiles (Treecover2000, loss-year and gain) (green), the land cover dataset GlobeLand30 (GL30) tiles (red), and the Aboveground Biomass (AGB) dataset tiles (blue). The orange filled shapes highlight countries within the tropical zone.

### 2.1.2. GlobeLand30

GL30 is the first global land cover dataset with 30 meter spatial resolution that provides a comprehensive view on the distribution of 10 different land cover classes (table 2.2) over

the entire globe [Chen et al. 2017]. We will use this dataset to classify tree cover loss and subsequently derive the PDDs of tropical forest. Currently, this dataset is available for two different time steps 2000 and 2010 [Chen et al. 2015]. The pixel values of this dataset are coded in unsigned 8-bit integers and as CRS, it uses WGS84 in Universal Transverse Mercator (UTM) projection. GL30 can be downloaded as a GTiff raster mosaic where each image covers 6x5 degrees [Chen et al. 2014]. For detecting the land cover classes Chen et al. used a so-called Pixel-Object-Knowledge oriented approach and satellite imagery from Landsat ETM+ [Chen et al. 2015]. The mapping process was divided into different stages where each land cover type is detected separately and deleted subsequently from the source image. The applied mapping order is the following: water bodies, wetland, snow and ice, cultivated land and forest, shrubland, grassland and bare land synchronous. To detect the pixels of a selected land cover type the following pixel-level classifiers are used: Decision Trees, Support Vector Machines or Maximum Likelihood Classifier. After pixel detection, the adjacent pixels are grouped as an aggregated land use object. These objects are subsequently validated by expert knowledge and the gained knowledge is used as a feedback loop to improve the automatized classification.

Chen et al. [2015] estimates an overall mapping accuracy of 80.33 % and 78.6 % for 2000 (only validated in Shaanxi, China) and 2010 (global), respectively. Several research groups besides Chen et al. validated the mapping accuracy of GL30 at different regions and scales. Arsanjani et al. estimates an overall accuracy of 77.9 % for Iran and an accuracy of >80 % is reported for Germany [Arsanjani et al. 2016a,b]. Yang et al. [2017], Cao et al. [2016] and Jacobson et al. [2015] estimate the accuracy of 82.4 %, 80.1 % and 83.1 % for China, Nepal, and East Africa, respectively. To the best of our knowledge, no study have focused on validating the mapping accuracy for regions exclusively within the tropical zone.

Chen et al. donated the GL30 land cover mapping to the United Nations (UN) but it is not accessible for public download unless the user registers on the dataset homepage. A registered user must fill an order application to get access to the image tiles. The application form must contain the tile identifiers and the selected time period. Tile identifiers have the following common pattern: "[NS]ZONE\_LAT\_NAME" where zone refers to the UTM zone between [1,60], N (north) or S (south) to the cardinal direction, and LAT (latitude) to the latitude coordinate of the top left corner. On the dataset homepage a vector file can be downloaded which contains the dataset tile polygons with assigned identifiers. This file was used to select all required tiles within the tropical zone between approximately [-23,23] degrees (WGS84). Figure 2.1 presents the selected images (middle panel-red). The corresponding image identifiers are converted to a single line string and copied to the application form. After submitting the form the order will be checked and approved within two weeks. After one week we received a two weeks limited access to a password protected FTP-server where we downloaded 716 raster images. Due to the several restrictions, this process of selecting

and downloading could not be automatized with one pipeline. Only the selection and string conversion were automatized with a throwaway script.

**Table 2.2. Land cover classification of the GlobeLand30 product:** The code column is the assigned pixel value, type refers to the corresponding land cover type and definition explains in broad terms which types of surfaces fall into each land cover type [Chen et al. 2017].

Code	Type	Definition
10	Cultivated land	Used for agriculture, horticulture and gardens, including paddy fields, irrigated and dry farmland, vegetable and fruit gardens, etc.
20	Forest	Covered by trees, vegetation covers over 30%, including deciduous and coniferous forest, and sparse woodland with cover 10-30%, etc.
30	Grassland	Covered by natural grass with cover over 10%, etc.
40	Shrubland	Covered by shrubs with cover over 30%, including deciduous and evergreen shrubs, and desert steppe with cover over 10%, etc.
50	Wetland	Covered by wetland plants and water bodies, including inland marsh, lake marsh, river floodplain wetland, forest/shrub wetland, peat bogs, mangrove and salt marsh, etc.
60	Water bodies	In land area, including river, lake, reservoir, fish pond, etc.
70	Tundra	Covered by lichen, moss, hardy perennial herb and shrubs in the polar regions, including shrub-, herbaceous-, wet- and barren-tundra, etc.
80	Artificial surfaces	Modified by anthropogenic influence, including all kinds of habitation, industrial and mining area, transportation facilities, and interior urban green zones and water bodies, etc.
90	Bareland	With vegetation cover lower 10%, including desert, sandy fields, Gobi, bare rocks, saline and alkaline land, etc.
100	Snow and ice	Covered by permanent snow, glacier and icecap

### 2.1.3. Aboveground Woody Biomass

The Aboveground live woody Biomass density (AGB) raster dataset is prepared by Global Forest Watch (GFW) by an adapted approach of Baccini et al. [2012, 2015, 2017]. We will use this raster dataset to determine AGB emissions by PDDs of the tropical tree cover. For the year 2000, this dataset estimates the aboveground biomass density per pixel in Mg C ha<sup>-1</sup> (megagram carbon per hectare), and the confidence per pixel at a spatial resolution of approximately 1 arc-second or 30 m. The dataset covers the global tropical zone as a mosaic of GTiff raster images, where each tile of the mosaic has the CRS WGS84 and is coded in a float. For deriving biomass density GFW used canopy metrics from Geoscience Laser Altimeter System (GLAS) Light Detection and Ranging (LIDAR) footprints and several regional and forest-specific allometric equations. The resulting GLAS AGB estimates

are used as labels to train regional-specific random forest models based on Landsat 7 ETM+ top-of-atmosphere reflectance, tree canopy density of GFC, elevation data, and climate data as predictor variables. Next, these models are subsequently applied to the entire study extent to predict the biomass content for each pixel. Additionally, an uncertainty layer is prepared accounting for the errors from allometric equations, the LIDAR-based model, and the random forest model.

The AGB raster mosaic is publicly available on the homepage of GFW. As mentioned, the dataset covers only the tropical zone, therefore we acquire the entire mosaic. The GFW homepage provides an Application Programming Interface (API) to receive the actual URL of each raster image. If a request is sent to this API the server responded with a Geographic JavaScript Object Notation (GeoJSON) feature collection. The collection contains as attributes the URLs of the biomass images, the URLs of the uncertainty layers, and the rectangular bounds of each image. The data acquisition is automatized by means of Python and the stdlib modules urllib, threading, and the open source library geopandas [van Rossum and Development 2018; McKinney 2010]. At first, the GeoJSON is downloaded via an API call and eventually stored on disk. Next, we iterate the features of the GeoJSON collection and extract the URLs (biomass and uncertainty) of each tile. These URLs are downloaded and subsequently stored on disk. During the downloads of the uncertainty layers, the GFW server answered repeatedly with a 404 (Not found). Therefore the uncertainty layers are not available. In total we downloaded 105 different image tiles, their extent and spatial location are shown in blue at the bottom panel of figure 2.1.

#### 2.1.4. Intact Forest Landscapes

An Intact Forest Landscapes (IFL) is defined as a mosaic of undisturbed forest patches or a naturally treeless ecosystems without signs of human activity and large enough to maintain all native biological diversity [Potapov et al. 2017]. We will use this dataset to distinguish between deforestation in primary or secondary forest to decide which Soil Organic Carbon (SOC) change coefficient from section 2.1.6 will be selected. On the fact that IFL comprises different intact natural landscape patterns like primary forests, non-forest ecosystems, temporary treeless areas after a natural disturbance, and water bodies the term is not congruent to the term primary forest defined by the FAO [FAO 2012]. However, as mentioned IFLs includes large patches of primary forests with a minimum extent of 500 Km<sup>2</sup>, therefore, primary forests can be extracted from the layer. Still, there are smaller fragments of primary forest outside of the IFLs. In regards to the extent, an IFL has a minimum size of 500 Km<sup>2</sup>, a minimum width of 10 Km, and a minimum corridor/appendage width of 2 Km. Further an IFL should not contain any of the following: ecosystem alternation, fragmentation by infrastructure and disturbance, and areas altered or managed through agriculture,

logging, and mining. For mapping and detecting IFLs Potapov et al. used Landsat imagery and several auxiliary data sources like GFC, and national transportation maps. The dataset can be downloaded as a Shapefile (SHP) file with the coordinate reference system WGS84. Each polygon in the SHP represents an IFL patch at a certain location on our planet at the time period 2000.

Data acquisition is pretty straight forward, since the IFL dataset is publicly accessible for download. On the fact that the layer is a single SHP only a single compressed archive must be downloaded, while the download is automatized with a Python script by using the stdlib modules urllib and threading [van Rossum and Development 2018].

### 2.1.5. Global Soil Organic Carbon Map

The Global Soil Organic Carbon map (GSOCmap) is a joint project between Global Soil Partnership (GSP) and Intergovernmental Technical Panel on Soils (ITPS) to produce a global SOC content map by a country-driven approach. We will use this dataset to determine the SOC content at deforested spots and to derive the SOC emissions by tropical tree cover loss. Until now 67 (approximately 63% of the global land mass) different countries have submitted their country-based SOC estimates. To foster the national SOC mappings the International Soil Reference and Information Center (ISRIC) provides several covariate datasets like national digital elevation maps, annual spectral remote sensing data or national soil type grids. Additionally, the contributors can join a mapping training and use the GSOCmap cookbook as guidance for their mapping efforts. As an exchange, each country shares its national GSOCmap by compliance of several criteria e.g. reporting of the metadata of the SOC sampling (sample timeline, sample depth, bulk density etc.), uncertainty assessment, and the applied methods for the estimation and interpolation of the SOC content. For interpolating, the leading organizations suggest the following approaches: simple geo-matching, class-matching, multiple linear regression, random forest or support vector machines. The national maps are aggregated to the final GSOCmap with a target resolution of 30 arc-seconds (approximately 1 km<sup>2</sup>) in the CRS WGS84. The dataset is one single raster image as GTiff coded in float covering the entire globe, where each pixel value is the SOC content in Mg C ha<sup>-1</sup> at a soil depth of 0-30 cm [FAO and ITPS 2018].

The product is validated by comparing the pixel level estimates with soil sampling data from various soil databases (WoSIS, HWSD, etc.). In total 312 122 samples were divided into three sub-levels (<150 Mg C ha<sup>-1</sup>, >150 Mg C ha<sup>-1</sup>, and the entire stratum) and the mean errors were computed . The mean error of the entire sample space and of the <150 Mg C ha<sup>-1</sup> suggests that the mean SOC content value is an overestimate of 1.6 and 4.5 Mg C ha<sup>-1</sup>, respectively. All samples with a Soil Organic Carbon Content (SOCC) content >150 Mg C

$\text{ha}^{-1}$  show an underestimation of approximately  $165 \text{ Mg C ha}^{-1}$  in the mean. In comparison with other global SOC products, the GSOCmap has the lowest root mean square error. In summary, the prepared validations show evidences that the GSOCmap is a conservative data product with a tendency to underestimate the SOC content.

The dataset is publicly available at the homepage of the FAO. As mentioned it consists of one raster image, therefore we download it by means of a Python script without any additional steps.

### 2.1.6. Soil Organic Carbon change

Don et al. [2010] performed the first study of tropical SOC changes resulting from Land Use (LU) change for a soil depth between 0 and 30 cm. We will use the empirical data to determine the cumulative SOC emissions by LC transitions through PDDs. For the study, a global meta-analysis is applied by using 358 (153 published an peer-reviewed) different studies to estimate SOC change for 12 major LU change types. The base data is derived from 39 different tropical countries covering all continents. All regions are not equally covered by the studies included in the meta-analysis: Africa and East-Asia are under-sampled, while South America has the best data coverage. The meta-analysis is restricted to mineral soils, therefore all wet soil types are excluded from the analysis. The 12 LU transitions encompass the following LU types: primary forest, secondary forest, grassland, cropland, and perennial crops. Primary forest is defined as natural vegetation without human impacts, which includes natural grassland and shrubland. Secondary forest represents managed forests and regrown forests after partial destruction of the old natural stand. Grassland comprises pastures for livestock but excludes natural grasslands. Cropland comprises annual crops like maize or beans, while perennial crop examples could be coffee or sugar cane. For our study we used only the SOC change estimates for these LU types which correspond to the GL30 and IFL classification scheme table 2.3.

**Table 2.3. Relative soil organic carbon change for certain land-use change types:** The land use change (LUC) columns from and to define the land use change type with the corresponding relative soil organic carbon (SOC) change and the standard error of the mean (SEM) [Don et al. 2010].

LUC type From→To	Relative SOC change	
	[%]	SEM [%]
Primary forest→Grassland	-12.1	±2.3
Primary forest→Cropland	-25.2	±3.3
Primary forest→Secondary forest	-8.6	±2.0
Secondary forest→Grassland	-6.4	±2.5
Secondary forest→Cropland	-21.3	±4.1

### 2.1.7. Ecosystem Service Values

We will use the ESV coefficients for different biomes to compute the following ESV dynamics for tropical forest cover change: ESV loss, gain, and net balance.

**Table 2.4. Ecosystem service values (ESV) per biome used in this study:** ESV per biome and its monetary value in 2007 Int'l\$  $y^{-1}$   $ha^{-1}$  also known as Geary-Khamis Dollar. Co refers to data from Costanza et al. [2014], Dg from de Groot et al. [2012], and Wb from Siikamaki et al. [2015].

Biome	Co	Dg	Wb
Cropland	5567	-	-
Tropical forest	5382	5264	1312
Grass/Rangelands	4166	2871	-
Urban	6661	-	-

## 2.2. Methods

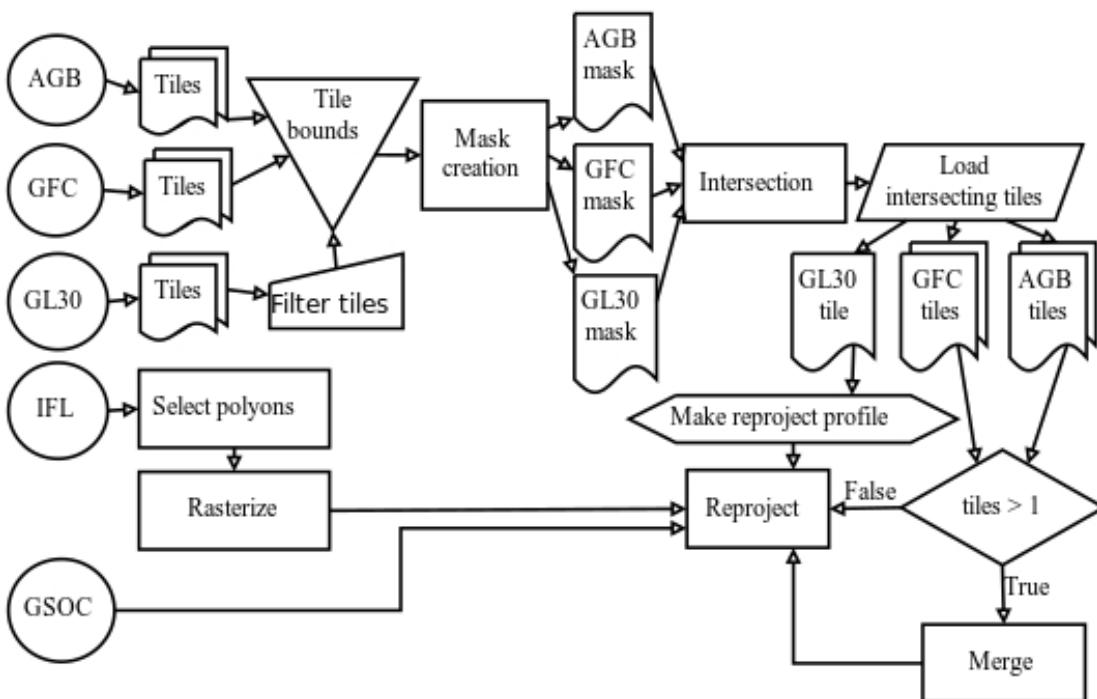
Figure [flowchart and reference](#) shows an overview of the entire processing pipeline. The following sections describe detailed the applied approach for each step in figure. The order of appearance is from left to right.

For implementing our processing algorithms we selected for each task the technology which fulfills best the requirements. Python is our core language for implementing our processing algorithms because it supports an easy implementation of multiprocessing which is heavily used in this project. From the Python stdlib we used the following libraries: urllib, re, unittest, time, math, logging, collections, bisect, and enum. Additionally for geo-processing we used the following open source libraries: numpy, pandas, fiona, geopandas, shapely, matplotlib. The entire frontend of our Python source code is aggregated in a Jupyter Notebook and available on GITHUB. This ensures that everyone interested can easily reproduce the findings of our project. JavaScript and the additional modules papaparse and the google maps api are used for programming a small web app for cross validation of land use predictions. R is used for hypothesis testing. Bash is used to aggregate large raster datasets as vrt files. To prepare map visualizations we used QGIS. Dia is used for preparing flowcharts and GIMP is used for image post-processing.

### 2.2.1. Preprocessing

Before we apply further analysis, we have to harmonize the used datasets. As introduced in the data section we use datasets which differ largely in their metadata properties, for example,

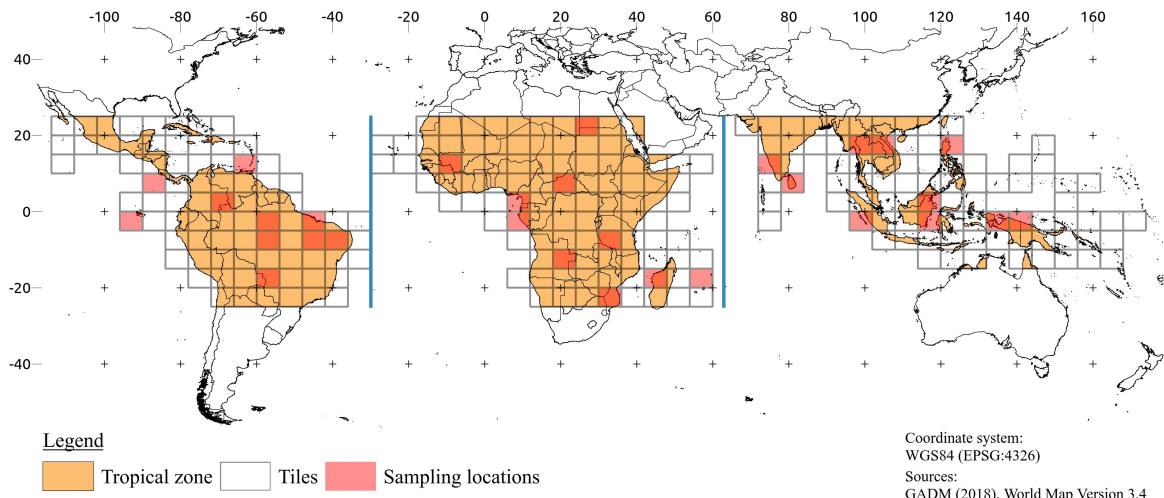
single-tiled or multi-tiled images, used CRS, spatial resolution, and file type. Therefore, our goal should be to develop a process which creates an image stack of equal meta-data for each location in our study extent. In further descriptions, we will refer to this stack as Aligned Image Stack Mosaic (AISM). As target CRS for our AISM we chose WGS84, and as target extent for the mosaic we use the bounding box of the GL30-2010 tiles. The following paragraph explains how we developed the alignment algorithm by means of Python and the additional open source libraries rasterio, geopandas, and shapely [van Rossum and Development 2018; McKinney 2010]. The figure 2.2 shows the applied steps to harmonize the different raster and vector datasets.



**Figure 2.2. Raster and vector harmonization process:** For the multi-tiled datasets represented by the multi-document symbols a mask is created by extracting the tile bounds. Next, the intersection between these masks is determined to identify superimposing data and the corresponding tiles are loaded from the disk. GlobeLand30 (GL30) tiles are used as a template by creating the re-project profile and subsequently applying it to the intersecting tiles. From the Intact Forest Landscapes (IFL) layer only polygons within the re-project area are selected and subsequently converted to a raster layer. The Global Soil Organic Carbon Map as a single tile raster file with a spatial resolution of 1 km<sup>2</sup> is re-projected and re-sampled by the nearest-neighbor approach.

The first exercise of the preprocessing algorithm is to detect all tiles covering the extent of our template tiles. First, we create for each multi-tiled dataset a polygon mask as SHP. This mask contains the spatial extent of each tile within a dataset and as attribute the corresponding file identifier. If the dataset tiles are not in WGS84 the extracted bounds are subsequently reprojected to this CRS. During the masking process, we recognized that the raster mosaic bounds of both GL30 datasets (2000 and 2010) generate re-projection errors. Further analysis revealed that all tiles located in UTM zone 1 and 60 overflowed the maximum and minimum longitude coordinates of these zones. To solve this we excluded all

tiles within UTM zone 1 and 60 from further processing. Now, as the figure 2.2 suggests we determine the intersection between these mask layers and group the intersecting tiles by our template tile. Next, we create for the template tile a re-projection profile (warp profile) and apply it subsequently to all intersecting tiles based on the following rules: if from one dataset more than one tile intersects merge them followed by re-projection; if only one tile intersects just re-project it. As introduced, the GSOCmap consists only of one single tile with a spatial resolution of approximately  $1 \text{ Km}^2$ , so it must only be re-project and re-sampled by the nearest-neighbor approach. We select from the IFL layer all polygons within our template warp profile and convert them to a raster layer where intact forest patches are coded by a one in an 8-bit unsigned integer. The last step of the alignment process is the rounding of the AISM bounds to full integer degrees and a subsequent clipping of each tile to this rounded bounds. Finally, we create a polygon mask of our AISM and store for each polygon as attributes the corresponding dataset tiles. This mask is used as a file index for the next algorithms. The figure 2.3 shows this mask and the extent of the harmonized dataset tiles. Each box in this figure highlights a raster tile stack of our AISM. The advantage of this tiled approach is that we don't have to store 8 large raster files which cover the entire tropical zone and occupy a large amount of disk space and exceed the available memory if loaded for further processing. Further, this approach enables us to parallelize most of our further computations because each tile from our AISM is a closed unit.



**Figure 2.3. Harmonized raster images and sampling locations:** The map shows the location of the aligned multi-image stack tiles as black-framed, square-sized polygons. The sampling locations for accuracy assessment are represented in red. Countries within tropical range appear in orange. Vertical blue lines separate the tiles into the three continental regions Latin America, Africa, and Asia/Australia.

## 2.2.2. Deforestation

### 2.2.2.1. Forest definition

To determine the proximate drivers of deforestation we combined the information of the two datasets GFC and GL30. However, both differ in their definition of tree cover by canopy cover threshold as introduced in section 2.1. GFC detects tree cover over the entire canopy density interval of  $(0, 100]$ , while the GL30 threshold is set to  $> 10\%$ . To successfully extract stable land cover transformation by superimposing both layers we must first harmonize the tree cover definition of both strata. We hypothesize that if both layers agree on tree cover they should also agree if a transition to a non-forest state occurs. To harmonize both definitions we have the opportunity to vary the canopy density of GFC to determine at which density class the similarity between them is at its maximum. Then, we use the examined maximum similarity canopy density to filter the tree cover loss and gain layer.

To determine the similarity between GL30 2000 and GFC reference tree cover we used the Jaccard Index (JI). The JI or coefficient of community is a simple measure of similarity between two pairs of a binary population or a measure of the degree of spatial overlap between two images [Sampat et al. 2009]. This index was first applied by Jaccard to compare distributions of rare alpine flora in 1912 [Jaccard 1912], and since it is a widely used metric across multiple fields. If we compare two binary images, let  $a$  be the magnitude where both images ( $\text{Img}_1, \text{Img}_2$ ) have an agreement represented as a pixel value of one. Let  $b$  the magnitude where  $\text{Img}_1$  is zero and  $\text{Img}_2$  is one, while  $c$  represents the inverse expression. Finally, assume that  $d$  is the magnitude of elements where both images are zero. The matrix in table 2.5 shows that the computation of this coefficients  $a, b, c$ , and  $d$  can be expressed as a set of boolean operations. Equation 2.1 shows how the JI is computed by substitute integer values for the variables. This computation can be reduced to two boolean operations for a major performance increase. The JI is always within the closed interval  $[0, 1]$ , where an index of one or zero means a complete similarity between both populations or a complete disagreement, respectively. The relationship between  $a$  and JI is near linear [Shi 1993]. The first step to compute the JI for our raster images is to extract the tree cover from the GL30 2000 land cover by setting all pixels with values  $\neq 20$  to zero and values = 20 to one. Next, we extract from the GFC reference tree-cover pixel values within the half-opened interval of the following canopy density classes and set them to one:  $(0, 100]$ ,  $(10, 100]$ ,  $(20, 100]$ , and  $(30, 100]$ . Therefore, we test four different tree cover definitions for GFC. The first excludes canopy densities  $\leq 0\%$ , the second  $\leq 10\%$ , the third  $\leq 20\%$ , and the fourth canopy densities  $\leq 30\%$ . We will refer to this JI of different canopy density classes as  $\text{JI}_0, \text{JI}_1, \text{JI}_2$ , and  $\text{JI}_3$ . For all 269 tiles of our AISM, we calculate the JI existing between GL30 and GFC for the four above-mentioned forest definitions by using equation 2.1. The algorithm is im-

plemented in Python by using numpy's ability to perform boolean operations between large matrices. As parameters, the function expects two matrices with the same dimensionality in  $R^{n*m}$  and a boolean indicating if the function should return the coefficient matrix as well. The previously described preprocessing steps are implemented as an extra function. This function requires as parameter two raster layers, a list of integer values to consider as GL30 forest cover (default is 20), and the lower bounds of the canopy density intervals to consider for computation.

**Table 2.5. Jaccard Index coefficient matrix:**  $a$  is the magnitude of agreement,  $d$  is the magnitude of disagreement,  $b$  and  $c$  are the magnitudes of partial disagreements among both images. The computation of these coefficients can be expressed as boolean operations on matrices.

		Img <sub>1</sub>	
		State	1      0
Img <sub>2</sub>	1	$a =  \mathbf{X}_1 \wedge \mathbf{X}_2 $	$b =  (\mathbf{X}_1 \wedge \mathbf{X}_2) \oplus \mathbf{X}_2 $
	0	$c =  (\mathbf{X}_1 \wedge \mathbf{X}_2) \oplus \mathbf{X}_1 $	$d =  \neg(\mathbf{X}_1 \vee \mathbf{X}_2) $

$$JI = \frac{a}{a+b+c} = \frac{|\mathbf{X}_1 \wedge \mathbf{X}_2|}{|\mathbf{X}_1 \vee \mathbf{X}_2|} \quad (2.1)$$

To optimize the overall tree cover similarity between both datasets we must test which canopy density class yields the highest agreement over our study extent. To test the significance of the difference between two correlated samples, we decided to apply the non-parametric Wilcoxon signed-rank test [Wilcoxon 1945]. This test requires paired data from the same population, at least an ordinal scale of measurement, each sample pair is independent, and the dependent variable can be expressed as a continuous probability [Lowry 2019]. Further, an advantage of this test is that we don't have to assume a normal distribution for our sample population. Our sample population fulfills these requirements. The test procedure is implemented in R because this language is mainly intended for this kind of statistical analysis. We exported the computed JI from our Python environment and applied a cross-testing in R. In our case, cross testing is defined as the test of all possible JI combinations. Further, we applied a two- and one-sided Wilcoxon test because we want to examine if there is a significant difference and which direction has the similarity distribution. To address the higher probability of family-wise error in multiple comparisons we used a Holm correction. Before we applied the examination of the distribution we separated our population into three independent regions, namely Latin America, Asia/Australia, and Africa highlighted by the vertical blue lines in figure 2.1. Latin America, Asia/Australia, and Africa comprised 82, 86, and 101 image tiles, respectively. Additionally, we excluded from the analysis all samples where  $JI_0$  is zero because this tiles from our AISM did not contain any pixels covered by trees. In Latin America, Asia/Australia, and Africa we excluded 6, 13, and 15 tiles. We tested continental differences in tree cover agreement to inspect regional dependencies and global differences. The results from the global testing are used to determine our definition of

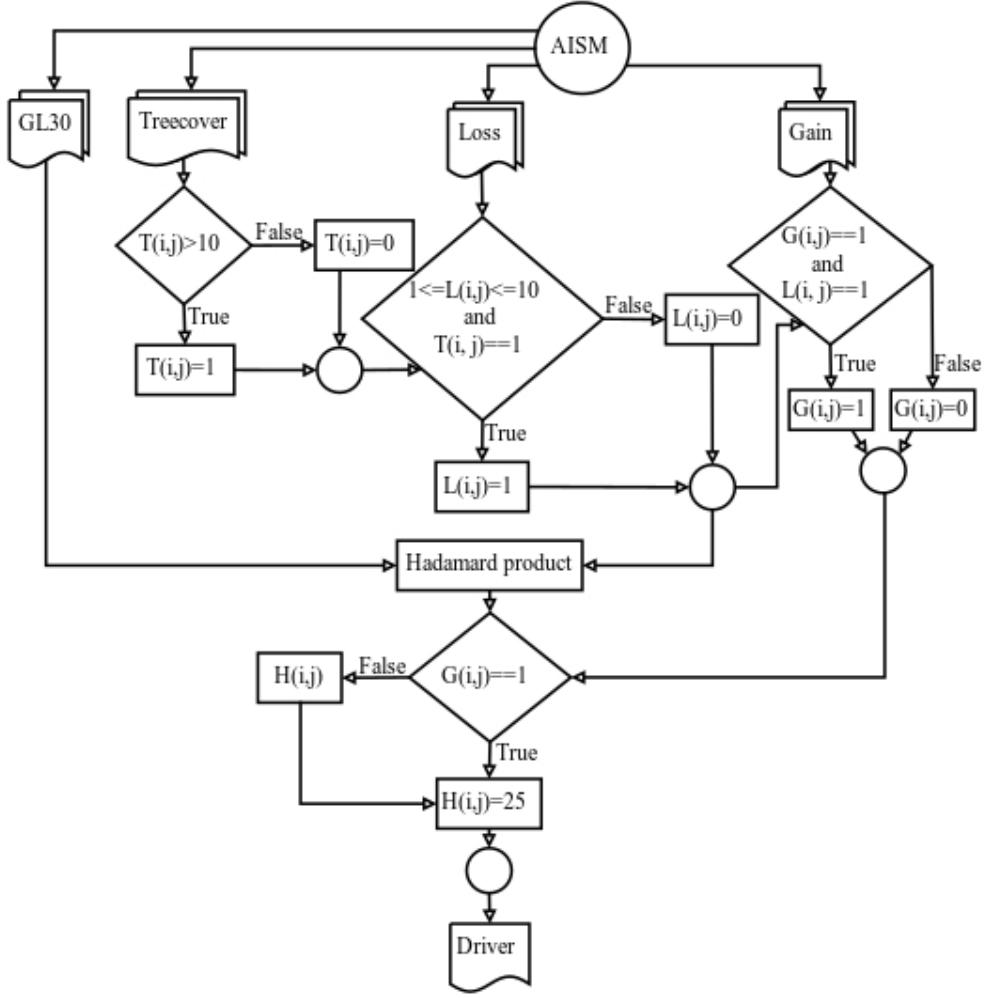
tree cover. Further, we compared differences of tree cover agreement between the continental regions by applying a Wilcoxon rank-sum test also known as Mann-Whitney U test. We applied a Benjamini and Hochberg correction to the test results.

#### 2.2.2.2. Proximate deforestation drivers

Based on our forest definition developed in the previous section we want to classify all the tropical deforestation occurring within a canopy density of  $(10, 100]$  percent between 2001 till 2010.

Figure 2.4 shows an overview of the classification process, which we used to derive the PDDs of tropical tree cover loss. For classifying the PDDs we selected the following raster images from our AIS: GFC tree cover, GFC annual losses, GFC gain, and the GL30 LC classification of 2010. Next, we apply to each raster image stack the following operations. From the reference tree-cover images, we select all pixels where the canopy density is within the half-open interval of  $(10, 100]$  percent and set them to one (true). The same exercise is applied on the annual losses stratum by setting all forest loss pixels within the time period 2001 till 2010 to one (true). After, both layers are combined with a logical AND operation to select our target deforestation pixels. Finally, we classify the pixels with a deforestation event by applying the Hadamard product (element-wise matrix multiplication) on the target deforestation layer and the GL30 LC stratum. This operation basically creates a new image matrix by superimposing both layers. As output we obtain the exact LC category of each deforestation pixel, which allow to understand what LC transitions is driving tree cover loss in different regions. For classifying forest regrowth we filtered the GFC gain layer to consider only tree cover gain within our target temporal resolution and target canopy density. After, the filtered stratum is aggregated with our classified deforestations by using the Hadamard product of both layers. The classification algorithm is implemented as a Python function which requires as parameters the previously named raster layers. Additionally the target canopy density and time period is freely selectable for experimental variations. The described filtering and aggregation steps are implements as binary matrix operations for fast processing of large data sizes by means of numpy.

Our PDDs classification scheme corresponds largely to the LC schema of GL30 in table 2.2. We introduced regrowth (pixel code is 25) as a new LC class from the GFC gain datasets. The LC type regrowth accounts for tree crops like oil palm plantations or forestry activities. Further, this class could be the natural regeneration of tree cover after using the area for other purposes like shifting agriculture. Tree cover loss classified as grassland account for the forest loss by the expansion of pastures for cattle ranching [Graesser et al. 2015]. Tree cover loss classified as wetland and water is forest loss by inundation by lakes and rivers [Sy



**Figure 2.4. Classification of proximate deforestation drivers:** For the classification of the proximate deforestation drivers the following layers are required: GlobeLand30 from 2010, Global Forest Change tree cover, annual losses, and gain. From the tree cover stratum we select all pixels within the canopy density interval  $(10, 100]$ . The tree cover mask is used to select the appropriate annual losses within the time interval [2001, 2010]. To predict a land cover change after a deforestation event we use the Hadamard product (element-wise matrix multiplication). As output we obtain the exact land cover category of each deforestation pixel, which allow to understand what LC transitions is driving tree cover loss in different regions.

et al. 2015]. Forest loss classified as forest by the GL30 layer could relate to false positives (predicts forest loss but there is no loss, type I error) of the GFC loss layer. On the fact that the GFC gain layer has a low overall accuracy and tends to underestimate tree cover gain the probability is higher that this pixels relate to false negatives (predicts no gain but there is gain, typ II error) of GFC gain layer. We will relate to these pixels as miss-classifications that account for a mean miss-classification rate of 52% (with large regional dependencies) if tree cover loss in the entire canopy density interval is considered for classification [Seydewitz 2017]. The next paragraph presents a approach to resolve this issue.

After classifying the proximate deforestation drivers we developed an approach to smooth the misclassified pixels based on LC change probabilities. This means the algorithm tries to

find for clusters of misclassified pixels and to reclassify them by finding the most frequent PDDs in the surroundings within a certain threshold. The first step of our reclassification is to cluster the misclassified pixels with the Hoshen-Kopelman algorithm [Hoshen 1998]. The clustering algorithm is implemented as a part of the Geospatial Data Abstraction Library (GDAL) library and can be called through the rasterio interface. For this project, we used the following parameters: connectivity 4 and a boolean mask where only pixels that relate to forest are set to true. Now the algorithm clusters only pixels which are set to true to one polygon. After, we created a squared-sized buffer with a side length of 500 m around the polygon centroid (the geometric midpoint of the polygon). Because WGS84 is not an equal area CRS we must compute for each tile the buffer size separately. To compute the buffer size we used the Haversine formula in equation 2.2. Let  $d$  be the great-circle distance between two latitude, longitude pairs  $(\varphi_n, \lambda_n)$  and  $r$  is the earth radius of approximately 6 378 137 meter. Because this computation is expensive we assumed that the pixel resolution is equal for an entire raster tile. After extracting the buffer we counted the most frequent class under exclusion of forest and no data pixels within the buffer. Finally, if a most frequent class is defined and we reassign this value to the cluster. The reclassification algorithm is implemented as a Python function which requires as parameters a PDD raster image, a list of elements which should be interpreted as occupied cells for the clustering, pixel values which should be excluded from counting, the side length of the buffer, and the on-ground resolution.

$$d = 2r \arcsin \left( \sqrt{\sin^2 \left( \frac{\varphi_2 - \varphi_1}{2} \right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2 \left( \frac{\lambda_2 - \lambda_1}{2} \right)} \right) \quad (2.2)$$

After preparing the predictions we aggregated our results on PDDs for country, continental and global scale by using the administrative bounds of the Natural Earth layer. To present our results as maps for the three continental regions we developed our own visualization approach that is explained in section 2.2.5.

#### 2.2.2.3. Accuracy assessment

For examining the accuracy of our PDD predictions we used a confusion matrix (also known as two-way frequency tables, error matrix or contingency tables). These matrices are commonly used for an accuracy assessment of land cover classifications and enable the computation of marginal and conditional distributions [Congalton 1991; Foody 2002]. Table 2.6 shows a general model of a confusion matrix. Foundation for an accuracy assessment by means of a confusion matrix is a collection of ground-truth samples which can be compared with the class predictions for these samples produced by a classification algorithm. For the preparation of our accuracy assessment, we have to extract a collection of pixel samples with

a deforestation occurrence from our proximate driver maps (further also called predictions). Next, we compose a set of ground-truth for these predictions (further also called references).

**Table 2.6. A general model of a confusion matrix:**  $X_1, \dots, X_n$  denote classification categories of two independent raters.  $x_{n,n}$  are the actual samples sorted into the categories where the values in the diagonal show the agreement between both raters. The remaining cell values account for the disagreement between the two raters.  $\Sigma$  column and row show the marginal distribution and N is the total number of samples.

		Reference				
		Cl <sub>s</sub>	X <sub>1</sub>	...	X <sub>n</sub>	$\Sigma$
Predict	X <sub>1</sub>	$x_{1,1}$	...	$x_{1,n}$	$x_{1.} = \sum_{i=1}^n x_{1,i}$	
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
	X <sub>n</sub>	$x_{n,1}$	...	$x_{n,n}$	$x_{n.} = \sum_{i=1}^n x_{n,i}$	
$\Sigma$		$x_{.1} = \sum_{i=1}^n x_{i,1}$	...	$x_{.n} = \sum_{i=1}^n x_{i,n}$	$\Sigma\Sigma = N$	

To create our collection of ground-truth data we draw randomly 10 image tiles from all three continental regions (Latin America, Africa, Asia/Australia) namely the following tiles: 25N 024E, 20N 096E, 20N 102E, 20N 120E, 15N 066W, 15N 012W, 15N 072E, 10N 090W, 10N 018E, 10N 078E, 05N 072W, 05N 006E, 05N 114E, 00N 096W, 00N 060W, 00N 048W, 00N 006E, 00N 096E, 00N 114E, 00N 132E, 00N 138E, 05S 048W, 05S 042W, 05S 060W, 05S 030E, 10S 018E, 15S 060W, 15S 042E, 15S 054E and 20S 030E where the first part is the latitude coordinate and the last part is the longitude coordinate of the upper left corner. From each tile, we sampled by random 200 pixels which total to 6000 samples over the entire study region. The sampling is realized with our own raster sampling algorithm build in Python by means of the open source libraries numpy and rasterio. As mentioned in the previous section do we superimpose two datasets and only a certain amount of pixels per tile is classified as a proximate driver. Therefore, the sampling algorithm should only draw samples from occupied/classified pixels without replacement. The algorithm expects as parameters a raster image, the total number of samples to draw, a list of pixel values which should be interpreted as occupied cells, the affine transformation matrix of the raster image, and a seed for the random number generator. If occupied cells are set the algorithm will create a binary mask where each occupied cell is set to one relative to the input raster image. Otherwise, it sets all pixel values greater or less than zero to one. After, the row and column coordinates of each one are extracted from the mask and converted to a flat list of coordinate tuples. Next, it draws the predefined number of samples from the list by a random order and uses the image coordinates to get the pixel value from the raster image. If an affine transformation matrix is provided the image coordinates are converted to real-world coordinates. The seed argument

ensures that on every algorithm rerun the samples are drawn. For our sampling we set the parameters to the following values: samples 200, occupied pixels GL30 class values and 25 for regrowth, the affine matrix of the corresponding raster image, and the seed is 42. The per tile samples are stored as a Comma Separated Values (CSV) file.

For the collection of ground-truth data, we used a visual interpretation of satellite and aerial imagery provided by Google Maps. We developed a small JavaScript web application to access the imagery via the Google Maps API. The application expects as input a CSV file with the sampling coordinates. After upload of a sample file the user can cycle through the entries and the map jumps automatically to the coordinates of the sample. Now a reference label can be assigned to the coordinates by visual interpretation of the imagery. We subsequently assigned to all 6000 samples a reference label and downloaded the results as CSV.

Finally, we developed a Python class to compute the confusion matrix. The constructor of the class requires a list of reference and prediction labels. With the provided arguments it creates the confusion matrix. Further, it computes the following marginal and conditional distributions: overall accuracy  $OvAc$  by dividing the sum of classification agreements by the sample total  $N$  (equation 2.3), the producer accuracy  $PAc_{.n}$  by dividing the category agreement by the column category total (equation 2.4), the error of commission  $Com_{.n}$  (Type II error) by dividing the category disagreement by the column category total (equation 2.5), the user accuracy  $UAc_{.n}$  by dividing the category agreement by the row category total (equation 2.6), the error of omission  $Om_{.n}$  (Type I error) by dividing the category disagreement by the row category total (equation 2.7), and the Cohens Kappa by substituting equation 2.8 and 2.3 into equation 2.9.

$$p_0 = OvAc = \frac{\sum_{i=1}^n x_{i,i}}{N} \quad (2.3)$$

$$PAc_{.n} = \frac{x_{i,i}}{x_{.n}} \quad (2.4)$$

$$Com_{.n} = \frac{FN_i}{x_{.n}} \quad (2.5)$$

$$UAc_{.n} = \frac{x_{i,i}}{x_{n.}} \quad (2.6)$$

$$Om_{n.} = \frac{FP_i}{x_{n.}} \quad (2.7)$$

$$p_c = \frac{1}{N^2} \sum_{i=1}^n x_{i.} \cdot x_{i.} \quad (2.8)$$

$$Kappa = \frac{p_0 - p_c}{1 - p_c} \quad (2.9)$$

### 2.2.3. Emissions

Land cover change respectively deforestation releases carbon emissions. These emissions can be grouped to different categories like emissions from transportation, biomass removal, changes of soil carbon dynamics, processing of certain kind of commodities etc.. During the previous sections we developed an approach to predict the change of tree cover driven by proximate causes like conversion to cropland or else. Now we can use these predictions to approximate the CO<sub>2</sub> emissions uprising from this land cover transitions. For this study, we focus on the emissions emitted by biomass removal and from changes of soil carbon stock. The first paragraph is focused on the estimation of emissions from biomass removal and the second section tries to approximate the impact of land cover change on the soil organic carbon content.

To obtain the gross CO<sub>2</sub> emissions through proximate deforestation driver we selected the following raster tiles from our AISIM: the AGB stratum and our classification of the PDD. By means of Python, we implemented a function which accepts as parameter two raster images, the area a pixel covers in m<sup>2</sup>, a factor to convert carbon to CO<sub>2</sub>, and a list of proximate driver classes to consider as deforestation. We considered the following PDD classes as deforestation for the computation: 10 (cropland), 25 (regrowth), 30 (grassland), 40 (shrubland), 50 (wetland), 60 (water bodies), 70 (Tundra), 80 (artificial), and 90 (bareland). The function computes the gross emissions by using equation 2.10. Let  $Y_{ij}$  be the AGB in Mg C ha<sup>-1</sup> and  $X_{ij}$  the PDD at an pixel index  $i, j$  obtained from a raster image matrix in  $R^{N*M}$ . Let  $A$  be the area in ha a pixel covers for a certain image tile. This area is calculated by using the Haversine function from equation 2.2. Factor 3.7 converts Carbon to CO<sub>2</sub>. Let  $AGBE_{tile}$  be the cumulative emissions emitted from the removal of tree cover. Then this value can be obtained by taking the sum of the product of  $Y_{ij}$  and  $f(X_{ij})$ . Whereas the piecewise function  $f$  only evaluates to one if the proximate deforestation driver is within our set of classes we want to consider as deforestation. To obtain the gross AGB emissions through the deforestation by proximate deforestation driver we aggregated the sum of  $AGBE_{tile}$  for the regions Latin America, Asia, and Africa.

$$AGBE_{tile} = 3.7A \sum_{i=0}^N \sum_{j=0}^M f(X_{ij})Y_{ij} \quad (2.10)$$

To obtain the gross CO<sub>2</sub> emissions emitted by the change of soil organic carbon content we selected the following raster tiles from our AISIM: the IFL stratum, the GSOCmap, and our prediction of PDD. We decided to predict the SOC emissions for two different scenarios. In scenario one SC<sub>1</sub> we assume that all tree covered areas concerned by a land cover change are primary forest. For scenario two SC<sub>2</sub> we used IFL stratum to determine the forest type. If land cover changes within an IFL patch it concerns primary forest otherwise it is secondary

forest. The SOC emissions of both scenarios can be computed by equation 2.11. Let  $X_{ij}$  be the PDD from our prediction,  $Y_{ij}$  the forest type determined by the IFL stratum, and  $Z_{ij}$  the SOC Mg C ha<sup>-1</sup> determined by GSOCmap at an pixel with index  $i, j$  obtained from a raster image matrix in  $R^{N*M}$ . Let  $A$  be the area in ha a pixel covers for a certain image tile. This area is calculated by using the Haversine function from equation 2.2. Factor 3.7 converts Carbon to CO<sub>2</sub>. Let  $SOCC_{tile}$  be the cumulative soil organic carbon emissions emitted by the change of forest to another land cover type. Then this value can be obtained by taking the sum of the product of  $Z_{ij}$  and  $h(X_{ij}, Y_{ij})$ . Whereas the piecewise function  $h$  returns the mean soil organic carbon change and the standard error in respect to the forest type and proximate driver class. The mappings of drive classes and forest type for both scenarios are shown in table 2.7 and 2.8. This algorithm is implemented by means of Python. The function needs as parameter the required layers whereas the IFL stratum is optional, the area a pixel covers in m<sup>2</sup>, a conversion factor for carbon to CO<sub>2</sub>, an identifier for the forest type, and if the standard error should be included during the computation of the emission. If the IFL stratum is provided the algorithm will rely on this layer to determine the forest type otherwise it uses forest type identifier. To obtain the gross SOC emissions by the transition of land cover we aggregated the sum of  $SOCE_{tile}$  for the regions Latin America, Asia, and Africa.

$$SOCE_{tile} = 3.7A \sum_{i=0}^N \sum_{j=0}^M h(X_{ij}, Y_{ij})Z_{ij} \quad (2.11)$$

**Table 2.7. Scenario one mapping of soil organic carbon change to proximate driver:** In scenario one we assume that deforestation always occurs in primary forest. Refer to table 2.2 for the description of the proximate driver class. Standard errors of the soil organic carbon change factors are denoted in table 2.3. The symbols in superscript denote the following transitions: † Primary forest→Cropland, ‡ Primary forest→Secondary forest, and ◊ Primary forest→Grassland

Forest type	Proximate driver class					
	10	25	30	40	70	90
Primary	.252 <sup>†</sup>	.086 <sup>‡</sup>	.121 <sup>◊</sup>	.121 <sup>◊</sup>	.121 <sup>◊</sup>	.121 <sup>◊</sup>

**Table 2.8. Scenario two mapping of soil organic carbon change to proximate driver:** In scenario two we use the Intact Forest Landscape stratum to distinguish between deforestation in primary and secondary forest. Refer to table 2.2 for the description of the proximate driver class. Standard errors of the soil organic carbon change factors are denoted in table 2.3. The symbols in superscript denote the following transitions: † Primary forest→Cropland, ‡ Primary forest→Secondary forest, ◊ Primary forest→Grassland, § Secondary forest→Cropland, and \* Secondary forest→Grassland

Forest type	Proximate driver class					
	10	25	30	40	70	90
Primary	.252 <sup>†</sup>	.086 <sup>‡</sup>	.121 <sup>◊</sup>	.121 <sup>◊</sup>	.121 <sup>◊</sup>	.121 <sup>◊</sup>
Secondary	.213 <sup>§</sup>	-	.064 <sup>*</sup>	.064 <sup>*</sup>	.064 <sup>*</sup>	.064 <sup>*</sup>

## 2.2.4. Ecosystem service values

For a comprehensive insight of the ESV dynamics, we quantified the loss of ESV from tree cover depletion within the tropical zone. This loss of forest cover is frequently followed by a transition to other land cover types, which is expressed through our PDDs. These transitions can be interpreted as a gain or loss of ESVs and are computed subsequently. Finally, to give an insight into the overall trend of both ESV dynamics we determined the balance among the monetary loss and gain. We estimate the ESVs dynamics by considering three different approaches found in the literature [Costanza et al. 2014; de Groot et al. 2012; Siikamaki et al. 2015]. The first part of this section describes our approach to determine the ESV loss, followed by the method to obtain the gain in monetary units, and finally we explain how to derive the balance between both values.

By applying equation 2.12 we compute the gross ESV loss from the loss of tropical tree cover for the entire set of our AISIM. Let  $X_{ij}$  be the PDDs from our prediction at a pixel with the index  $i, j$  (image coordinates) obtained from a raster image matrix in  $R^{N*M}$ . Let  $ESV_{Forest}$  be the ESV of tropical forest from one of our selected source datasets from table 2.9. Let  $A$  be the area in ha a pixel covers for a certain image tile. The pixel area is calculated by using the Haversine function from equation 2.2. Let  $ESV_{loss,tile}$  be the cumulative loss in ESV for a certain tile from our AISIM. Then this value can be determined by adding the product of  $f(X_{ij})$  and  $ESV_{Forest, Dataset}$ . The function  $f$  returns only a one if the PDD is considered as deforestation by the mapping in table 2.9. The computation of ESV loss is implemented as a Python function. The function accepts as parameters a raster image of PDD predictions or a pandas data frame object. Further, the function requires as parameter the area a pixel covers in ha and the monetary value of tropical forest. Additionally, the function requires a list of PDD classes considered as the loss of tropical forest cover. We considered the following PDD classes as anthropogenic tree cover loss: cultivated land, regrowth, grassland, shrubland, artificial surfaces, and bareland as table 2.9 suggests. We excluded pixel classified as forest by our PDD prediction because within this class we are uncertain if a deforestation event occurred. Further, we excluded transitions of tree cover to wetland or water because we assume this LC changes are largely driven by natural causes.

$$ESV_{loss,tile} = A \sum_{i=0}^N \sum_{j=0}^M f(X_{ij}) ESV_{Forest, Dataset} \quad (2.12)$$

To estimate the gain in ESV from the transition of tropical forest to other land cover classes per AISIM tile we applied equation 2.13. Let  $X_{ij}$  be the PDD from our prediction at a pixel with the index  $i, j$  obtained from a raster image matrix in  $R^{N*M}$ . Let  $A$  be the area in ha a pixel covers for a certain image tile. The pixel area is calculated by using the Haversine

**Table 2.9. Biome ESV types corresponding to PDD classes:** The monetary values are given in 2007 Int’l\$  $y^{-1} \text{ ha}^{-1}$ , also known as Geary-Khamis Dollar. Mapping of biome types to PDD classes have the following schema: cultivated to cropland biome, regrowth to tropical forest biome, grassland to grassland biome, and artificial surfaces to the urban biome. The abbreviations in dataset column refer to the following publications: Costanza et al. [2014] (Co), de Groot et al. [2012] (Dg), and Siikamaki et al. [2015] (Wb)

Dataset	Proximate deforestation driver class					
	Cultivated	Regrowth	Grassland	Shrubland	Artificial	Bareland
Co	5567	5382	4166	-	6661	-
Dg	-	5264	2871	-	-	-
Wb	-	1312	-	-	-	-

function from equation 2.2. Let  $ESV_{gain,tile}$  be the cumulative gain of ESV per tile. Then this value can be determined by taking the sum of  $h(X_{ij})$ . The function  $h$  returns for a selected PDD class the corresponding monetary value. The algorithm is implemented in Python. The function accepts as parameters a raster image of PDD predictions or a pandas data frame object. Further, the function requires as parameter a mapping of ESVs to PDD classes from table 2.9. Additionally, the function can be called with a exclude list of PDD classes.

$$ESV_{gain,tile} = A \sum_{i=0}^N \sum_{j=0}^M h(X_{ij}) \quad (2.13)$$

By applying equation 2.14 we compute the ESV balance for each tile of our AISIM. Let  $ESV_{gain}$  be the total ESV gain per continental region and  $ESV_{loss}$  the total ESV loss per region. Then the ESV balance  $ESV_{balance}$  can be obtained by the difference of  $ESV_{gain}$  and  $ESV_{loss}$ .

$$ESV_{balance} = ESV_{gain} - ESV_{loss} \quad (2.14)$$

For our study we aggregated the ESV loss, gain, and net balance dynamics on two spatial ranges with the following extent: per continental region (Latin America, Africa, and Asia/Australia) highlighted in figure 2.3 and on global scale.

## 2.2.5. Binning analysis and visualization

During the previous sections, we were focused on the exercise of creating large scale spatial explicit predictions for land cover transitions and the following consequences for humanity. Now, an appropriate method must be developed to analyze and visualize these spatial explicit datasets by generalizing the problem domain. By the nature of fine resolution raster images and the large area of our study extent we must handle a large  $N$  (many samples) and the resulting high dimensionality respectively complexity of relationships among the samples [Carr 1990]. Raster image maps can be interpreted as multivariate scatter plots. In our case this scatter plot has the three dimensions  $x$  is the longitude,  $y$  the latitude coordinate

of a pixel, and  $z$  is the nominal scaled pixel value in case of the PDD prediction. Drawing scatter plots with large multidimensional  $N$  commonly leads to overplotting and hidden point densities [Carr et al. 1987]. Additionally, it is to assume that the distribution of PDD is not equally distributed over the entire study extent. Hence, there should be regions with sparse data densities and with high densities but our goal is to visualize land cover changes on a continental level. As mentioned the ground resolution of one pixel covers an area of approximately 30x30 m and as an example, the bounding box of Latin America covers an area of  $5 * 10^7$  Km<sup>2</sup>. The large frame size as well the unequal distributed data leads to the issue that only large scale land cover changes are representable and small scale isolated changes stay hidden.

By referring to the latter paragraph our goal should be to develop a process to solve the representation issues and generates satisfying maps. In the case of raster data, one opportunity could be a re-sampling to a coarser on-ground resolution. This approach may solve the overplotting as well resolution issues and normalize unequal distributed data. For nominal scaled data the commonly used re-sampling methods are nearest neighbor or majority wins [Reference](#). Both approaches are not appropriate because they would negate spatial patterns and eliminate important land cover class frequency distributions. Another well-accepted method is binning of spatially explicit data with a regular polygon that can tessellate the plane [Carr et al. 1992]. Polygon tessellations provide numerous opportunities for presenting multivariate statistical and visual summaries. The scale of a polygon may be used to visualize pixel densities within the bounds and a color gradient may be used to prepare a choropleth map for nominal or ordinal scaled data. Additionally, the interior of a polygon may be used to prepare a pie chart. Hence, binning enables convenient visualization of multidimensional data. For preparing a regular tessellation only three types of convex polygons can be used to tessellate the plane: squares, equilateral triangles, and hexagons [Carr et al. 1992]. Square tessellations are the most common method in comparison with hexagons for binning and visualizing spatial data. Every raster image is already a square tessellation of the mapped object and most of the image processing algorithms are focused on squares. Hexagon mosaic maps have two major advantages over square tessellations: visual appeal and representational accuracy. Binning of data by a square or hexagon mosaic creates visual lines. These lines compete with the data generated patterns. Especially humans have a strong visual response to horizontal and vertical lines. Hence, the line artifacts of square tessellations are distracting and should be avoided. Thus, we decided to use hexagon mosaic maps to represent the visual and statistical results of our study. For bivariate representations, we select the combination of scaling and color gradient. Multivariate data is visualized by hexagonal pie charts. The following paragraphs describe our algorithmic approach to create these mosaic maps. We used Python and the open source library shapely to implement our algorithms.

The first step to construct a hexagon tessellation is to define the vertices of the polygon.

There are two common orientations of hexagons in  $R^2$  flat topped and pointy topped. For our hexagon construction we decided to use pointy topped polygons. For flexibility our algorithm accepts one out of four parameter to construct a hexagon polygon. The unit of the parameter is always in map units. Let  $D$  be the long diagonal (diameter of the circumscribing circle),  $d$  the short diagonal (diameter of the inscribed circle),  $A$  the area the hexagon should cover, and  $e$  the edge length of a hexagon. Let  $R$  be the radius of the circumscribing circle. Then  $R$  can be obtained by applying equation 2.15 with one out of the parameter set  $D, d, A$ , or  $e$ .  $R$  is used to compute the center vector  $\vec{m} = \langle c_x, c_y \rangle$  of the polygon by applying equation 2.16 and 2.17. The polygon center is always located in the first quadrant of the Cartesian coordinate system. Now, by using  $R, c_x$ , and  $c_y$  we can obtain  $\mathbf{H}$  the anti-clockwise orientated vertex matrix of a hexagon. The construction of a hexagon by using the introduced method is shown in the left bottom corner of figure 2.5. The next paragraph describes how we derive a tessellation from the constructed hexagon.

$$R = \frac{\sqrt{2A}}{\sqrt[4]{27}} = \frac{D}{2} = \frac{d}{\sqrt{3}} = e \quad (2.15)$$

$$c_x = \frac{R\sqrt{3}}{2} \quad (2.16)$$

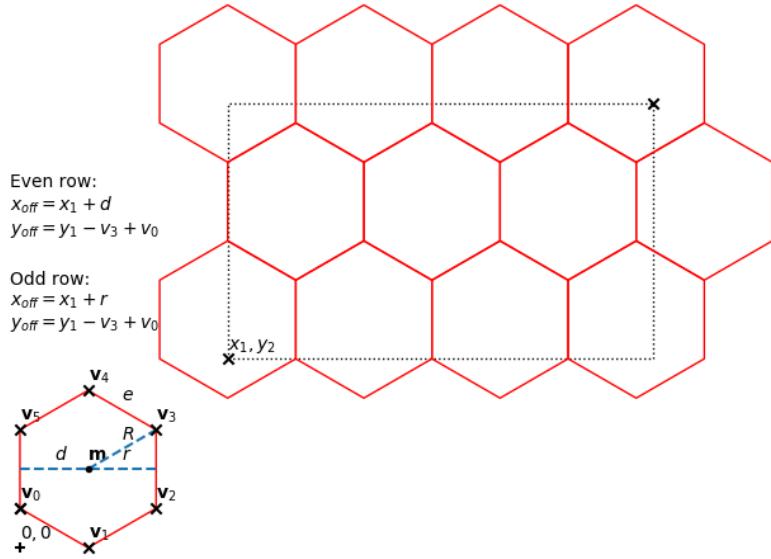
$$c_y = R \quad (2.17)$$

$$\mathbf{H} = \begin{bmatrix} 0 & c_x & 2c_x & 2c_x & c_x & 0 \\ R\sin\left(\frac{7\pi}{6}\right) + c_y & 0 & R\sin\left(\frac{11\pi}{6}\right) + c_y & R\sin\left(\frac{\pi}{6}\right) + c_y & 2R & R\sin\left(\frac{5\pi}{6}\right) + c_y \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (2.18)$$

To create a polygon grid of a plane image we must align several hexagons to cover the image. For our tessellation algorithm we use the vertex matrix  $\mathbf{H}$  computed by the previously described approach and subsequently translate it to its position within the grid. We expect to receive the bounds matrix  $\mathbf{B}$  of the raster image which should be tessellated by hexagons, equation 2.19. Let  $x_1, y_1$  be the left bottom corner coordinates and  $x_2, y_2$  the right top corner coordinates of an image. Let  $x_{off}(0), y_{off}(0)$  in equation 2.20 and 2.21 be the initial coordinates for creating a polygon grid over a plane. Then we can obtain  $x_{off}(n+1)$  the  $x$  coordinates for even rows by applying equation 2.22 and  $x_{off}(n+1)$  the  $x$  coordinates for odd rows by equation 2.23. Where  $r$  is the radius of an inscribed circle in a hexagon and can be obtained by dividing  $d$  by 2. Then  $\mathbf{H}$  can be translated to the vertex matrix  $\mathbf{T}$  by applying the dot product of an affine transformation matrix and  $\mathbf{H}$ , equation 2.25.

$$\mathbf{B} = \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} \quad (2.19)$$

$$x_{off}(0) = x_1 \quad (2.20)$$



**Figure 2.5. Hexagon tessellation:** Located at the left bottom corner in red a hexagon defined by its geometric properties the 6 vertex vectors  $\{\vec{v}_0, \dots, \vec{v}_5\}$  (black crosses), with center vector  $\vec{m}$ , edge length  $e$ ,  $R$  radius of the circumscribing circle,  $r$  radius of the inscribed circle and  $d$  the short diagonal (diameter of the inscribed circle). Top right black dotted box are the bounds of an area which is tessellated by a hexagon grid in red. Each grid cell is translated from the origin hexagon at its position by computing the  $x_{off}$  and  $y_{off}$  offset with the presented equations at the left-hand side of the grid.

$$y_{off}(0) = y_1 \quad (2.21)$$

$$x_{off}(n+1) = x_{off}(n) + d \quad (2.22)$$

$$x_{off}(n+1) = x_{off}(n) - r + d \quad (2.23)$$

$$y_{off}(n+1) = y_{off}(n) - v_{0,2} + v_{3,2} \quad (2.24)$$

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & x_{off}(n) \\ 0 & 1 & y_{off}(n) \\ 0 & 0 & 1 \end{bmatrix} \circ \mathbf{H} \quad (2.25)$$

Goal: A pie chart within the area of a hexagon, split the hexagon in horizontal pieces which represent a ratio

How: Compute the y coordinate of the split line from the ratio, the distance between  $y_1$  and  $y_2$ , compute from this the x coordinates

$$y = \frac{P(y_2 - y_1)}{100} + y_1 \quad (2.26)$$

$$f^{-1}(y) = \begin{cases} -\frac{y-y_1}{\tan(\frac{\pi}{6})} + \frac{x_1+x_2}{2} & \text{if } y_1 \leq y < y_1 + R \sin(\frac{5\pi}{6}) \\ x_1 & \text{if } y_1 + R \sin(\frac{5\pi}{6}) \leq y < R(\sin(\frac{5\pi}{6}) + 1) \\ \frac{y-y_2}{\tan(\frac{\pi}{6})} + \frac{x_1+x_2}{2} & \text{if } R(\sin(\frac{5\pi}{6}) + 1) \leq y \leq y_2 \end{cases} \quad (2.27)$$

$$g^{-1}(y) = \begin{cases} \frac{y-y_1}{\tan(\frac{\pi}{6})} + \frac{x_1+x_2}{2} & \text{if } y_1 \leq y < y_1 + R \sin(\frac{5\pi}{6}) \\ x_2 & \text{if } y_1 + R \sin(\frac{5\pi}{6}) \leq y < R(\sin(\frac{5\pi}{6}) + 1) \\ -\frac{y-y_2}{\tan(\frac{\pi}{6})} + \frac{x_1+x_2}{2} & \text{if } R(\sin(\frac{5\pi}{6}) + 1) \leq y \leq y_2 \end{cases} \quad (2.28)$$

$$\mathbf{L} = \begin{bmatrix} f^{-1}(y) & g^{-1}(y) \\ y & y \end{bmatrix} \quad (2.29)$$

Goal: All use hexagons with an area of 0.5 degrees i square, aggregated tiles per continent americas, africa, asia

Treecover: Count tree cover pixels within hexagon in canopy interval (10,100], count total pixels within hexagon, compute pixel area with haversine, divide tree covered area by total area, 5 ratio bins 0.2 0.4 0.6 0.8 1.0 for scaling, store the area for interval, mean canopy density over occupied pixels, use mean canopy density for choropleth map

Loss: count frequency of pdd recognized as deforestation

PDD: count frequencies within a hexagon, compute frequency ratios, segment hexagons by the ratio of pdd driver, order the drivers in decreasing order, most common is first, tried with scaling but not feasible cause few big sized and many small sized, just scale them a bit down cause visual appeal,

Hexagonal Country-boundaries: Hexagon grid over the continent aggregate cells which cover a country, vastly made by hand

# 3. Results

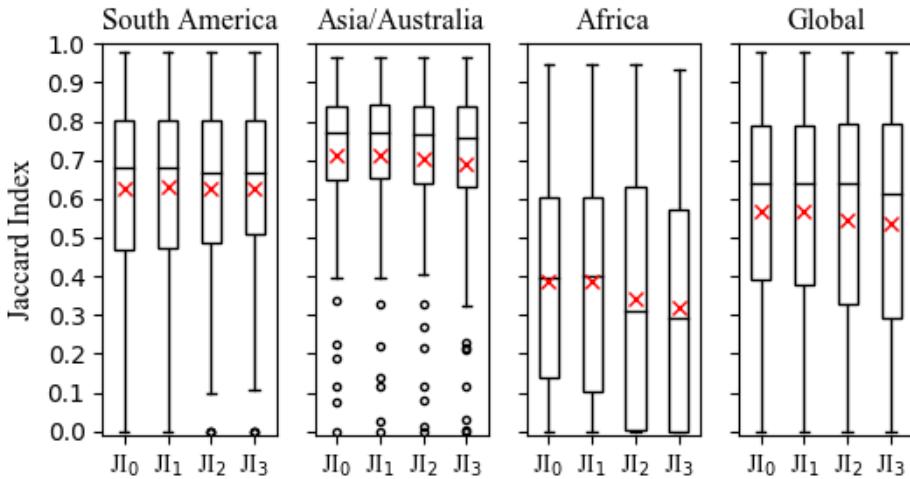
## 3.1. Deforestation

### 3.1.1. Forest definition

Our goal is to determine at which canopy cover density the agreement between GL30 and GFC tree cover is greatest to receive the subsequent PDD for stable LC transitions introduced by anthropogenic causes. This process should ensure that we keep the largest number of tree cover loss samples from the GFC dataset while harmonizing the tree cover definition between both layers. We applied the JI to determine the similarity between each tile pair from our AIS. The JI computation is grouped by the continental regions Latin America (82 tiles), Asia/Australia (86 tiles), and Africa (101 tiles). We determined the similarity for the following canopy density intervals: (0, 100], (10, 100], (20, 100], and (30, 100]. Later we excluded all tiles with a initial JI (canopy density interval (0, 100]) from our analysis because these tile pair does not contain any tree cover. We excluded 6, 9, and 15 tiles for Latin America, Asia/Australia and Africa, respectively. To determine the canopy density interval where the agreement is at maximum we applied the non-parametric tests Wilcoxon signed-rank test and Wilcoxon rank-sum test. Both of the tests are performed as a one- and two-sided to deduce, if there is a difference in agreement (equality) and which direction (less or greater) has this difference. To address the higher probability of family-wise error rate in multiple comparisons we applied a Holm correction for Wilcoxon signed-rank tests and a Benjamini and Hochberg correction for Wilcoxon rank-sum test. We applied continental and global testing to deduce regional differences and to determine the optimum for our subsequent PDD predictions. Further, we compared the tree cover agreement of the three continental regions. Our initial hypothesis was that the tree cover agreement is at its maximum within the canopy density interval of (30, 100] for the entire study extent. We assumed that for Latin America and Asia/Australia the best results could be achieved with the same canopy density threshold. For Africa we assumed that the highest agreement could be achieved within the interval of (10, 100] because this region comprises a higher frequency of sparse woodland cover. The following paragraphs present our results for the three regions in the following order: Latin America, Asia/Australia and Africa. The last paragraph discusses the results for

the entire study extent and determines which canopy density we used for the following PDD prediction.

For Latin America, Asia/Australia and Africa as well the entire study extent figure 3.1 shows the distribution of the computed JIs for all tile pairs within the canopy density intervals. The x-axis are the different canopy density intervals where the label  $JI_0$  accounts for  $(0, 100]$ ,  $JI_1$   $(10, 100]$ ,  $JI_2$   $(20, 100]$ , and  $JI_3$   $(30, 100]$ , respectively. The y-axis is the corresponding JI between 0 and 1 where 0 highlights a complete disagreement and 1 a full agreement. The sample mean is labelled by a red cross and the boxes comprises the  $Q_1$  (25 %),  $Q_2$  (50 %), and  $Q_3$  (75 %) sample interval, respectively.



**Figure 3.1. Tree cover similarity distribution:** This box-plot shows the distribution of Jaccard Indexes for each raster image tile pair of GlobeLand30 and Global Forest Change tree cover from 2000. The labels  $JI_0$ ,  $JI_1$ ,  $JI_2$ , and  $JI_3$  on the x-axis account for the canopy density classes  $(0, 100]$ ,  $(10, 100]$ ,  $(20, 100]$ , and  $(30, 100]$ , respectively. The y-axis is the Jaccard Index of the corresponding raster image pair, where 0 is a total disagreement and 1 a total agreement. Red crosses within the  $Q_{25}$ ,  $Q_{50}$ , and  $Q_{75}$  boxes highlight the sample mean. Whiskers are 1.5 times the  $IQR$ .

For Latin America the sample mean does not change significantly within the four canopy density experiments. It is approximately 0.62 while sample median decreases from 0.68 to 0.66 from the interval class to the last. The upper 25 % of the first experiment interval have a tree cover similarity ranging between approximately 0.8 and 1. This holds true over the other three experiments while only the maximum similarity slightly increases from 0.9787 to 0.9798. The figure B.1 in the appendix B suggests, that exclusions of canopy densities smaller than 11% increases slightly the tree cover agreement of the upper 25% but the exclusion of higher canopy density provides no benefit. This can be explained by the fact that the upper percentile already contain samples with a high tree cover agreement and it is to assume that only a small number of pixels have canopy density smaller than 30. Therefore, the interval change has only a small impact on this samples. For the first two canopy density intervals the range of the lower 25% percentile is between approximately 0.0003 and 0.45.

Whereas the range increases from 0.0 to 0.5 for the last two interval classes. The figure B.1 reveals a strong regional dependency of tree cover agreement within different canopy density classes of the lower percentiles. Whereas samples from the northern hemisphere show a decline in agreement the samples of the southern hemisphere show an increase of agreement. The strong up-shift of the southern samples within the last two intervals increases the range of lower percentile. This suggests in general that samples with low tree cover agreement benefit by the exclusion of lower canopy densities. The mobility of the samples within the  $Q_1$  and  $Q_3$  percentile show no general trend. For the first two experiments it ranges between 0.5 and 0.8 and for the last two it shows a small decline in  $IQR$ . This suggests that tiles within this group could benefit from a local optimization of canopy density exclusion. We applied a Wilcoxon signed-rank test to deduce which canopy density class yields the highest tree cover agreement overall samples in Latin America. Table 3.1 shows the results for the two-sided test and table 3.1 for the one-sided test. The two-sided test reveals that only the similarity distribution between  $Jl_0$  and  $Jl_1$  is significantly different ( $p < 0.01$ ), while the comparison of the other distributions suggest that they are equal. The directional test of  $Jl_0$  and  $Jl_1$  suggests that the regional tree cover agreement is significantly greater ( $p < 0.005$ ) if canopy densities smaller than 10 are excluded. The directional test does not confirm that the distribution of  $Jl_1$  is significantly greater than  $Jl_2$  and  $Jl_3$ . Therefore it can be assumed that the exclusion of lower canopy densities could yield better results for certain tiles. The result of our experiments suggests that the tree cover agreement between GL30 and GFC at canopy densities greater than 10 is at its maximum for Latin America. In case of local studies or for smaller extents the canopy density should be selected by a single tile approach to optimize the tree cover agreement by maximizing the number of data points of the GFC dataset.

**Table 3.1. Continental experiment group comparison:** This table shows the results of a two-sided Wilcoxon signed-rank test to detect continental differences in the tree cover agreement by considering different canopy densities between GlobeLand30 and Global Forest Change at 2000. The classes  $Jl_0$ ,  $Jl_1$ ,  $Jl_2$ , and  $Jl_3$  as row and column headings account for the canopy density classes (0,100], (10,100], (20,100], and (30,100], respectively. The test hypothesis is  $H_0: X_1 = X_2$  where  $X_1$  is the column  $Jl_n$  class and  $X_2$  the row  $Jl_n$  class. The significance is indicated by  $p^* < 0.05$ ,  $p^{**} < 0.02$ , and  $p^{***} < 0.01$ .

Cls	Latin America			Asia/Australia			Africa		
	$Jl_0$	$Jl_1$	$Jl_2$	$Jl_0$	$Jl_1$	$Jl_2$	$Jl_0$	$Jl_1$	$Jl_2$
$Jl_1$	.00***	-	-	.71	-	-	.22	-	-
$Jl_2$	.30	1.	-	.00***	.00***	-	.09	.09	-
$Jl_3$	.64	1.	1.	.00***	.00***	.00***	.00***	.00***	.00***

For Asia/Australia, the sample mean scatters around 0.7 as the red crosses in figure 3.1 suggest. The sample mean decreases slightly at higher canopy density intervals. Further, the median is approximately 0.8 by showing a slight decrease at higher canopy density intervals too. The range of the upper percentiles for all experiments groups is between approximately 0.85 and 0.96 while the maximum agreement decreases slightly from 0.9654 to 0.9634. Fig-

**Table 3.2. Continental experiment group directional comparison:** This table shows the results of a one-sided Wilcoxon signed-rank test to detect the direction of continental differences in the tree cover agreement by considering different canopy densities between GlobeLand30 and Global Forest Change at 2000. The classes  $Jl_0$ ,  $Jl_1$ ,  $Jl_2$ , and  $Jl_3$  as row and column headings account for the canopy density classes (0,100], (10,100], (20,100], and (30,100], respectively. The test hypothesis is  $H_0: X_1 \leq X_2$  and  $H_0: X_2 \geq X_1$  where  $X_1$  is the column  $Jl_n$  class and  $X_2$  the row  $Jl_n$  class. The significance is indicated by  $p^* < 0.05$ ,  $p^{**} < 0.025$ ,  $p^{***} < 0.01$ , and  $p^\dagger < 0.005$ .

Cls	Latin America				Asia/Australia				Africa			
	$Jl_0$	$Jl_1$	$Jl_2$	$Jl_3$	$Jl_0$	$Jl_1$	$Jl_2$	$Jl_3$	$Jl_0$	$Jl_1$	$Jl_2$	$Jl_3$
$Jl_0$	-	.00 <sup>\dagger</sup>	.14	.33	-	1.	1.	1.	-	.65	1.	1.
$Jl_1$	1.	-	.55	.55	.36	-	1.	1.	.89	-	1.	1.
$Jl_2$	1.	1.	-	.55	.00 <sup>\dagger</sup>	.00 <sup>\dagger</sup>	-	1.	.03*	.03*	-	1.
$Jl_3$	1.	1.	1.	-	.00 <sup>\dagger</sup>	.00 <sup>\dagger</sup>	.00 <sup>\dagger</sup>	-	.00 <sup>\dagger</sup>	.00 <sup>\dagger</sup>	.00 <sup>\dagger</sup>	-

ure B.2 in the appendix B, reveals that in general the tree cover agreement increases if canopy densities below 10 % are excluded but the exclusion of densities above 20 % reverts this. For Asia/Australia the lower percentile ranges between approximately 0 and 0.65 while most of the samples show a decline in tree cover agreement if canopy densities above 20 are excluded. Within the  $Q_{1,3}$  percentile a per tile relationship is detectable. The range of this percentile is between 0.65 and 0.85 for the first two classes  $Jl_0$  and  $Jl_1$  while the range increases for the last two experiments. As mentioned no clear trend is observable some of the samples benefit if the considered canopy density interval is increased till 30 % and some show a decrease in agreement if the canopy density is lift over 10 %. The two-sided Wilcoxon test in table 3.1 reveals that the similarity distribution is significantly different ( $p<0.01$ ) between each experiment group except the pair of  $Jl_1$  and  $Jl_0$ . The directional test in table 3.2 reveals that the tree cover agreement distributions of  $Jl_2$  and  $Jl_3$  are significantly smaller than  $Jl_0$  and  $Jl_1$  ( $p<0.005$ ). Further this results show that the distributions of  $Jl_0$  and  $Jl_1$  have no directional differences. This could be explained by a regional or tile-wise agreement component. While some of the tiles show strong increase in similarity if canopy density is set to 10 %, others show a decrease. A more detailed analysis could be performed by applying a smaller canopy density step-size. For studies targeting the region Asia/Australia the results of the directional tests suggest to use all data from GFC within the canopy density interval of (0,100]. While the figure B.2 suggests to include all data form the interval (10,100].

The box-plot for Africa in figure 3.1 shows that the similarity range of the upper 75 % is the greatest among our study regions. It ranges between 0.15 and 1.0 for  $Jl_0$  while the range increases if smaller canopy densities are excluded. The first two experiment groups  $Jl_0$  and  $Jl_1$  have a nearly similar mean and median of approximately 0.38 and 0.4, respectively. Whereas both metrics show a strong decline to 0.33 and 0.3 in the last two experiment groups  $Jl_2$  and  $Jl_3$ , respectively. Figure B.3 in the appendix B reveals that the upper 25 % different to Latin America and Asia not benefit in general by exclusion of lower canopy densities. Africa has the highest amount of tiles where there tree cover agreement is smaller than 0.1. For the sec-

ond experiment group  $J_1$  11 tiles from the AISIM have a complete disagreement ( $J_1 = 0$ ) if the canopy density interval is reduced to  $(10, 100]$ . This trend continues if the canopy density interval is further reduced. This already suggests that reducing the canopy density interval for Africa is not feasible. The two-sided test in table 3.1 reveals that  $J_3$  is significantly different ( $p < 0.01$ ) in its similarity distribution compared to the other experiment groups  $J_0$ ,  $J_1$ , and  $J_2$ , respectively. The other experiment groups could origin from the same similarity distribution. Especially, the tree cover agreement between  $J_0$  and  $J_1$  show evidences that it could origin from the same distribution. This is comparable with the Asia/Australian region. The table 3.2 shows the directional component of the tree cover agreement between the experiments groups for Africa. The tree cover agreement of  $J_2$  and  $J_3$  is significantly smaller than  $J_0$  or  $J_1$  as the table suggest. Therefore the exclusion of canopy densities above 10 % reduces the continental tree cover agreement in Africa. For the two experiment groups  $J_0$  and  $J_1$  no directional agreement component could be proofed but the figure B.3 shows a strong regional component of tree cover agreement. The strong regional component can be explained by the high share of sparse woodland in Africa. The figure 3.6 in section 3.1.2 shows that different from Asia/Australia and Latin America a vast amount of African tree cover has a canopy density between 0 and 46 %. The results for Africa suggests to set the canopy density interval to  $(0, 100]$  to optimize the tree cover agreement between GL30 and GFC on a continental level. Preferable the strong regional component suggests to optimize the tree cover agreement per tile to achieve a higher tree cover agreement for the entire continent.

The table B.1 in the appendix B shows that in Asia/Australia the tree cover similarity between GL30 and GFC is the greatest out of all three regions. Only within the last experiment group  $J_3$  the median tree cover agreement between Asia/Australia and Latin America could be the same shown in table B.2. Africa has the poorest tree cover agreement out of our three continental regions. Asia's higher tree cover agreement compared to Latin America could relate to the smaller total landmass and vice versa forest cover within this region and the high share of dense forest cover. Therefore both global LC datasets have a smaller chance to miss forest covered pixels. That both regions achieve better tree cover similarity as Africa could be explained by the high amount of dense forest cover within this regions. It could be assumed that the probability of misclassification of forest cover increases as lower the canopy density is. In regards of tile-wise tree cover agreement optimization would Africa benefit at most out of the three regions. In Asia/Australia and Latin America only certain tiles would benefit from a exclusive optimization as the figures B.1 and B.2 suggests.

On the far right of figure 3.1 the tree cover agreement of the entire study extent is shown. The sample mean and median differ between the first two experiment groups and the last two groups. For the first two groups the mean and median account for 0.56 and 0.63, respectively. The last two experiment groups show a decline to 0.53 and 0.53 for mean and median.

For the upper percentile the same statement as for the regions holds true. In general, the samples with high tree cover agreement benefit from the exclusion of lower canopy densities. The lower percentile shows strong regional or tile-wise tree agreement dependencies. If we consider the entire sample range the mid percentile steadily increases its range from 0.4 to 0.8 till 0.3 to 0.8. As mentioned in the regional analysis, this percentile is characterized by inhomogeneous changes of tree cover agreement by the step-wise change of the canopy density. In general the trend points downwards as the decrease in median and the range increase show. The results of the two-sided Wilcoxon test in table 3.3 show a significant difference in distribution ( $p < 0.02$  and  $p < 0.01$ ) between each experiment group except  $Jl_0$  and  $Jl_2$  where the similarity distribution could be the same. Table 3.4 highlights the directional component of these distribution differences. At a global scale the tree cover agreement is at its maximum if we set the canopy density interval to (10, 100]. The second experiment group  $Jl_1$  is significantly greater than  $Jl_0$  ( $p < 0.005$ ) and the last two groups  $Jl_2$  and  $Jl_3$  are significantly smaller ( $p > 0.005$ ) than  $Jl_1$ . Further, the tree cover agreement of  $Jl_0$  is significantly greater than ( $p < 0.005$  and  $p < 0.05$ )  $Jl_2$  and  $Jl_3$ . Therefore it is proved that canopy densities above 20 % reduce the tree cover agreement between GL30 and GFC on a global level. On a global scale as the analysis on tree cover agreement suggests the highest agreement can be achieved within the  $Jl_1$  interval. Therefore we decided to proceed for our study of PDD and the derived products with this definition of tree cover.

**Table 3.3. Global experiment group comparison:** This table shows a two-sided Wilcoxon signed-rank test to detect differences in the tree cover agreement by considering different canopy densities between GlobeLand30 and Global Forest Change at 2000. The classes  $Jl_0$ ,  $Jl_1$ ,  $Jl_2$ , and  $Jl_3$  as row and column headings account for the canopy density classes (0,100], (10,100], (20,100], and (30,100], respectively. The test hypothesis is  $H_0: X_1 = X_2$  where  $X_1$  is the column  $Jl_n$  class and  $X_2$  the row  $Jl_n$  class. The significance is indicated by  $p^* < 0.05$ ,  $p^{**} < 0.02$ , and  $p^{***} < 0.01$ .

Cl <sub>s</sub>	Jl <sub>0</sub>	Jl <sub>1</sub>	Jl <sub>2</sub>
Jl <sub>1</sub>	.01**	-	-
Jl <sub>2</sub>	.07	.04*	-
Jl <sub>3</sub>	.00***	.00***	.00***

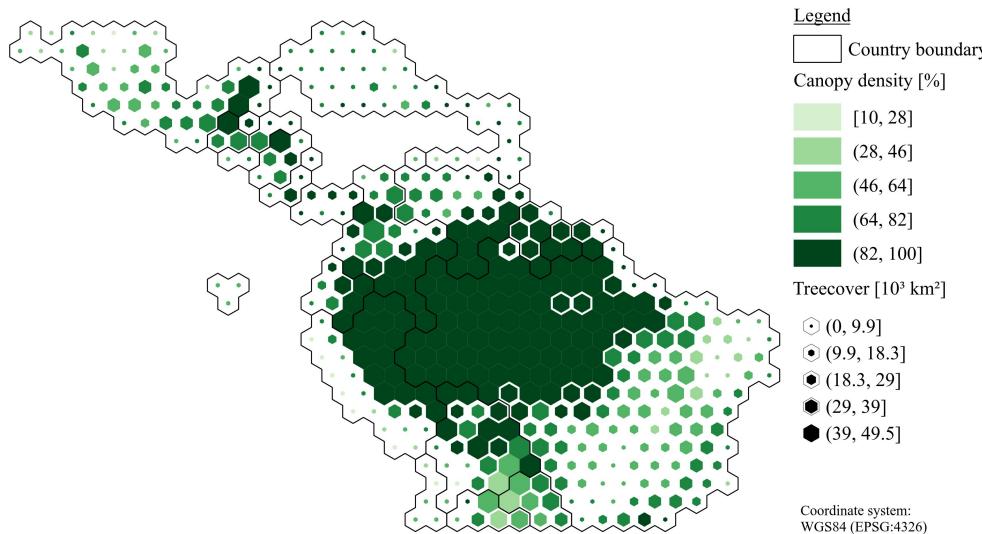
**Table 3.4. Global experiment group directional comparison:** This table shows a one-sided Wilcoxon test to determine the direction of global differences in the tree cover agreement by considering different canopy densities between GlobeLand30 and Global Forest Change at 2000. The test hypothesis is  $H_0: X_1 \leq X_2$  and  $H_0: X_2 \geq X_1$  where  $X_1$  is the column  $Jl_n$  class and  $X_2$  the row  $Jl_n$  class. The significance is indicated by  $p^* < 0.05$ ,  $p^{**} < 0.025$ ,  $p^{***} < 0.01$ , and  $p^\dagger < 0.005$ .

Cl <sub>s</sub>	Jl <sub>0</sub>	Jl <sub>1</sub>	Jl <sub>2</sub>	Jl <sub>3</sub>
Jl <sub>0</sub>	-	.01***	1.	1.
Jl <sub>1</sub>	1.	-	1.	1.
Jl <sub>2</sub>	.07	.03*	-	1.
Jl <sub>3</sub>	.00 <sup>\dagger</sup>	.00 <sup>\dagger</sup>	.00 <sup>\dagger</sup>	-

### 3.1.2. Patterns of tree cover and deforestation

This section is intended to present a comprehensive insight in the tropical tree cover distribution within our study extent at 2000 over the tree continental regions. Further, we highlight at which sites the tree cover loss peaked between 2000 and 2010. The tree cover maps are derived from the GFC tree cover 2000 layer while we selected only pixels within our forest definition which refers to the canopy density interval (10, 100]. For an appropriate visualization of multivariate spatial data on a large extent we selected our hexagonal binning approach. For the tree cover maps we computed the total area covered by trees within a polygon and divide by the total area of the hexagon to determine the scaling. Additionally we aggregate the canopy density within a hexagon by applying the arithmetic mean. To arrange the tree cover loss maps we used our PDD products. We computed the loss area within each hexagon for the following PDD classes: cultivated land (10), regrowth (25), grassland (30), shrubland (40), artificial surfaces (80), and bareland (90). To determine the polygon scaling we divided the per hexagon loss area by the highest observed loss within a continental region. Forest cover, losses, and hexagon areas are computed by applying the Haversine equation. A hexagon in unscaled shape covers an area of 0.5 decimal degrees. The maps in this section should be interpreted as precursor to our PDD predictions to detect regional and continental patterns of deforestation and as an example how large multivariate spatial data can be visualized and evaluated by a more advanced aggregation approach.

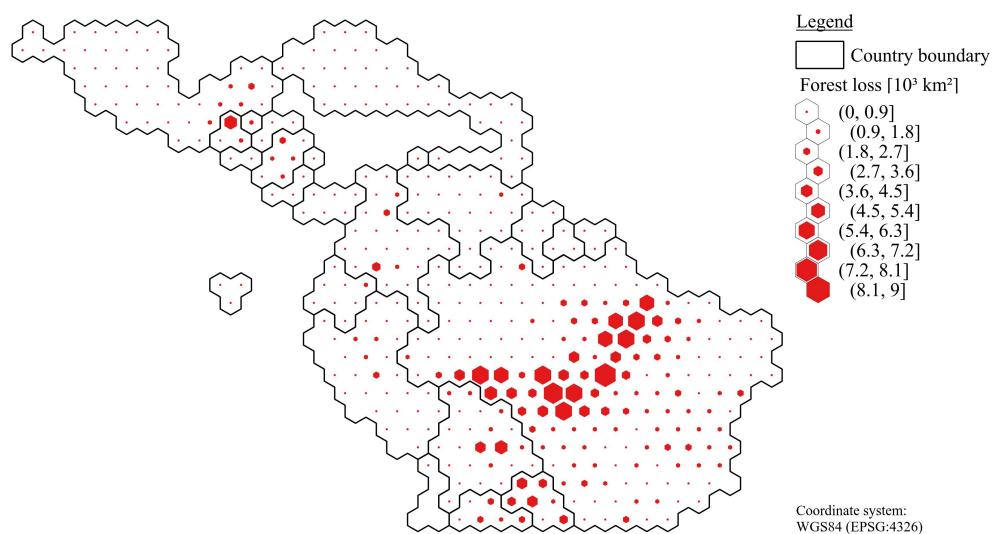
The map in figure 3.2 shows the tree cover and canopy density distribution within our study extent for Latin America at 2000. The center of the map is covered by the core tropical rain forest characterized by high tree cover per hexagon between approximately 39 and 49 thousand km<sup>2</sup> and a mean canopy density between 82 and 100 %. The core rain forest zone is distributed over 9 Latin American countries, namely: Colombia, Venezuela, Guiana, Suriname, French Guiana, Brazil, Bolivia, Peru, and Ecuador. In Brazil, within the tropical core zone two hexagons are smaller scaled which highlights a tree cover between 29 and 39 thousand km<sup>2</sup> at a canopy density between 82 and 100 %. These two hexagons comprising floodplain forest at the borders of the Amazon river located in the state Pará. This river basin enclosed by the three cities Santarém, Almeirim, and Óbidos suffered severe deforestation since the 16th century [Renó et al. 2011]. Renó et al. suggests that major forest losses took place between 1950 and 1975 followed by lower annual deforestation till 2008. Located in the lower left part of the map crossing the borders of Bolivia, Paraguay, and Argentina is the Gran Chaco a hot semi-arid wooded grassland also known as tropical dry forest characterized by tree cover between approximately 29 and 49 thousand km<sup>2</sup> per hexagon and canopy densities between 28 and 46 % [Caldas et al. 2013]. The tropical rain forest is surrounded by tropical moist forest characterized by tree cover up to approximately 49 thousand km<sup>2</sup> per hexagon while the canopy density is between 10 and 82 %.



**Figure 3.2. Tree cover and canopy density in Latin America at 2000:** This map shows the tree cover and mean canopy density distribution at 2000. An unscaled hexagon covers an area of 0.5 decimal degrees which translates to an area of approximately 49 thousand  $\text{km}^2$  at the equator. Tropical rain forest in the center of the map is characterized by a tree cover between approximately 29 and 49  $\text{km}^2$  and canopy densities between 82 and 100 %. The rain forest is surrounded by tropical moist forest with tree cover up to 49  $\text{km}^2$  but lower mean canopy densities between 10 and 82 % as the rain forest. The tropical dry forest also known as Gran Chaco is located in the lower left of the map distributed over the countries Bolivia, Paraguay, and Argentina.

The figure 3.3 shows the distribution of tree cover losses over Latin America within the time frame of 2000 till 2010. During this period an area of approximately 388 thousand  $\text{km}^2$  was deforested as table C.1 in the appendix C shows. We could identify for several countries deforestation hot-spots where the deforested area is between approximately 2.7 and 9 thousand  $\text{km}^2$  as the map suggests. Tropical countries with deforestation hot-spots are: Paraguay, Argentina, Bolivia, Brazil, Colombia, Peru, and Guatemala. In Brazil the hot-spots of tree cover loss are known as arc of deforestation and cover the states of Acre, Rondônia, Mato Grosso, and Pará while the deforestation starts to move into the state of Amazonas [Wood 2002]. The tree cover loss within this arc develops along the highway network in this regions [Alves 2002; Müller et al. 2016]. Several paved highways like BR-163, BR-219, BR-230 etc. foster the agricultural and infrastructural development and lead to high deforestation rates. In Brazil an area of approximately 274 thousand  $\text{km}^2$  is deforested between 2000 and 2010. Deforestation hot-spots in Paraguay and Argentina are located in the Gran Chaco region which was one of the least disturbed forests worldwide and is exposed to high deforestation rates since 1969 [Caldas et al. 2013; Zak et al. 2004]. During our study time frame deforestation accounts for an area of approximately 7097  $\text{km}^2$  and 21 thousand  $\text{km}^2$  in the Chaco region in Argentina and Paraguay, respectively. In Bolivia the deforestation hot-spot is located in the department of Santa Cruz. Bolivia had till the 1960s relative low deforestation rates which increased moderately after and increased sharply during the 1990s and remain high as the map suggests [Pacheco 2002; David Kaimowitz and Vanclay 2002]. During the first decade of 2000 an area of approximately 19 thousand  $\text{km}^2$  is exposed to tree

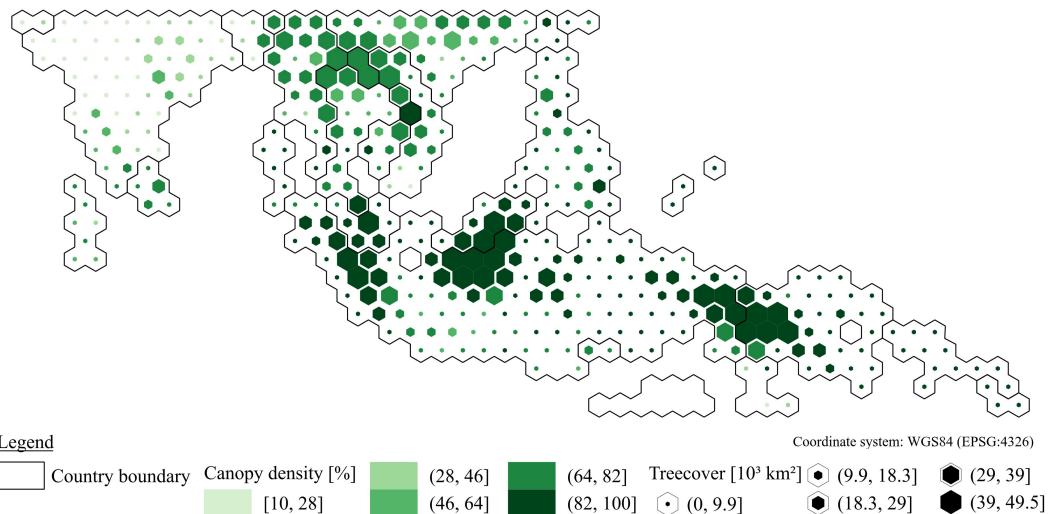
cover loss. The department of Petén in Guatemala is committed to deforestation since the 1980s [Beach 1998]. Since 2000 the deforestation rates in Guatemala are among the highest in Latin America and after 2005 the rates increased further [McSweeney et al. 2014]. During 2000 and 2010 an area of approximately  $6515 \text{ km}^2$  is exposed to tree cover loss as table C.1 shows. McSweeney et al. highlights that deforestation hot-spots often spatially overlap with drug trafficking nodes. Especially in the department of Petén where large ranches are owned by narco-traffickers. In Peru deforestation hot-spots covering the provinces of Huánuco, Loreto, San Martín, and Ucayali, while the cumulative forest loss account for an area of approximately 8 thousand  $\text{km}^2$  in the entire country. An area of approximately 15 thousand  $\text{km}^2$  is exposed to deforestation in Colombia, while deforestation hot-spots are located in Cacquetá, Bolívar, and Antioquia.



**Figure 3.3. Tree cover loss in Latin America between 2000 and 2010:** This map shows the tree cover loss within our study extent between 2000 and 2010. An unscaled hexagon covers an area of 0.5 decimal degrees which translates to an area of approximately 49 thousand  $\text{km}^2$  at the equator. Deforestation hot-spots with tree cover loss about 2.7 thousand  $\text{km}^2$  per hexagon are located in Paraguay, Argentina, Bolivia, Brazil, Colombia, Peru, and Guatemala.

The map in figure 3.4 shows the tree cover and canopy density distribution for Asia/Australia. In Asia the tropical rain forest is distributed over several southeast Asian islands like the Indonesian islands Sumatra, Borneo, Java, Papua etc., the Philippines, Malaysia, and Papua New Guinea. Generally the rain forest is characterized by high tree covers comparable to Latin America but in our map most of the islands are smaller than a single hexagon. Therefore, by our method to compute the hexagon scaling the size of the polygons don't reflect the tree cover as share of the landmass and most of them appear smaller. The mean canopy density of tropical rain forest is between 82 and 100 % and the tree cover is above approximately  $9.9 \text{ km}^2$ . Tropical moist and dry forest are distributed over the continental Asia covering the countries India, Vietnam, Cambodia, Laos etc. These forests are characterized lower tree cover among  $9.9 \text{ km}^2$  and canopy densities between 10 and 82 % while the lower tree cover

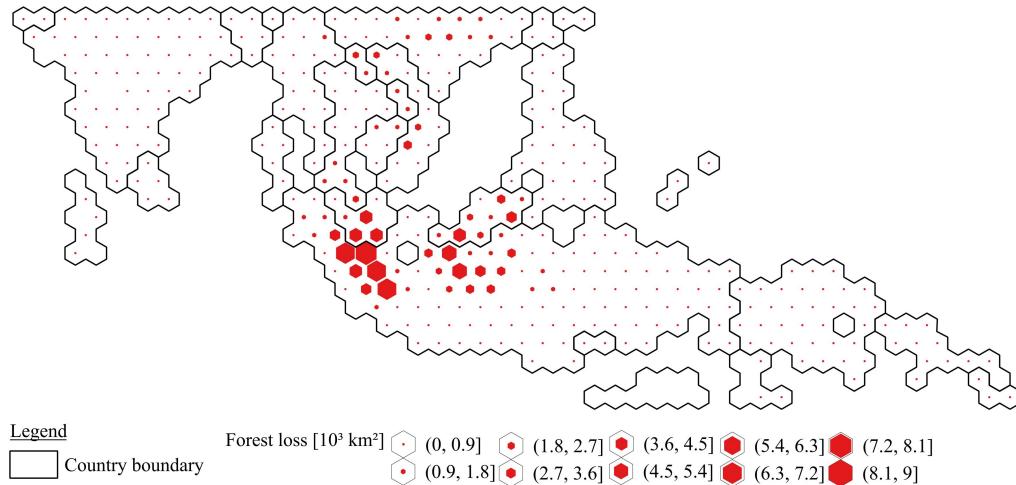
indicates that in southeast Asia deforestation occurs since the 1950s [Kummer and Turner 1994].



**Figure 3.4. Tree cover and canopy density in Asia/Australia at 2000:** This map shows the tree cover and mean canopy density distribution within our study extent at 2000. An unscaled hexagon covers an area of 0.5 decimal degrees which translates to an area of approximately 49 thousand  $\text{km}^2$  at the equator.

The map in figure 3.5 shows the distribution of tree cover loss in Asia/Australia for the time period 2000 till 2010. During this period a forest area of approximately 196 thousand  $\text{km}^2$  is exposed to deforestation as table C.1 in the appendix C suggests. For this region we identified the following countries as deforestation hot-spots with deforestation areas per hexagon of approximately 1.9 to 9 thousand  $\text{km}^2$ : Indonesia, continental and insular Malaysia, Vietnam, and Laos. Indonesia is known as a country with one of the highest rates of primary forest loss for the time period 2000 till to 2016 [Austin et al. 2019]. In the first decade of 2000 an area of approximately 100 thousand  $\text{km}^2$  is exposed to deforestation as table C.1 suggests. Tree cover loss is predominantly distributed over the Indonesian islands of Sumatra and Borneo (Kalimantan) as the map 3.5 shows. Indonesia's forests are exposed to deforestation by dynamic causes since the 1950s [Nawir and Rumboko 2007]. Malaysia has lost an area of approximately thousand 33  $\text{km}^2$  by deforestation during the time frame 2000 till 2010. The deforestation hot-spots in Malaysia are distributed over the Malaysian Borneo (Sarawak/Sabah) and the continental Malaysia. Comparable to Indonesia, Malaysia's forests are exposed to deforestation since the 1950s by dynamic causes like logging followed by agricultural activities [Kummer and Turner 1994]. From 2000 till 2010 an area of approximately thousand  $\text{km}^2$  was exposed to deforestation in Vietnam as table C.1 suggests. We identified the central highland provinces Dak Lak, Dak Nong, Gia Lai, and Lam Dong as deforestation hot-spots. Meyfroidt et al. [2013] confirms this in his study on trajectories of deforestation and highlights that the losses in this area could reduce the benefits of national-scale forest recovery. During the early 1990s a transition to reforestation was encouraged

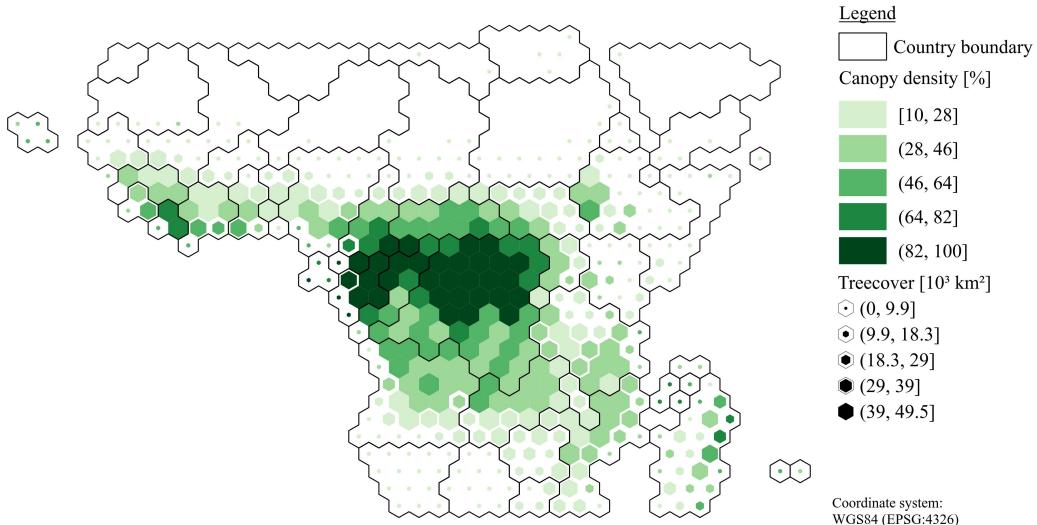
and natural forests expanded in Vietnam but it is connected with a increased loss in the highland regions [Meyfroidt et al. 2013; Chazdon 2008]. In Laos a forest area of approximately 8 thousand km<sup>2</sup> is lost during the time period 2000 till 2010. The map in figure 3.5 suggests that northern Laos is predominantly exposed to tree cover loss while Hirsch [2000] confirms. Laos is known for a steady loss of forest cover since the early 1960s. Whereas the causes for this LC transitions are dynamically changing till now.



**Figure 3.5. Tree cover loss in Asia/Australia between 2000 and 2010:** This map shows the tree cover loss within our study extent between 2000 and 2010. An unscaled hexagon covers an area of 0.5 decimal degrees which translates to an area of approximately 49 thousand km<sup>2</sup> at the equator. Deforestation hot-spots with tree cover loss about 1.9 thousand km<sup>2</sup> per hexagon are located in Indonesia, Malaysia, Vietnam, and Laos.

Figure 3.6 shows the tree cover and canopy density distribution over Africa. In Africa the tropical rain forest is distributed over the following central African countries: Democratic Republic of the Congo, Republic of the Congo, Equatorial Guinea, Gabon, Cameroon, and partial in Ghana, Ivory Coast, and Liberia. In our map this type of forest is characterized by a dense tree cover between 39 and 49 thousand km<sup>2</sup> per hexagon while the canopy density between 64 and 100 % does not separate it from the moist type as good as in Latin America and Asia/Australia. The rain forest is surrounded by tropical moist forest with a dense tree cover between 18 and 49 thousand km<sup>2</sup> while the mean canopy density between 10 and 82 % is sparser than tropical rain forest. Countries covered by the moist tropical forest type are the following examples: Angola, Uganda, Central African Republic, Zambia, Madagascar etc. Tropical dry forest is characterized by a sparse tree cover among approximately 18 thousand km<sup>2</sup> per hexagon and canopy densities between 10 and 46 %. Mozambique, Tanzania, and Nigeria are examples for countries covered partial by tropical dry forest.

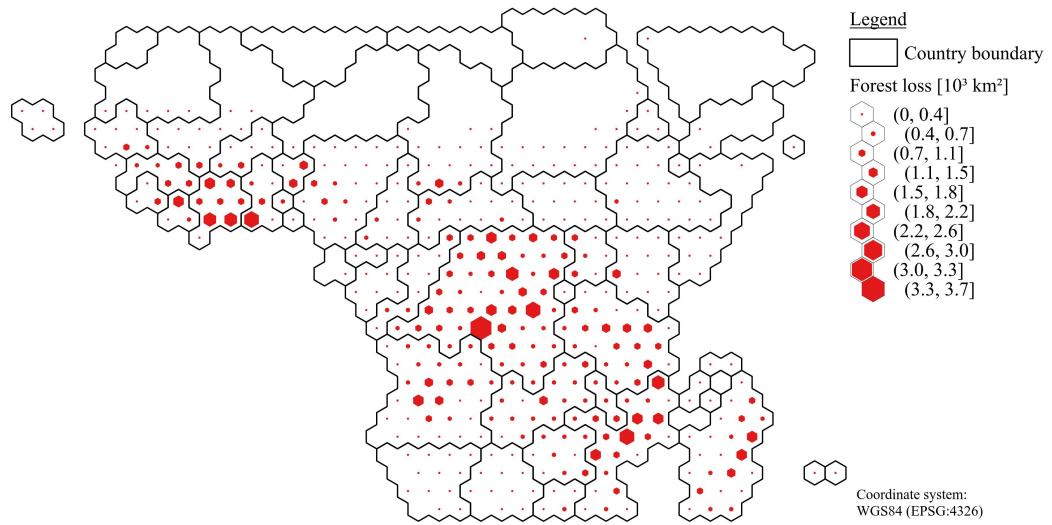
The map in figure 3.7 shows the regions which are exposed to tree cover loss in Africa for the time period 2000 till 2010. During the first decade of 2000 an area of approximately 174 km<sup>2</sup> was deforested as table C.1 in the appendix C suggests. With tree cover loss between 1.1



**Figure 3.6. Tree cover and canopy density in Africa at 2000:** This map shows the tree cover and mean canopy density distribution within our study extent at 2000. An unscaled hexagon covers an area of 0.5 decimal degrees which translates to an area of approximately 49 thousand km<sup>2</sup> at the equator. Tropical rain forest is characterized by dense tree cover between 39 and 49 thousand km<sup>2</sup> and high canopy density above 64 % in the center of Africa. The rain forest is surrounded by tropical moist forest with a dense tree cover as well but a lower mean canopy density between 10 and 82 %.

and 3.7 thousand km<sup>2</sup> per hexagon we identified the following countries as deforestation hot-spots: Ivory Coast, Democratic Republic of the Congo, Angola, Mozambique, Madagascar, and Tanzania. At the Ivory Coast an area of approximately 10 thousand km<sup>2</sup> was exposed to deforestation between 2000 and 2010 as table C.1 shows. The map in figure 3.7 shows that the deforestation hot-spots are located in the southern and northern parts of the country. Evidences for this are given by Goetze et al. [2006] and Barima et al. [2016]. Lower deforestation in the center of the country can be explained by a military conflict started in 2002 that divides the country in a northern and southern zone with a buffer zone in between controlled by Un forces and French soldiers [Barima et al. 2016]. At the Ivory coast continuous loss of tree cover by deforestation started approximately in the 1958s [Chatelain et al. 1996]. The deforestation accounts for an area of approximately 45 thousand km<sup>2</sup> in the Democratic Republic of the Congo. The DR Congo home of the second largest tropical forest in the world has compared to other tropical countries relatively low deforestation dynamics but recent studies show that deforestation could accelerate in the future [Ickowitz et al. 2015]. The deforestation hot-spots are located in the eastern Congo and around medium-sized cities along the Congo river in the lower center of the country. During the first decade of the 2000s an area of about 13 thousand km<sup>2</sup> is deforested in Angola. Deforestation hot-spots are more oriented to the center of the country covering roughly the provinces Huambo, Bie, and Moxico as the map in figure 3.7 suggests. For the time period 1990 till 2009 Cabral et al. [2011] performed a study on forest change in the province of Huambo where a decrease in dense forest cover is observed while the cover of sparse forest is increasing [Cabral et al. 2011]. Another study by Schneibel et al. [2017] on deforestation dynamics in south-central Angola

reveals that the tree cover loss develops along anthropogenic infrastructure and forests are exploited over a long term until (fuel-wood collection etc.) a LC transitions to other types like cropland occurs. To best of our knowledge no studies on historic deforestation dynamics exist for Angola due to the civil war from 1975 till 2002. In Mozambique an area of approximately 18 thousand km<sup>2</sup> was exposed to tree cover loss as table C.1 suggests. During the war between 1976 and 1992 the forests of Mozambique were not largely exposed to deforestation but since end of the war the deforestation rates are increasing [Sitoé et al. 2012]. We identified the following provinces as deforestation hot-spots between 2000 and 2010: Zambezia, Nampula, and Cabo Delgado. This is largely confirmed by Sitoé et al. where it is stated that deforestation is concentrated in the center and north of Mozambique. These hot-spots could be related to the higher population densities within this areas. Tree cover loss accounts for an area of about 11 thousand km<sup>2</sup> in Madagascar during the study period. In Madagascar deforestation hot-spots are concentrated at the north-east coast of the island as the map in figure 3.5 suggests. Madagascar's central highland forests are exposed to anthropogenic deforestation since 1600, reportedly [Harper et al. 2007]. By the nineteenth century the deforestation advanced over the entire island forests. In Tanzania an area of approximately 14 thousand km<sup>2</sup> was exposed to tree cover loss. Since the 1900s approximately 19.4 % of the forest cover is lost [Kideghesho 2015]. The map in figure 3.7 suggests that deforestation hot-spots are located in the country center covering the following provinces: Lindi, Singida, Tabora, Dodoma, Tanga, Pwani, and the island of Zanzibar.



**Figure 3.7. Tree cover loss in Africa between 2000 and 2010:** This map shows the tree cover loss within our study extent between 2000 and 2010. An unscaled hexagon covers an area of 0.5 decimal degrees which translates to an area of approximately 49 thousand km<sup>2</sup> at the equator. Deforestation hot-spots with tree cover loss about 1.1 thousand km<sup>2</sup> per hexagon are located in Ivory Coast, Democratic Republic of the Congo, Angola, Mozambique, and Tanzania.

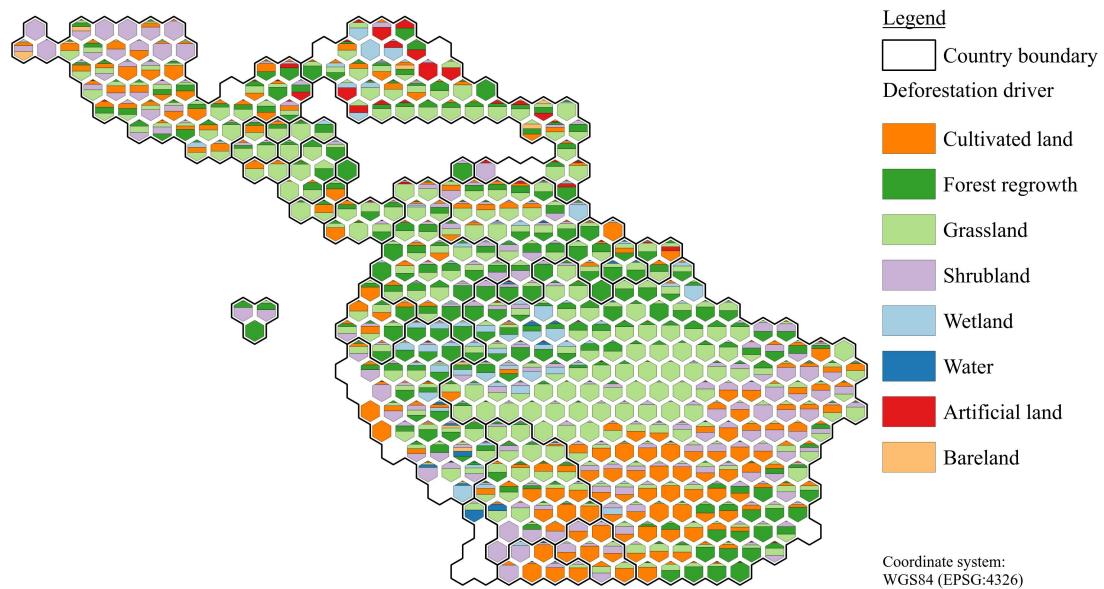
### 3.1.3. Proximate deforestation drivers

Our goal is to estimate the distribution of PDDs over the tropical zone, the continental range, and at a country scale for the time frame of 2000 till 2010. We achieved this estimate by superimposing the annual tree cover losses and aggregated gains of the GFC datasets and the GL30 land cover map from 2000. We carefully selected our global definition of tree cover in the canopy density interval  $(10, 100]$  by applying the Jaccard Index and statistical testing detailed in section 3.1.1. By using this canopy density interval we filtered the GFC annual losses and we considered tree cover gains only within previously lost tree cover. After superimposing we applied a reclassification of tree cover losses still classified as forest by the GL30 layer. We aggregated this structures by clustering and applied a square sized buffer of 500 meter side length. Next, we reclassified the structure by determining the highest frequent LC class within the buffer. For further details on the PDD prediction refer to section 2.2.2.2. The PDD distribution choropleth-maps for Latin America, Asia/Australia, and Africa we derived by applying our hexagonal-binning approach in combination with a hexagon-pie-chart. A hexagon in unscaled shape covers an area of 0.5 decimal degrees. Section 2.2.5 describes detailed how we derived these cartograms. In the following paragraphs we will describe our results for Latin America, Asia/Australia, and Africa on continental and regional scale and close this section by discussing our results on global scale.

In Latin America the transition of tree cover to cropland and pastures account for 21.3% ( $95929.6 \text{ km}^2$ ) and 40.8% ( $183841.4 \text{ km}^2$ ) as the table C.1 in the appendix C shows. Therefore, approximately 62% ( $279771 \text{ km}^2$ ) of the tree cover were cleared for agricultural purposes in Latin America. Around 12.6% ( $56909.3 \text{ km}^2$ ) of the tree cover loss is followed by tree cover regrowth, while the transition to shrubland account for 11.1% ( $50260.2 \text{ km}^2$ ) of the tree cover loss, respectively. Minor PDDs are water, artificial surfaces, and bareland which account for 1.6% ( $7169.9 \text{ km}^2$ ), 0.3% ( $1561.5 \text{ km}^2$ ), and 0.1% ( $405.4 \text{ km}^2$ ) of the forest transitions. Sy et al. [2015] estimates that agriculture accounts for 88.5% of the tree cover loss, while transitions of forest cover to other LC accounts for 11.5%. Further, Hosonuma et al. [2012] achieve similar results for the conversion of tree cover to agricultural land where it accounts for approximately 90% of the tree cover loss. Both studies estimate that the expansion of artificial surfaces account for approximately 1% of the forest loss. The large difference could be explained by the definition of the forest transition classes in both studies. In both studies agriculture comprises pastures, cropland, and tree plantations. Additionally, both studies refer to the change of LU and we determine the PDDs by the change of LC. Further, Sy et al. uses a sample based approach on  $10 \times 10 \text{ km}$  FAO FRA-2010 RSS data which could yield overestimates, while Hosonuma et al. uses an empirical approach based on FAO data as well. A recent study on PDDs estimates that commodity-driven deforestation, shifting agriculture, forestry, wildfire, and urbanization account for 56%, 31%, 13%, 1%, and

<1% of the tree cover loss, respectively [Curtis et al. 2018]. Commodity-driven deforestation refers to tree cover loss as long-term permanent transition of forest to a non-forest LU like agriculture which includes cropland, pastures, plantations and so forth. Comparable to the previously mentioned research on PDDs the difference between our estimates and Curtis et al. arise from class definitions. If we would aggregate our LC classes to their schema it would yield nearly the same results. In particular the same estimate for tree cover loss by urbanization shows that the similar classes without aggregation yield similar results. In Latin America the LC change to cropland is mainly distributed over southern part, while the transition to pastures is concentrated in the central part of the continent as figure 3.8 suggests. Cropland expansion mainly took place in the south of Brazil, Paraguay, Argentina, and Bolivia, while forest loss by pasture expansion is concentrated in the center and north of Brazil. The spatial distribution of cropland/pasture dynamics largely corresponds to the findings of Graesser et al. [2015]. Large quantities of tree cover regrowth can be observed in the southeast and center of Latin America namely the southeast coast of Brazil and within the tropical rain forest covering Brazil, Peru, and Colombia. In south-east Brazil the findings of Curtis et al. suggests that the regrowth dynamics are driven by forestry actions, while the coastline is exposed to shifting agriculture. Cropland and pasture expansion account for 19.1% and 49.7% of the tree cover loss, while regrowth dynamics and artificial surfaces account for 11.8% and 0.3% of the tree cover loss in Brazil. Figure 3.8 shows that deforestation by cropland expansion is mainly concentrated in the southern part of Brazil in the provinces Mato Grosso, Goiás, and Mato Grosso do Sul, which is largely confirmed by Zalles et al. [2018] and Graesser et al. [2015]. Additionally, in the northern part of Mato Grosso and the southern part of Mato Grosso do Sul deforestation by grassland expansion can be observed [Graesser et al. 2015; Sy et al. 2015]. The tree cover loss in the arc of deforestation can be attributed to pasture expansion, which is confirmed by Sy et al. [2015] and Graesser et al. [2015]. For the Chaco region of Paraguay the main PDD is the expansion of cropland, which accounts for 48.9% of the tree cover loss. The findings of Graesser et al. [2015] and Caldas et al. [2013] suggest that the main PDD in this region is the expansion of pastures, while findings by Graesser et al. [2018] suggests that pastures are largely replaced by cropland LU between 1990 and 2015. Therefore, the initial cause for tree cover loss could be the expansion of pastures but the LU changed already to cropland at our image date of 2010. In the Argentinian part of the Chaco cultivated and grassland account for 65.8% and 5.3% of the tree cover loss. This could also be observed by Sy et al. [2015]. In Bolivia cropland and pasture expansion account for 37.2% and 22.4% of the forest loss, while transitions of forest cover to artificial surfaces account for 0.4% of the tree cover loss, respectively. Main PDD for the deforestation hot-spot in the province Santa Cruz is the expansion of cropland as figure 3.8 suggests. Further, in north Bolivia and at the Brazilian border the deforestation is driven by pasture expansion. Both patterns are largely confirmed by Graesser et al. [2015] and Sy et al. [2015]. In Guatemala cropland, pasture, and artificial expansion account for 23.9%, 38.6%, 0.5% of

the forest loss, respectively. The main PDDs for the deforestation hot-spot in the province of Petén are pasture expansion followed by cropland expansion as the map 3.8 suggests. Pasture expansion is the main force for deforestation in south of Guatemala. Regrowth dynamics in Guatemala which account for 12% of the forest loss could be attributed to the establishment of oil palm plantations and shifting agriculture [Furumo and Aide 2017; Curtis et al. 2018]. In Peru cropland, pasture, and artificial expansion account for 6.6%, 23%, and 0.2% of the forest loss, while regrowth dynamics account for 28.2% of the forest loss, respectively. For the deforestation hot-spots located in the provinces of Huánuco, San Martín, and Ucayali the major PDDs are the expansion of pastures and regrowth dynamics. [Sy et al. 2015] findings suggests that the expansion of cropland is the main deforestation driver in this central region of Peru. Regrowth dynamics in this region could be related to deforestation for oil palm plantations as the findings of Vijay et al. [2018] and Furumo and Aide [2017] suggest. Vijay et al. estimates that an area of approximately 845 km<sup>2</sup> is cleared for oil palm plantations between 2007 till 2013, while Furumo and Aide states that approximately an area of 156 km<sup>2</sup> forest is cleared for palm oil plantations between 2001 and 2014. Cultivated land, grassland, and regrowth dynamics account for 6.2%, 33.4%, and 24.7% of the tree cover loss in Colombia, respectively. For the Colombian deforestation hot-spot located in the province of Cacquetá the expansion of pastures is the main PDD, which is confirmed by Graesser et al. [2015]. This is also the case for the deforestation hot-spots located in the provinces of Bolívar and Antioquia.

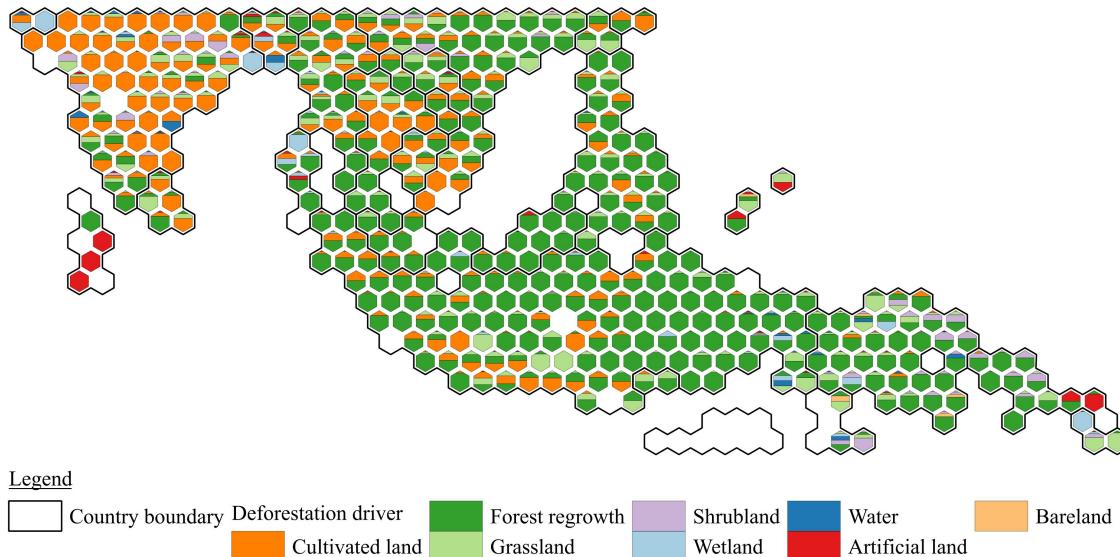


**Figure 3.8. Map of proximate deforestation drivers in Latin America:** The map shows the distribution of proximate deforestation drivers in Latin America. The different sized and colored quantities within each hexagons interior shows the relative tree cover loss by a proximate deforestation driver. Scaling of a the hexagons is only intended for improving visual appeal.

For Asia/Australia the table C.1 in the appendix C shows cultivated and grassland account for 16% (36819.3 km<sup>2</sup>) and 7.1% (16302.6 km<sup>2</sup>) of the tree cover loss, respectively. There-

fore, the transition of LC to agriculture usage accounts for 23.1% ( $53121.9 \text{ km}^2$ ) of the tree cover losses. The transition of tree cover to artificial surfaces and the forest loss by inundation by lakes and rivers account for 1.1% ( $2431.9 \text{ km}^2$ ) and 0.4% ( $890 \text{ km}^2$ ), respectively. In Asia/Australia the largest PDD is the regrowth which accounts for 61.2% ( $140653.4 \text{ km}^2$ ) of the cumulative tree cover loss. Forest transitions to shrubland account for 6.7 % (), respectively. Hosonuma et al. [2012] estimates that LC transitions by agriculture account for approximately 70% of the forest loss, while Curtis et al. [2018] estimates that commodity-driven deforestation, shifting agriculture, forestry, wildfire, and urbanization account for 13%, 78%, 9%, 13%, <1%, and <1% of the tree cover loss in Southeast Asia, respectively. As mentioned in the previous paragraph the differences in PDDs estimates relate mainly to the applied methodology and the aggregation of LC classes. The figure 3.9 shows an overview of the PDDs distribution in Asia/Australia and indicates that regrowth dynamics are largely concentrated at the east of Asia/Australia namely in the following countries: Indonesia, Malaysia, Philippines, and Papua New Guinea. Conversion of tree cover to cropland can be observed in Vietnam, Cambodia, the north of Thailand, and India in its entire extent. Curtis et al. predicts that in Indonesia and Malaysia tree cover loss is largely driven by commodity-driven deforestation, while in Vietnam and Cambodia the forest loss is driven by commodity-driven deforestation and forestry. In Indonesia cropland and pasture expansion accounts for 15.4% and 4.6% of the forest cover loss, while regrowth dynamics account for the largest share of forest loss with 68.6%. In Malaysia comparable to Indonesia regrowth dynamics account for the largest share of forest loss with 79.4%. Cropland and grassland expansion account for 7.5% and 2.6% of the tree cover loss, respectively. In Vietnam the PDDs cropland and pastures account for 32.8% and 18.3% of the tree cover loss, respectively. Further, regrowth dynamics account for 30.8% of the tree cover loss. Cropland and pasture expansion account for 14.8% and 16.4% of the tree cover loss, while regrowth dynamics account for 57% of the forest loss in Laos.

In Africa the table C.1 in the appendix C suggests that cropland and grassland account for 22.8% ( $44289.5 \text{ km}^2$ ) and 46% ( $89301.4 \text{ km}^2$ ) of the tree cover loss, respectively. Therefore, transitions to agricultural land are the major PDD in Africa and account for 68.8% ( $133590.9 \text{ km}^2$ ) of the tree cover loss. Forest transitions to water, artificial surfaces, and bareland account for 1.2% ( $2409.9 \text{ km}^2$ ), 0.6% ( $1238.6 \text{ km}^2$ ), and 0.1% ( $146 \text{ km}^2$ ) of the tree cover loss, respectively. In Africa on 17% ( $32980.8 \text{ km}^2$ ) of the area exposed to tree cover loss regrowth could be detected, while the transition to shrubland account for 3.4% ( $6599.5 \text{ km}^2$ ) of the tree cover loss, respectively. The study on PDDs by Hosonuma et al. [2012] estimates that agriculture and urbanization account for approximately 75% and 2% of the tree cover loss, respectively. For Africa Curtis et al. [2018] shows that commodity-driven deforestation, shifting agriculture, forestry, wildfire, and urbanization account for 4%, 92%, 4%, <1%, <1% of the tree cover loss. The figure 3.10 shows the distribution of PDDs in Africa and indicates



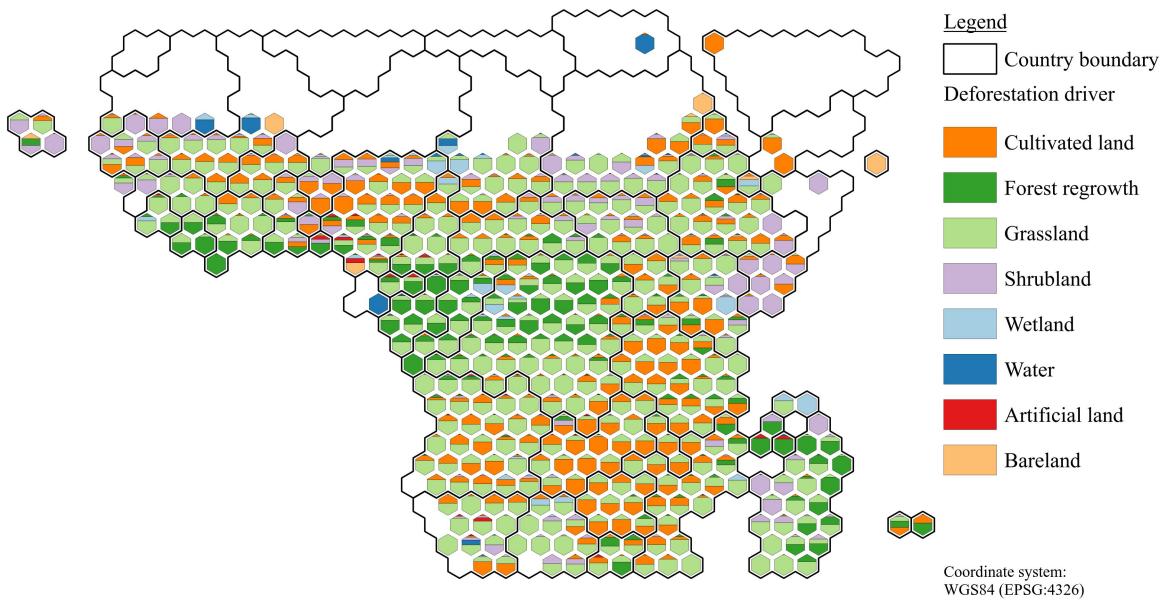
**Figure 3.9. Map of proximate deforestation drivers in Asia/Australia:** The map shows the distribution of proximate deforestation drivers in Asia/Australia. The different sized and colored quantities within each hexagons interior shows the relative tree cover loss by a proximate deforestation driver. Scaling of a the hexagons is only intended for improving visual appeal.

that the forest loss in the east of the continent is largely driven by the expansion of cultivated land. This can be observed in the following east African countries: Tanzania, Mozambique, Zambia, Malawi, and Zimbabwe. The central African countries like Democratic Republic of the Congo, Central African Republic, Congo, and South Sudan are mainly exposed to expansion of pastures as the map suggest. The findings of Curtis et al. suggest that overall countries in Africa the major deforestation driver is shifting agriculture. At the Ivory coast cropland and pasture expansion account for 12.5% and 46.9% of the forest loss, respectively. The expansion of cropland and pastures account for 9.4% and 52% of the forest loss, while regrowth dynamics account for 26.6% of the tree cover loss in the Democratic Republic of the Congo. In Angola cultivated and grassland account for 32.7% and 54.7% of the deforestation, respectively. Cropland and pasture expansion account for 32.7% and 45.6% of the tree cover loss, respectively. Largest deforestation driver in Tanzania are cropland and pastures, which account for 42.1% and 40.6% of the tree cover loss. In Madagascar regrowth dynamics account for 40.6% of the tree cover loss, while grassland account for 44.6% of forest loss.

On the global range the major PDDs are cropland and pastures which account for 20.2% ( $177038.5 \text{ km}^2$ ) and 33.1% ( $289445.5 \text{ km}^2$ ) as the table C.1 in the appendix C suggests.

### 3.1.4. Accuracy assessment

**Goal (review):** Goal is the assessment of the accuracy of our proximate deforestation driver predictions. We created a set of ground truth data by sampling our proximate deforestation



**Figure 3.10. Map of proximate deforestation drivers in Africa:** The map shows the distribution of proximate deforestation drivers in Africa. The different sized and colored quantities within each hexagons interior shows the relative tree cover loss by a proximate deforestation driver. Scaling of a the hexagons is only intended for improving visual appeal.

driver layers. In each continental region we select per random 10 tiles and draw 200 samples per tile. The 200 samples comprises pixels over the full value range of our proximate deforestation driver classes. We imported the prepared sample to our JavaScript application and subsequently classified each sample with a label by visual interpretation of google earth imagery. To determine the accuracy we used a confusion matrix and the derived metrics.

**Results (review):** Table 3.5 shows the confusion matrix to determine the accuracy of our predictions where the term reference refers to the labeling of pixel by our visual interpretation and predictions refer to the labeling of our proximate driver predictions. The abbreviations PAc, UAc, OvAc, Com, Om, Tot, and Kappa refer to the terms Producers-Accuracy, Users-Accuracy, Overall-Accuracy, Error of Commission, Error of Omission, row or column total, and Kappa Coefficient. From the 6000 samples we draw from our study extent 14 %, 20 %, 22 %, 32 %, 8 %, 2 %, 0.5 %, 2 %, and 0.5 % account for cultivated land (10), tree cover (20), regrowth (25), shrubland (40), wetland (50), water (60), artificial land (80), and bareland (90), respectively. Our method predicts a distribution of 15 %, 18 %, 27 %, 31 %, 7 %, 1 %, 0.8 %, 1 %, and 0.2 % for the land cover classes 10, 20, 25, 30, 40, 50, 60, 80, and 90, respectively. Highest producers accuracy is achieved the regrowth class with 88 percent. The prediction of this class was achieved by including global forest change gain data within our target canopy density. During building our reference data set per visual interpretation we determined this class by following these rules. We checked the surroundings and the corresponding pixel for signs of road networks and infrastructure. Further we checked if the canopy shows signs of age class forest and line patterns which show the establishment of

artificial introduced forest cover. During the visual interpretation we recognized that a large portion of the regrowth class is occupied by plantations especially in Americas and Asia. For Asia a major share was occupied by palm oil or other plantations. It must be mentioned that here the deforestation within our temporal frame is not coercively the clearing of natural cover/primary forest it shows also rotational cycles of deforestation reforestation. The second highest accuracy with 85 (producers accuracy) was achieved in the prediction of cultivated land where only 15 percent could be identified as error of commission. 8.2 percent are classified as forest or regrowth which could reveal zones of shifting agriculture or temporal issues. To identify this class during our visual interpretation we applied nearly the same rules as the detection of regrowth checking for infrastructure but here no canopy should be identified. Further we checked for large scale square patterns which is shown by large scale commodity driven agriculture. Especially the small scale agriculture was hard detect because these class of agriculture often show no regular patterns of forest clearing and are often accompanied by selective logging. Further it should be mentioned common practice of small scale agriculture is inhomogeneous over the regions Americas, Asia and Africa. Another good indicator for cultivated land is the appearance of tillage patterns on high resolution imagery. Cultivated land comprises a large variety of different use cases. Grassland shows an accuracy of 77 at an error of commission of 23 percent. The error could lead back to classification errors by GlobeLand30 or land cover change dynamics and the temporal time frame. For visual interpretation cultivated land and grassland share some common properties like square shaped clearings and infrastructure in the surroundings. But grassland is also defined by the occurrence of shrubs and single trees within its extent. To really determine the differences between both classes high resolution imagery is required. In Americas the grassland patches could often be identified as used for cattle ranching. During visual interpretation we recognized frequently the occurrence of water holes. Africa showed the occurrence of natural grasslands and Asia bla.

## 3.2. Emissions

**Goal (review):** Our goal is to estimate the emissions emitted by proximate deforestation driver. We estimated the emissions from the removal of aboveground biomass and through the change of soil organic carbon content. Detailed below are the pdd classes we assumed as deforestation to estimate the agb emissions: 10, 25, 30, 40, 80, 90. To approximate the emissions by soil organic carbon change we used the transition pathways from don et al. We estimated soc for two different scenarios with the following conditions: scenario one (SC1) all transitions occur in primary forest and scenario two (SC2) by using the intact forest landscape layer we distinguished between primary forest and secondary forest. The

**Table 3.5. Confusion matrix for accuracy assessment:** We draw 6000 samples from 10 random selected tiles from the three regions Americas, Asia and Africa. Labels refer to our proximate deforestation driver classes which correspond to GlobeLand30 classification schema in table 2.2. Reference refers to the samples we classified by visual interpretation of external imagery and predictions refer to the label the sample has in our proximate driver product. The abbreviations PAc, UAc, OvAc, Com, Om, Tot, and Kappa refer to the terms Producers-Accuracy, Users-Accuracy, Overall-Accuracy, Error of Commission, Error of Omission, row or column total, and Kappa Coefficient.

		Reference											
		10	20	25	30	40	50	60	80	90	Tot	UAc	Om
Prediction	10	730	37	62	15	16	2	3	5	0	870	.84	.16
	20	41	744	56	189	31	12	0	15	4	1092	.68	.32
	25	29	202	1155	172	22	10	5	11	4	1610	.72	.28
	30	36	187	32	1466	73	21	0	17	0	1832	.80	.20
	40	14	21	4	41	352	1	1	2	1	437	.81	.19
	50	0	5	3	10	4	50	0	1	0	73	.68	.32
	60	2	1	0	3	0	2	18	2	0	28	.64	.36
	80	3	3	0	1	1	1	0	40	0	49	.82	.18
	90	0	0	0	1	0	0	0	3	5	9	.56	.44
	Tot	855	1200	1312	1898	499	99	27	96	14	6000		
	PAC	.85	.62	.88	.77	.71	.51	.67	.42	.36	Kappa		OvAc
	Com	.15	.38	.12	.23	.29	.49	.33	.58	.64	.69		.76

first scenario can be seen as liberal estimate and the second scenario as conservative estimate. We aggregated our results on global and continental scale.

### 3.3. Ecosystem service values

Our goal is to estimate the monetary loss of ESV by the aggregated PDD at a global and continental scale between 2001 and 2010. Whereas a study targeting regional scale aggregation could easily performed with our data. Additionally, we approximated the monetary gain if the former tree cover is converted to a certain type of LC. We refer to this gain as ESV gain. Further, we compute the balance between ESV loss and gain to estimate the net change of ESV in the tropical zone. We applied three datasets which estimate the monetary value of ecosystems on a global scale namely the following data: de Groot et al. [2012], Costanza et al. [2014], and Siikamaki et al. [2015]. We refer to this datasets as Co, Dg, and Wb. To compute the monetary loss of ESV we selected the following PDD classes as anthropogenic deforestation: cultivated land (10), regrowth (25), grassland (30), shrubland (40), artificial surfaces (80), and bareland (90). We excluded pixel classified as forest (20) by our PDD prediction because within this class we are uncertain if a deforestation event occurred. Further, we excluded transitions to wetland or water bodies because we assume this LC changes are largely driven by natural causes. The cumulative ESV gain depends on the particular ESV dataset because the datasets don't define a monetary value for each PDD class. For detailed

informations on the methodology refer to section 2.2.4. The ESV net change or balance is the difference of cumulative ESV loss and gain. The monetary unit for each value is the Geary–Khamis dollar at 2007 per year also known as international dollar ( $2007 \text{ Int.\$ } y^{-1}$ ). The table 3.6 shows the gross ESV loss and gain and the net balance for continental and global scale. We will discuss the ESV losses exclusively for the Co and Wb datasets because the difference between Co and Dg is just 118 international dollar per year. Further, we discuss the ESV gain for all three datasets because they provide a range of different estimates per biome.

In Latin America a forest loss of approximately 396 thousand  $\text{km}^2$  accounts for 208.3, 204, or 51 billion  $2007 \text{ Int.\$ } y^{-1}$  of ESV loss as table 3.6 suggests. By applying the Costanza et al. [2014] monetary value for tropical forest the transition of tree cover to cropland accounts for a loss of 51 billion  $Int'\text{\$ } y^{-1}$ , while transitions to pastures account for a loss of approximately 126 billion dollar per year. The tropical deforestation by the expansion of artificial and bareland surfaces account for a loss of 1 billion  $Int'\text{\$ } y^{-1}$ , while regrowth dynamics convert to a loss of 30 billion dollar. By using the monetary value of tropical forest from Siikamaki et al. [2015] the transitions to cropland and pastures cost 12.6 and 30.6 billion dollar, respectively. Further, the aggregated urbanization and bareland transition accounts for 0.3 billion dollar per year, while regrowth dynamics account for 7.5 billion international dollar per year. The variability in the estimates is related to the difference of monetary value for tropical forest between both datasets. Costanza et al. monetary prediction for tropical forest is approximately 4 times greater than Siikamaki et al.. We estimated a gross monetary gain of 161 billion dollar per year for the Co dataset. The table 2.9 in section 2.2.4 suggests that we calculated the monetary gain from LC transitions to cropland, grassland, artificial surfaces, and regrowth which account for 53.4, 76.6, 1, and 30.6 billion dollar per year, respectively. By applying de Groot et al. [2012] ESV estimates the gross gain from LC transitions account for 82.5 billion dollar per year. This converts to a monetary gain from grassland and regrowth transitions of approximately 52.8 and 29.9 billion dollar. The last dataset we tested enabled us to compute only the ESV gain from tree cover regrowth which accounts for a monetary gain of 7.3 billion dollar per year, respectively. In Latin America the net balance of ESV changes account for -47.3, -121.5, and -43.7 billion dollar for Co, Dg, and Wb, respectively. The negative signs indicates that overall three datasets the we could observe a net loss of ESV in Latin America for the time period 2000 till 2010. The highest net loss can be observed for the Dg dataset but this relate to the small number of LC transitions defined by a monetary value.

For our study period in Asia/Australia tree cover loss of approximately 230 thousand  $\text{km}^2$  accounts for an ESV loss of 111.6, 109.1, or 27.2 billion dollar per year for the three datasets as table 3.6 shows. The first dataset predicts that forest cover transitions to cultivated- and grassland account for a loss of 19.8 and 9.8 billion dollar per year, respectively. Further, the

**Table 3.6. Ecosystem service value balance:** The dataset column refers to the three global ecosystem service value datasets by Costanza et al. [2014] (Co), de Groot et al. [2012] (Dg), and Siikamaki et al. [2015] (Wb). Loss refers to the monetary value of tropical forest deforested by the following aggregated anthropogenic proximate deforestation drivers: cropland, grassland, regrowth, shrubland, artificial surfaces, and bareland. Gain is the monetary value of the land cover transition and balance is difference of gain and loss. Monetary unit of the ecosystem service values is  $10^9$  2007 Int'l\$  $y^{-1}$  (billion international dollar 2007 per year).

Dataset	Latin America	Asia/Australia	Africa	Global
Co <sub>loss</sub>	208.3	111.6	94.2	414.1
Co <sub>gain</sub>	161.0	109.1	80.6	350.6
Co <sub>balance</sub>	-47.3	-2.5	-13.6	-63.5
Dg <sub>loss</sub>	204.0	109.1	92.1	405.2
Dg <sub>gain</sub>	82.5	83.1	43.1	208.7
Dg <sub>balance</sub>	-121.5	-26.0	-49.0	-196.5
Wb <sub>loss</sub>	51.0	27.2	22.9	101.1
Wb <sub>gain</sub>	7.3	19.4	4.3	31.0
Wb <sub>balance</sub>	-43.7	-7.8	-18.6	-70.1

expansion of artificial surfaces account for a loss of 0.5 billion dollar, while the regrowth dynamics account for a loss of approximately 75.7 billion dollar per year. By applying the monetary ESV estimate for tropical forest of Siikamaki et al. [2015] cropland and pastures account for a loss of 4.8 and 2.4 billion dollar per year, respectively. The transition of forest cover to artificial surfaces account for 0.2 billion dollar, while regrowth dynamics account for 18.4 billion dollar. In regards of ESV gain dynamics and by applying the first dataset cropland and pastures account for a monetary gain of 20.5 and 6.8 billion dollar per year, respectively. Further LC transitions to regrowth and artificial surfaces account for a gain of 75.7 and 0.5 billion dollar. By considering de Groot et al. [2012] estimates grassland and regrowth account for a gain of approximately 4.6 and 74 billion dollar per year, respectively. By applying the last dataset in table 3.6 regrowth account for a gain of 19.4 billion dollar per year. In summary by applying the three datasets we estimated a ESV gain of 80.6, 43.1, and 4.3 billion dollar per year, respectively. In regards of the net ESV balance in Asia/Australia the three datasets account for a net loss of -2.5, -26, and -7.8 billion dollar per year, respectively. The low net losses can be explained by the high share of regrowth for this region. In Asia/Australia the high share of regrowth could be the establishment of plantations or the rotational cycle as management practice of existing plantations as explained in the section 3.1.3 on PDDs. It is to assume that if we would use for regrowth the monetary value of cultivated land from the first dataset the net balance could have a positive sign.

For Africa the table 3.6 shows that a tree cover loss of approximately 177 km<sup>2</sup> account for a ESV loss of 94.2, 92.1, and 22.9 billion dollar per year for the three datasets, respectively. For the first dataset tree cover losses by cultivated- and grassland expansion convert to a loss of 23.7 and 51.5 billion dollar, while the expansion of artificial surfaces account for 0.7 billion dollar. In Africa regrowth dynamics are responsible for a loss of approximately

17.7 billion dollar per year. By applying the monetary value for tropical forest of Siikamaki et al. [2015] cropland and pasture expansion account for a loss if 5.8 and 12.7 billion dollar per year, respectively. The transition to artificial surfaces account for 0.2 billion dollar, while regrowth dynamics account for a loss of 4.3 billion dollar. The monetary gain by LC transitions account for 80.6, 43.1, and 4.3 billion dollar per year, respectively. For the Co-dataset the forest transition to cropland, pastures, artificial surfaces, and regrowth dynamics account for 24.6, 37.2, 0.8, and 17.7 billion dollar of the total ESV gain. By applying the second dataset grassland and regrowth dynamics account for a gain of 25.6 and 17.4 billion dollar, while the last dataset estimates a gain of approximately 4.3 billion dollar for regrowth dynamics. In Africa the net balance is negative and account for a net loss of -13.6, -49, and -18.6 billion dollar overall datasets.

The highest ESV losses can be observed in Latin America followed by Asia/Australia and Africa. This ESV loss dynamic is mainly attributed to the large area deforested in Latin America. If we normalize the Co ESV gain by dividing the gross gain by the assumed cumulative anthropogenic deforestation the normalized gain account for 414, 553, and 462 thousand Int'I\$  $y^{-1} km^{-2}$  for Latin America, Asia/Australia, and Africa, respectively. Therefore, the highest ESV gain per deforested  $km^2$  of tropical forest can be observed in Asia/Australia followed by Africa, while Latin America has the lowest ESV gain per  $km^2$ . This can be attributed to the large share of regrowth dynamics in Asia/Australia. In Africa the ESV gain is mainly driven by large share of agricultural transitions and regrowth dynamics. The ESV net losses are smallest in Asia/Australia followed by Africa, while Latin America has the greatest net losses. Differences in ESV estimates between the datasets of Costanza et al. [2014], de Groot et al. [2012], and Siikamaki et al. [2015] are mainly attributed to the completeness of biomes represented and the valuation. For ESV loss estimates of tropical forest the difference between Co and Dg is marginal, while the completeness of the Co dataset outranks the Dg coefficients.

Between 2000 and 2010 on a global scale tree cover loss of approximately 772 thousand  $km^2$  accounts for a ESV loss of 414.1, 405.2, and 101.1 billion dollar per year as table 3.6 shows. By applying the ESV coefficient for tropical forest from Costanza et al. [2014] the expansion of cropland and pastures account for a loss of approximately 95.3 and 128 billion dollar per year, respectively. The transition of forest cover to artificial surfaces and regrowth dynamics account for 2.3 and 124 billion dollar per year. If we consider the monetary value of tropical forest from Siikamaki et al. [2015] cropland, grassland, artificial surfaces, and regrowth dynamics account for a loss of approximately 23.2, 46.7, 0.5, and 30.2 billion dollar per year, respectively. The gross gain in ESV from LC transitions of tropical forest cover accounts for 350.6, 208.7, and 31.0 billion dollar per for the three datasets. By using the coefficients from the first dataset cropland, pastures, artificial surfaces, and regrowth dynamics account for a gain of 98.5, 120.6, 2.4, and 124.1 billion dollar per year, respectively. If we consider

de Groot et al. [2012] grassland and regrowth account for a gain of 83 and 121 billion dollar per year, while regrowth dynamics account for a gain of approximately 31 billion dollar for the last dataset. On global scale the net balance of tropical forest change account for -63.5, -196, and -70.1 billion dollar per year for the three datasets Co, Dg, and Wb, respectively. The greatest net loss can be observed by applying de Groot et al. ESV coefficients which could be mainly attributed to the small number of PDDs classes which correspond to a evaluated ESV biome. The relative small net loss in ESV by using the coefficients from the first dataset is attributed on the fact that approximately a fifth of the global tree cover is lost by the expansion of cropland. Costanza et al. estimates  $5567 \text{ Int'IS} \text{ y}^{-1}$  as the ESV for cropland which is 1.03 times greater than the ESV of tropical forest. Further, nearly a quarter of tropical forest cover is exposed to regrowth dynamics which uses the equivalent ESV as tropical tree cover. Therefore, approximately a half of the tropical forest loss is replaced by LC which has a greater or equal ESV than the ESV of tropical forest. For the time frame 2000 till 2012 Song [2018] estimates that the net loss of tropical forest cover accounts for a ESV loss of 550.7 billion dollar per year. The difference between our ESV loss can be attributed to discrepancies in LC change estimates, forest definition, and temporal resolution. Song considered the tree cover loss in all canopy densities and used for his net loss estimate tree cover gains within former forested area and the establishment of new tree cover. Further, Costanza et al. [2014] estimates that tropical forest loss account for a ESV loss of 3.5 trillion dollar per year between 1997 and 2011. The large differences can be attributed to forest loss area estimates and the time scale of the study.

# 4. Discussion

## 4.1. Deforestation

### 4.1.1. Forest definition

- For a regional approach a better solution could be to select for each region independently the right canopy density. For America a good agreement between the tree cover could be achieved by selecting the second class. For Asia by selecting the first class and for Africa the second. Even better would be to decide per tile individually which canopy density should be selected. This would eliminate regional effects of different forest densities.
- discuss regions independently Asia and America have large tree cover agreement
- Africa has the lowest agreement only core forest zones show high similarity
- To improve we should apply for each tile a canopy class decision based on our analysis
- This could improve the similarity (accuracy) by maximizing the sample count
- Extent method by analyzing the differences per clustering
- Could be used for Sannier et al. method as precursor to identify canopy densities which are worth for a detailed investigation
- Algorithm draft for single similarity: Compute Jaccard indexes for tile pair at different canopy densities, put results in a list, sort the list in decreasing order, pick the class where Jaccard index is max
- Use this Jaccard method to exclude tiles where the tree cover similarity fall below a certain threshold

#### 4.1.2. Tree cover and deforestation

- Improved tree cover map divide tree cover within hexagon by landmass within hexagon
- Improved loss map divide loss by the tree cover within a hexagon

#### 4.1.3. Proximate deforestation driver

- Reclassification is not a good idea cause this approach leads not to consistent results
- limitation the conversion of tree cover to shrubland or grassland in Africa can be mapping error

#### 4.1.4. Accuracy assessment

- Is largely subjective because it is prepared from the study author
- Better if someone independent does it
- Even better if you have ground truth prepared by field studies
- Class variability errors source the reclassification
- Time frame of our classification we classified our ground truth data at google image data
- Land cover change dynamics stay hidden
- Sample size is not scientific chosen
- Confusion matrix neglect uncertainty, apply approach in olofsson et al should improve the reasoning
- mention that kappa coefficient should be neglected cause issues reference olofsson

## 4.2. Emissions

## 4.3. Ecosystem service values

- resilience of esv loss could be achieved over optimizing total value of the new land-use
- target optimization is use the clearcut by maximizing profit and minimizing the esv loss
- large differences between the datasets
- till now the most complete data for global estimates is Costanza et al.
- de Groot et al. global values are redundant because nearly same estimate for tropical forest and smaller number of biomes is represented by this dataset
- Siikamaki et al. provides forest values which consider only a small amount of ecosystem services but it provides for each country a individual value

## 4.4. Binning analysis and visualization

- Cut polygon by line the Scala (1992) approach explained with parametric separation function, and bezier
- A approach where ratio is also ratio of the hexagon area

## 5. Conclusion

# Bibliography

- Alves D. S. An analysis of the geographical patterns of deforestation in the Brazilian Amazon in the period 1991-1996. In *Deforestation and land use in the Amazon*, chapter An analysis of the geographical patterns of deforestation in the Brazilian Amazon in the period 1991-1996, pages 95–107. University Press of Florida, 2002.
- Arjasakusuma S., Kamal M., Hafizt M., and Forestriko H. F. Local-scale accuracy assessment of vegetation cover change maps derived from Global Forest Change data, ClasLite, and supervised classifications: case study at part of Riau province, Indonesia. *Applied Geomatics*, 10(3): 205–217, May 2018. doi: 10.1007/s12518-018-0226-2.
- Arsanjani J. J., See L., and Tayyebi A. Assessing the suitability of GlobeLand30 for mapping land cover in Germany. *International Journal of Digital Earth*, 9(9):873–891, March 2016a. doi: 10.1080/17538947.2016.1151956.
- Arsanjani J. J., Tayyebi A., and Vaz E. GlobeLand30 as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries. *Habitat International*, 55:1–7, July 2016b. doi: 10.1016/j.habitatint.2016.02.003.
- Austin K. G., Schwantes A., Gu Y., and Kasibhatla P. S. What causes deforestation in Indonesia? *Environmental Research Letters*, 14(2):024007, February 2019. doi: 10.1088/1748-9326/aaf6db.
- Baccini A., Goetz S. J., Walker W. S., Laporte N. T., Sun M., Sulla-Menashe D., Hackler J., Beck P. S. A., Dubayah R., Friedl M. A., Samanta S., and Houghton R. A. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, 2(3):182–185, January 2012. doi: 10.1038/nclimate1354.
- Baccini A., Walker W., Carvahlo L., Farina M., Sulla-Menashe D., and Houghton R. Tropical forests are a net carbon source based on new measurements of gain and loss. Online accessed through Global Forest Watch, 2015. URL <https://www.globalforestwatch.org>.
- Baccini A., Walker W., Carvalho L., Farina M., Sulla-Menashe D., and Houghton R. A. Tropical forests are a net carbon source based on aboveground measurements of gain and loss. *Science*, 358(6360):230–234, September 2017. doi: 10.1126/science.aam5962.
- Barima Y. S. S., Kouakou A. T. M., Bamba I., Sangne Y. C., Godron M., Andrieu J., and Bogaert J. Cocoa crops are destroying the forest reserves of the classified forest of Haut-Sassandra (Ivory Coast). *Global Ecology and Conservation*, 8:85–98, October 2016. doi: 10.1016/j.gecco.2016.08.009.
- Beach T. Soil catenas, tropical deforestation, and ancient and contemporary soil erosion in the Petén, Guatemala. *Physical Geography*, 19(5):378–405, September 1998. doi: 10.1080/02723646.1998.10642657.

- Boucher D., Elias P., Lininger K., May-Tobin C., Roquemore S., and Saxon E. *The root of the problem: What's driving tropical deforestation today?* Union of concerned scientists, 2011.
- Cabral A. I. R., Vasconcelos M. J., Oom D., and Sardinha R. Spatial dynamics and quantification of deforestation in the central-plateau woodlands of Angola (1990-2009). *Applied Geography*, 31 (3):1185–1193, July 2011. doi: 10.1016/j.apgeog.2010.09.003.
- Caldas M. M., Goodin D., Sherwood S., Krauer J. M. C., and Wisely S. M. Land-cover change in the Paraguayan Chaco: 2000-2011. *Journal of Land Use Science*, 10(1):1–18, July 2013. doi: 10.1080/1747423x.2013.807314.
- Cao X., Li A., Lei G., Lei G., Tan J., Zhang Z., Yan D., Xie H., Zhang S., and Yang Y. Land cover mapping and spatial pattern analysis with remote sensing in Nepal. *Journal of Geo-information Science*, 18:1384–1398, 2016.
- Carr D. B. Looking at large data sets using binned data plots. resreport, U.S. Department of Energy, 1990.
- Carr D. B., Littlefield R. J., Nicholson W. L., and Littlefield J. S. Scatterplot matrix techniques for large N. *Journal of the American Statistical Association*, 82(398):424–436, June 1987. doi: 10.1080/01621459.1987.10478445.
- Carr D. B., Olsen A. R., and White D. Hexagon mosaic maps for display of univariate and bivariate geographical data. *Cartography and Geographic Information Systems*, 19(4):228–236, January 1992. doi: 10.1559/152304092783721231.
- Carter S., Herold M., Avitabile V., Bruun S., de, Sy V. D., Kooistra L., and Rufino M. C. Agricultural-driven deforestation in the tropics from 1990-2015: emissions, trends and uncertainties. *Environmental Research Letters*, 13(1):014002, December 2018. doi: 10.1088/1748-9326/aa9ea4.
- Chatelain C., Gautier L., and Spichiger R. A recent history of forest fragmentation in southwestern Ivory Coast. *Biodiversity and Conservation*, 5:37–53, 1996.
- Chazdon R. L. Beyond deforestation: restoring forests and ecosystem services on degraded lands. *Science*, 320(5882):1458–1460, June 2008. doi: 10.1126/science.1155365.
- Chen J., Chen J., Liao A., Cao X., Chen L., Chen X., He C., Han G., Peng S., Lu M., Zhang W., Tong X., and Mills J. *30-meter Global Land Cover Dataset - Product Description*. National Geomatics Center of China, May 2014.
- Chen J., Chen J., Liao A., Cao X., Chen L., Chen X., He C., Han G., Peng S., Lu M., Zhang W., Tong X., and Mills J. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:7–27, 2015. doi: 10.1016/j.isprsjprs.2014.09.002. URL <http://www.globallandcover.com>.
- Chen J., Cao X., Peng S., and Ren H. Analysis and applications of GlobeLand30: a review. *ISPRS International Journal of Geo-Information*, 6(8):230, July 2017. doi: 10.3390/ijgi6080230.
- Congalton R. G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1):35–46, July 1991. doi: 10.1016/0034-4257(91)90048-b.
- Costanza R., d'Arge R., Groot R., de, Farber S., Grasso M., Hannon B., Limburg K., Naeem S., O'Neill R. V., Paruelo J., Raskin R. G., Sutton P., and Belt M., van den. The value of the world's ecosystem services and natural capital. *Nature*, 387(6630):253–260, May 1997. doi: 10.1038/387253a0.

Costanza R., Groot R., de, Sutton P., Ploeg S., van der, Anderson S. J., Kubiszewski I., Farber S., and Turner R. K. Changes in the global value of ecosystem services. *Global Environmental Change*, 26:152–158, May 2014. doi: 10.1016/j.gloenvcha.2014.04.002.

Curtis P. G., Slay C. M., Harris N. L., Tyukavina A., and Hansen M. C. Classifying drivers of global forest loss. *Science*, 361(6407):1108–1111, September 2018. doi: 10.1126/science.aau3445.

David Kaimowitz, Patricia Mendez A. P. and Vanclay J. Spatial regression analysis of deforestation in Santa Cruz, Bolivia. In *Deforestation and land use in the Amazon*, pages 41–65. University Press of Florida, 2002.

Groot R., de, Brander L., Ploeg S., van der, Costanza R., Bernard F., Braat L., Christie M., Crossman N., Ghermandi A., Hein L., Hussain S., Kumar P., McVittie A., Portela R., Rodriguez L. C., Brink P., ten, and Beukering P., van. Global estimates of the value of ecosystems and their services in monetary units. *Ecosystem Services*, 1(1):50–61, July 2012. doi: 10.1016/j.ecoser.2012.07.005.

DeFries R. S., Rudel T., Uriarte M., and Hansen M. Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience*, 3(3):178–181, February 2010. doi: 10.1038/ngeo756.

Don A., Schumacher J., and Freibauer A. Impact of tropical land-use change on soil organic carbon stocks - a meta-analysis. *Global Change Biology*, 17(4):1658–1670, November 2010. doi: 10.1111/j.1365-2486.2010.02336.x.

FAO. Global forest resources assessment 2010. resreport, Food and Agriculture Organization of the United Nations, 2010.

FAO. FRA 2015 terms and definitions. resreport, Food and Agriculture Organization of the United Nations, 2012.

FAO. Global forest resources assessment 2015. resreport, Food and Agriculture Organization of the United Nations, 2016.

FAO and ITPS. Global Soil Organic Carbon Map. resreport, FAO and ITPS, 2018.

Foody G. M. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1):185–201, April 2002. doi: 10.1016/s0034-4257(01)00295-4.

Furumo P. R. and Aide T. M. Characterizing commercial oil palm expansion in Latin America: land use change and trade. *Environmental Research Letters*, 12(2):024008, February 2017. doi: 10.1088/1748-9326/aa5892.

Geist H. J. and Lambin E. F. *What drives tropical deforestation*. LUCC, 2001.

Goetze D., Horsch B., and Porembski S. Dynamics of forest-savanna mosaics in north-eastern Ivory Coast from 1954 to 2002. *Journal of Biogeography*, 33(4):653–664, April 2006. doi: 10.1111/j.1365-2699.2005.01312.x.

Graesser J., Aide T. M., Grau H. R., and Ramankutty N. Cropland/pastureland dynamics and the slowdown of deforestation in Latin America. *Environmental Research Letters*, 10(3):034017, March 2015. doi: 10.1088/1748-9326/10/3/034017.

Graesser J., Ramankutty N., and Coomes O. T. Increasing expansion of large-scale crop production onto deforested land in sub-Andean South America. *Environmental Research Letters*, 13(8): 084021, August 2018. doi: 10.1088/1748-9326/aad5bf.

- Hansen M. C., Potapov P. V., Moore R., Hancher M., Turubanova S. A., Tyukavina A., Thau D., Stehman S. V., Goetz S. J., Loveland T. R., Kommareddy A., Egorov A., Chini L., Justice C. O., and Townshend J. R. G. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160):850–853, November 2013. doi: 10.1126/science.1244693. URL [https://earthenginepartners.appspot.com/science-2013-global-forest/download\\_v1.0.html](https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.0.html).
- Harper G. J., Steininger M. K., Tucker C. J., Juhn D., and Hawkins F. Fifty years of deforestation and forest fragmentation in Madagascar. *Environmental Conservation*, 34(04), December 2007. doi: 10.1017/s0376892907004262.
- Hijmans R., Garcia N., Kapoor J., Rala A., Maunahan A., and Wieczorek J. GADM database of Global Administrative Areas. Online, May 2018. URL <https://www.gadm.org>. Version 3.6.
- Hirsch P. Underlying causes of deforestation in the Mekong region. resreport, Australian Mekong Resource Centre, Sydney, 2000.
- Hoshen J. On the application of the enhanced Hoshen-Kopelman algorithm for image analysis. *Pattern Recognition Letters*, 19(7):575–584, May 1998. doi: 10.1016/s0167-8655(98)00018-x.
- Hosonuma N., Herold M., Sy V. D., Fries R. S. D., Brockhaus M., Verchot L., Angelsen A., and Romijn E. An assessment of deforestation and forest degradation drivers in developing countries. *Environmental Research Letters*, 7(4):1–12, October 2012. doi: 10.1088/1748-9326/7/4/044009.
- Ickowitz A., Slayback D., Asanzi P., and Nasi R. A. Agriculture and deforestation in the Democratic Republic of the Congo: a synthesis of the current state of knowledge. resreport, CIFOR, 2015.
- Jaccard P. The distribution of the flor in the alpine zone. *The New Phytologist*, 11(2), February 1912.
- Jacobson A., Dhanota J., Godfrey J., Jacobson H., Rossman Z., Stanish A., Walker H., and Riggio J. A novel approach to mapping land conversion using Google Earth with an application to East Africa. *Environmental Modelling & Software*, 72:1–9, October 2015. doi: 10.1016/j.envsoft.2015.06.011.
- Kideghesho J. R. Realities on deforestation in Tanzania-trends, drivers, implications and the way forward. In *Precious Forests - Precious Earth*. InTech, September 2015. doi: 10.5772/61002.
- Kreuter U. P., Harris H. G., Matlock M. D., and Lacey R. E. Change in ecosystem service values in the San Antonio area, Texas. *Ecological Economics*, 39(3):333–346, December 2001. doi: 10.1016/s0921-8009(01)00250-6.
- Kummer D. M. and Turner B. L. The human causes of deforestation in Southeast Asia. *BioScience*, 44(5):323–328, May 1994.
- Li Y., Sulla-Menashe D., Motesharrei S., Song X.-P., Kalnay E., Ying Q., Li S., and Ma Z. Inconsistent estimates of forest cover change in China between 2000 and 2013 from multiple datasets: differences in parameters, spatial resolution, and definitions. *Scientific Reports*, 7(1), August 2017. doi: 10.1038/s41598-017-07732-5.
- Lowry R. *Concepts and applications of inferential statistics*. Vassar College, 2019. URL <http://www.vassarstats.net/textbook/>.
- McKinney W. Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 2010.

McRoberts R. E., Vibrans A. C., Sannier C., Næsset E., Hansen M. C., Walters B. F., and Lingner D. V. Methods for evaluating the utilities of local and global maps for increasing the precision of estimates of subtropical forest area. *Canadian Journal of Forest Research*, 46(7):924–932, July 2016. doi: 10.1139/cjfr-2016-0064.

McSweeney K., Nielsen E. A., Taylor M. J., Wrathall D. J., Pearson Z., Wang O., and Plumb S. T. Drug policy as conservation policy: narco-deforestation. *Science*, 343(6170):489–490, January 2014. doi: 10.1126/science.1244082.

Meyfroidt P., Vu T. P., and Hoang V. A. Trajectories of deforestation, coffee expansion and displacement of shifting cultivation in the Central Highlands of Vietnam. *Global Environmental Change*, 23(5):1187–1198, October 2013. doi: 10.1016/j.gloenvcha.2013.04.005.

Müller H., Griffiths P., and Hostert P. Long-term deforestation dynamics in the Brazilian Amazon—Uncovering historic frontier development along the Cuiabá–Santarém highway. *International Journal of Applied Earth Observation and Geoinformation*, 44:61–69, February 2016. doi: 10.1016/j.jag.2015.07.005.

Nawir A. A. and Rumboko L. History and state of deforestation and land degradation. In *Forest rehabilitation in Indonesia*, pages 11–33. Center for International Forestry Research, 2007.

Pacheco P. Deforestation and forest degradation in lowland Bolivia. In *Deforestation and land use in the Amazon*, pages 66–95. University Press of Florida, 2002.

Potapov P., Hansen M. C., Laestadius L., Turubanova S., Yaroshenko A., Thies C., Smith W., Zhuravleva I., Komarova A., Minnemeyer S., and Esipova E. The last frontiers of wilderness: Tracking loss of intact forest landscapes from 2000 to 2013. *Science Advances*, 3(1), January 2017. doi: 10.1126/sciadv.1600821. URL <http://www.intactforests.org/>.

Renó V. F., Novo E. M. L. M., Suemitsu C., Rennó C. D., and Silva T. S. F. Assessment of deforestation in the Lower Amazon floodplain using historical Landsat MSS/TM imagery. *Remote Sensing of Environment*, 115(12):3446–3456, December 2011. doi: 10.1016/j.rse.2011.08.008.

Sampat M. P., Wang Z., Gupta S., Bovik A. C., and Markey M. K. Complex wavelet structural similarity: a new image similarity index. *IEEE Transactions on Image Processing*, 18(11): 2385–2401, November 2009. doi: 10.1109/tip.2009.2025923.

Sannier C., McRoberts R. E., and Fichet L.-V. Suitability of Global Forest Change data to report forest cover estimates at national level in Gabon. *Remote Sensing of Environment*, 173:326–338, February 2016. doi: 10.1016/j.rse.2015.10.032.

Sannigrahi S., Bhatt S., Rahmat S., Paul S. K., and Sen S. Estimating global ecosystem service values and its response to land surface dynamics during 1995–2015. *Journal of Environmental Management*, 223:115–131, October 2018. doi: 10.1016/j.jenvman.2018.05.091.

Schneibel A., Frantz D., Röder A., Stellmes M., Fischer K., and Hill J. Using annual landsat time series for the detection of dry forest degradation processes in south-central Angola. *Remote Sensing*, 9(9):905, August 2017. doi: 10.3390/rs9090905.

Seydewitz T. Applicability of GlobeLand30 and Global Forest Change data products for forest land cover change studies on global and regional scales. resreport, Potsdam Institute for Climate Impact Research, 2017. Internship - report.

Shi G. R. Multivariate data analysis in palaeoecology and palaeobiogeography - a review. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 105(3–4):199–234, November 1993. doi: 10.1016/0031-0182(93)90084-v.

- Siiakamaki J., Santiago-Avila F. J., and Vail P. Global assessment of non-wood forest ecosystem services. resreport, Program on Forests (PROFOR), 2015.
- Sitoe A., Salomao A., and Wertz-Kanounnikoff S. The context of REDD+ in Mozambique: drivers, agents and institutions. resreport, CIFOR, 2012.
- Song X.-P. Global estimates of ecosystem service value and change: taking into account uncertainties in satellite-based land cover data. *Ecological Economics*, 143:227–235, January 2018. doi: 10.1016/j.ecolecon.2017.07.019.
- Sy V. D., Herold M., Achard F., Beuchle R., Clevers J. G. P. W., Lindquist E., and Verchot L. Land use patterns and related carbon losses following deforestation in South America. *Environmental Research Letters*, 10(12):124004, November 2015. doi: 10.1088/1748-9326/10/12/124004.
- Rossum G., van and Development T. The Python language reference: release 3.5.6. Online, 2018. URL <https://docs.python.org/3.5/download.html>. Python Software Foundation.
- Vijay V., Reid C. D., Finer M., Jenkins C. N., and Pimm S. L. Deforestation risks posed by oil palm expansion in the Peruvian Amazon. *Environmental Research Letters*, 13(11):114010, November 2018. doi: 10.1088/1748-9326/aae540.
- Wang Z., Zhang B., Zhang S., Li X., Liu D., Song K., Li J., Li F., and Duan H. Changes of land use and of ecosystem service values in Sanjiang Plain, Northeast China. *Environmental Monitoring and Assessment*, 112(1-3):69–91, jan 2006. doi: 10.1007/s10661-006-0312-5.
- Wilcoxon F. Individual comparisions by ranking methods. *Biometrics Bulletin*, 1(6):80–83, December 1945.
- Wood C. H. Introduction - deforestation and land use in the Amazon. In *Deforestation and land use in the Amazon*, pages 1–38. University Press of Florida, 2002.
- WWF. *Living Planet Report 2016: Risk and resilience in a new era*. WWW International, 2016. ISBN 978-2-940529-40-7.
- Yang Y., Xiao P., Feng X., and Li H. Accuracy assessment of seven global land cover datasets over China. *ISPRS Journal of Photogrammetry and Remote Sensing*, 125:156–173, March 2017. doi: 10.1016/j.isprsjprs.2017.01.016.
- Zak M. R., Cabido M., and Hodgson J. G. Do subtropical seasonal forests in the Gran Chaco, Argentina, have a future? *Biological Conservation*, 120(4):589–598, December 2004. doi: 10.1016/j.biocon.2004.03.034.
- Zalles V., Hansen M. C., Potapov P. V., Stehman S. V., Tyukavina A., Pickens A., Song X.-P., Adusei B., Okpa C., Aguilar R., John N., and Chavez S. Near doubling of Brazils intensive row crop area since 2000. *Proceedings of the National Academy of Sciences*, 116(2):428–435, December 2018. doi: 10.1073/pnas.1810301115.
- Zhao B., Kreuter U., Li B., Ma Z., Chen J., and Nakagoshi N. An ecosystem service value assessment of land-use change on Chongming Island, China. *Land Use Policy*, 21(2):139–148, April 2004. doi: 10.1016/j.landusepol.2003.10.003.

# List of Figures

1.1.	Tropical zone and forest . . . . .	11
1.2.	Examples of proximate deforestation drivers . . . . .	13
2.1.	Map of dataset tiles . . . . .	17
2.2.	Raster and vector harmonization process . . . . .	24
2.3.	Harmonized raster images and sampling locations . . . . .	25
2.4.	Classification of proximate deforestation drivers . . . . .	29
2.5.	Hexagon tessellation . . . . .	39
3.1.	Tree cover similarity distribution . . . . .	42
3.2.	Tree cover and canopy density in Latin America at 2000 . . . . .	48
3.3.	Tree cover loss in Latin America between 2000 and 2010 . . . . .	49
3.4.	Tree cover and canopy density in Asia/Australia at 2000 . . . . .	50
3.5.	Tree cover loss in Asia/Australia between 2000 and 2010 . . . . .	51
3.6.	Tree cover and canopy density in Africa at 2000 . . . . .	52
3.7.	Tree cover loss in Africa between 2000 and 2010 . . . . .	53
3.8.	Map of proximate deforestation drivers in Latin America . . . . .	56
3.9.	Map of proximate deforestation drivers in Asia/Australia . . . . .	58
3.10.	Map of proximate deforestation drivers in Africa . . . . .	59
B.1.	Caption . . . . .	VII
B.2.	Caption . . . . .	VIII
B.3.	Caption . . . . .	IX

# List of Tables

2.1. Datasets used in this study . . . . .	15
2.2. Land cover classification of the GlobeLand30 product . . . . .	19
2.3. Relative soil organic carbon change for certain land-use change types . . . . .	22
2.4. Ecosystem service values (ESV) used in this study . . . . .	23
2.5. Jaccard Index coefficient matrix . . . . .	27
2.6. A general model of a confusion matrix . . . . .	31
2.7. Scenario one mapping of soil organic carbon change to porixmate driver . . . . .	34
2.8. Scenario two mapping of soil organic carbon change to porixmate driver . . . . .	34
2.9. Biome ESV types corresponding to PDD classes . . . . .	36
3.1. Continental experiment group comparison . . . . .	43
3.2. Continental experiment group directional comparison . . . . .	44
3.3. Global experiment group comparison . . . . .	46
3.4. Global experiment group directional comparison . . . . .	46
3.5. Confussion matrix . . . . .	61
3.6. Ecosystem service value balance . . . . .	63
B.1. Comparison of tree cover agreement between regions . . . . .	VI
B.2. Comparison of tree cover agreement between regions . . . . .	VI
C.1. Estimates of proximate drivers of deforestation for countries, continents, and tropical zone . . . . .	XI

# List of Abbreviations

<b>AGB</b>	Aboveground live woody Biomass density
<b>AISM</b>	Aligned Image Stack Mosaic
<b>API</b>	Application Programming Interface
<b>CRS</b>	Coordinate Reference System
<b>CSV</b>	Comma Separated Values
<b>ESV</b>	Ecosystem Service Value
<b>ETM+</b>	Enhanced Thematic Mapper Plus
<b>FAO</b>	Food and Agriculture Organization of the United Nations
<b>GADM</b>	Global Administrative Areas Map
<b>GFC</b>	Global Forest Change
<b>GFW</b>	Global Forest Watch
<b>GIS</b>	Geographic Information System
<b>GL30</b>	GlobeLand30
<b>GLAS</b>	Geoscience Laser Altimeter System
<b>GSOCmap</b>	Global Soil Organic Carbon map
<b>GSP</b>	Global Soil Partnership
<b>GTiff</b>	Geo-Tiff
<b>GeoJSON</b>	Geographic JavaScript Object Notation
<b>IFL</b>	Intact Forest Landscapes
<b>IPCC</b>	Intergovernmental Panel on Climate Change
<b>ISRIC</b>	International Soil Reference and Information Center
<b>ITPS</b>	Intergovernmental Technical Panel on Soils
<b>LC</b>	Land Cover
<b>LCC</b>	Land Cover Change
<b>LIDAR</b>	Light Detection and Ranging
<b>LU</b>	Land Use
<b>LUC</b>	Land-use Change
<b>LULC</b>	Land Use/Land Cover
<b>ME</b>	Mean Error
<b>MLC</b>	Maximum Likelihood Classifier
<b>MLR</b>	Multiple Linear Regression
<b>REGEX</b>	Regular Expression
<b>RMSE</b>	Root Mean Square Error
<b>SD</b>	Standard Deviation
<b>SHP</b>	Shapefile
<b>SOC</b>	Soil Organic Carbon
<b>SOCC</b>	Soil Organic Carbon Content

<b>UN</b>	United Nations
<b>URL</b>	Uniform Resource Locator
<b>UTM</b>	Universal Transverse Mercator
<b>WGS84</b>	World Geodetic System 1984
<b>stdlib</b>	Standard Library
<b>GDAL</b>	Geospatial Data Abstraction Library
<b>JI</b>	Jaccard Index
<b>PDD</b>	Proximate Deforestation Driver
<b>GHG</b>	Greenhouse gas

## A. Hexagonal country boundaries

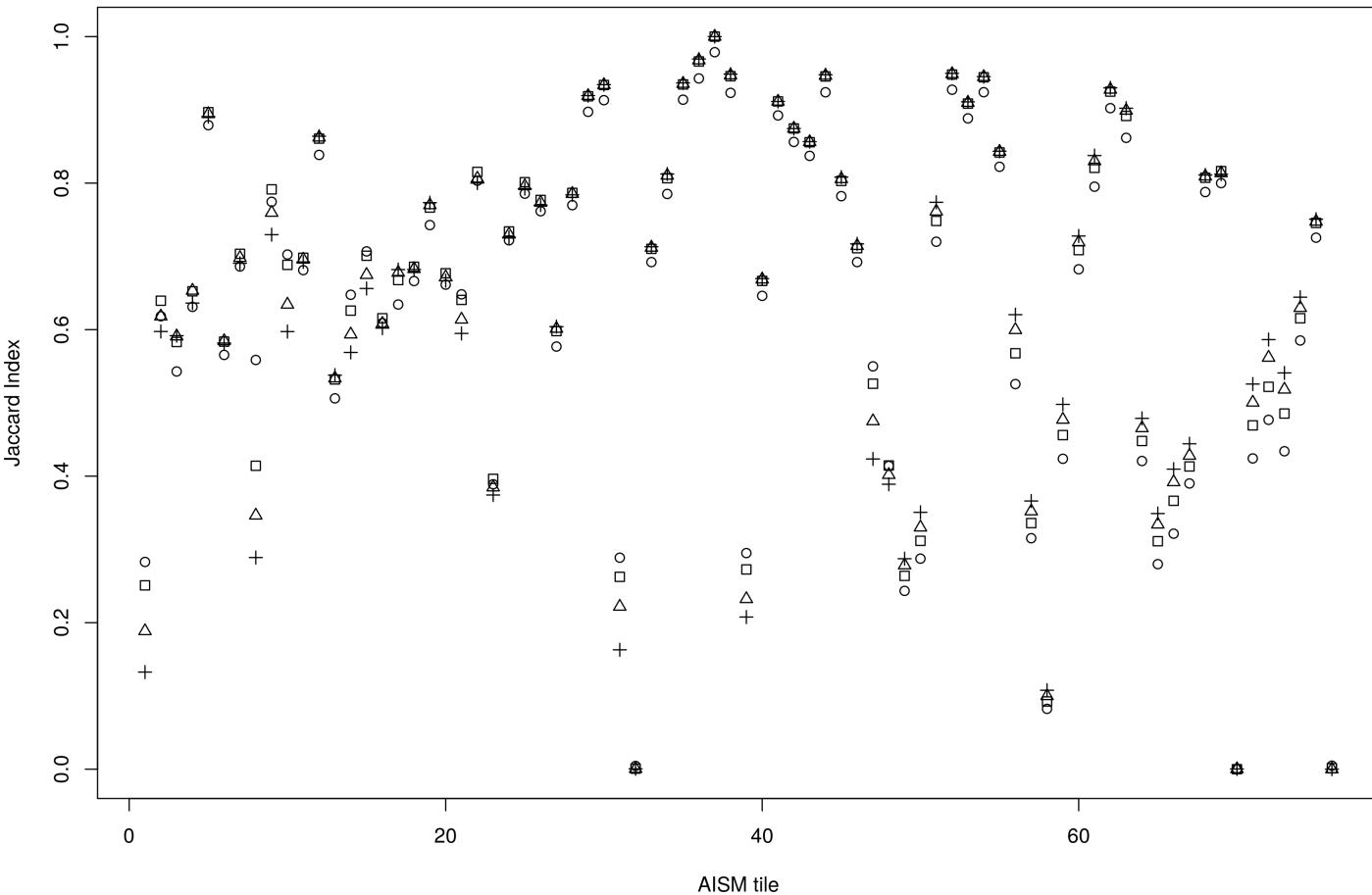
## B. Forest definition

**Table B.1. Comparison of tree cover agreement between regions:** This table shows, a comparison of tree cover agreement between regions. The classes  $Jl_0$ ,  $Jl_1$ ,  $Jl_2$ , and  $Jl_3$  as row and column headings account for the canopy density classes (0,100], (10,100], (20,100], and (30,100], respectively. The test hypothesis is  $H_0: X_1 = X_2$  where  $X_1$  is the column  $Jl_n$  class and  $X_2$  the row  $Jl_n$  class. The significance is indicated by  $p^* < 0.05$ ,  $p^{**} < 0.02$ , and  $p^{***} < 0.01$ . We applied a Benjamini and Hochberg correction for multiple-pairwise testing.

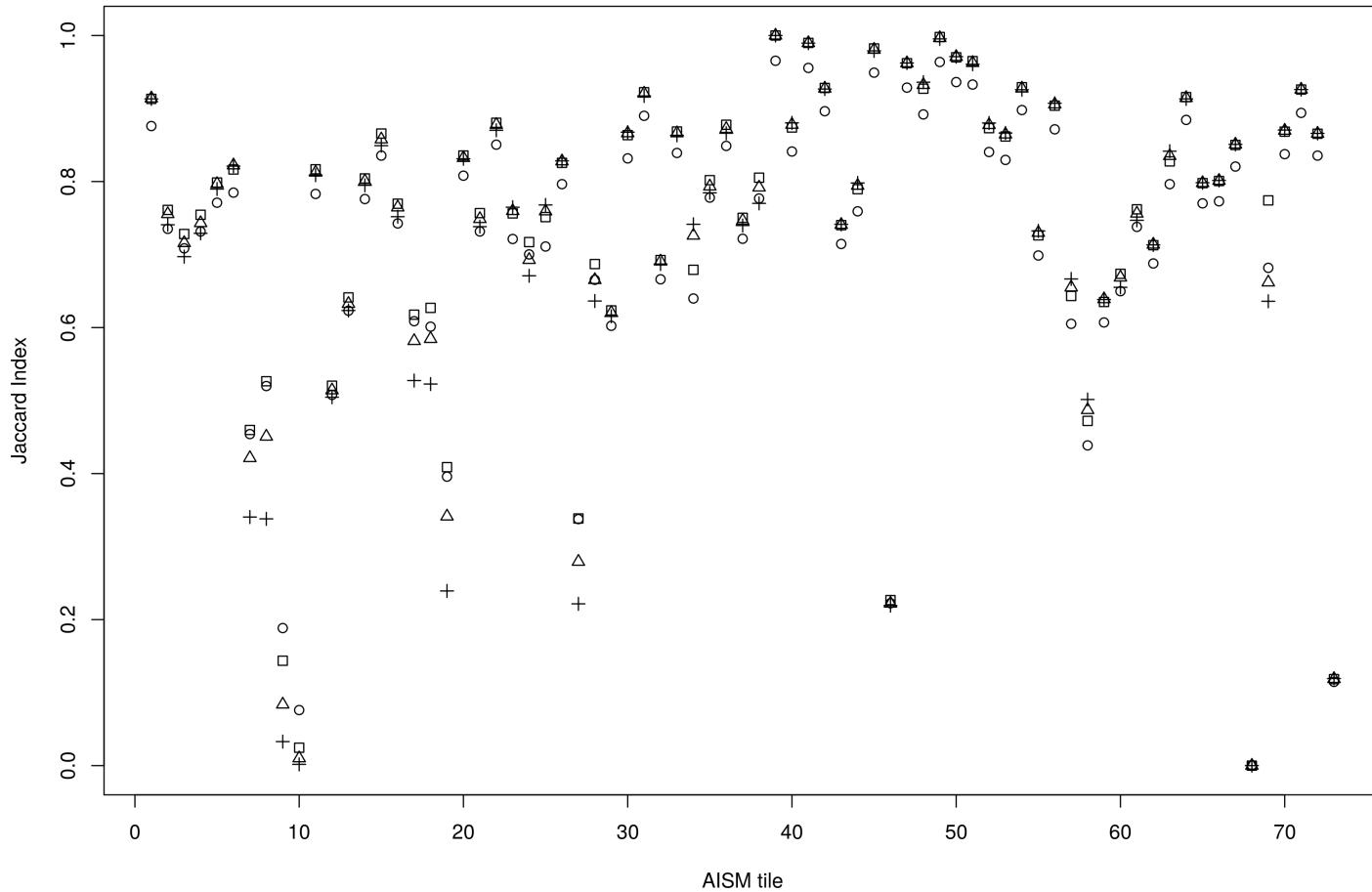
		Latin America				Asia/Australia			
	Cls	$Jl_0$	$Jl_1$	$Jl_2$	$Jl_3$	$Jl_0$	$Jl_1$	$Jl_2$	$Jl_3$
Asia	$Jl_0$	.04*	-	-	-	-	-	-	-
	$Jl_1$	-	.04*	-	-	-	-	-	-
	$Jl_2$	-	-	.05*	-	-	-	-	-
	$Jl_3$	-	-	-	.07	-	-	-	-
Africa	$Jl_0$	.00***	-	-	-	.00***	-	-	-
	$Jl_1$	-	.00***	-	-	-	.00***	-	-
	$Jl_2$	-	-	.00***	-	-	-	.00***	-
	$Jl_3$	-	-	-	.00***	-	-	-	.00***

**Table B.2. Comparison of tree cover agreement between regions:** This table shows, a comparison of tree cover agreement between regions and the direction of differences. The classes  $Jl_0$ ,  $Jl_1$ ,  $Jl_2$ , and  $Jl_3$  as row and column headings account for the canopy density classes (0,100], (10,100], (20,100], and (30,100], respectively. The test hypothesis is  $H_0: X_1 \leq X_2$  where  $X_1$  is the column  $Jl_n$  class and  $X_2$  the row  $Jl_n$  class. The significance is indicated by  $p^* < 0.05$ ,  $p^{**} < 0.025$ ,  $p^{***} < 0.01$ , and  $p^\dagger < 0.005$ . We applied a Benjamini and Hochberg correction for multiple-pairwise testing.

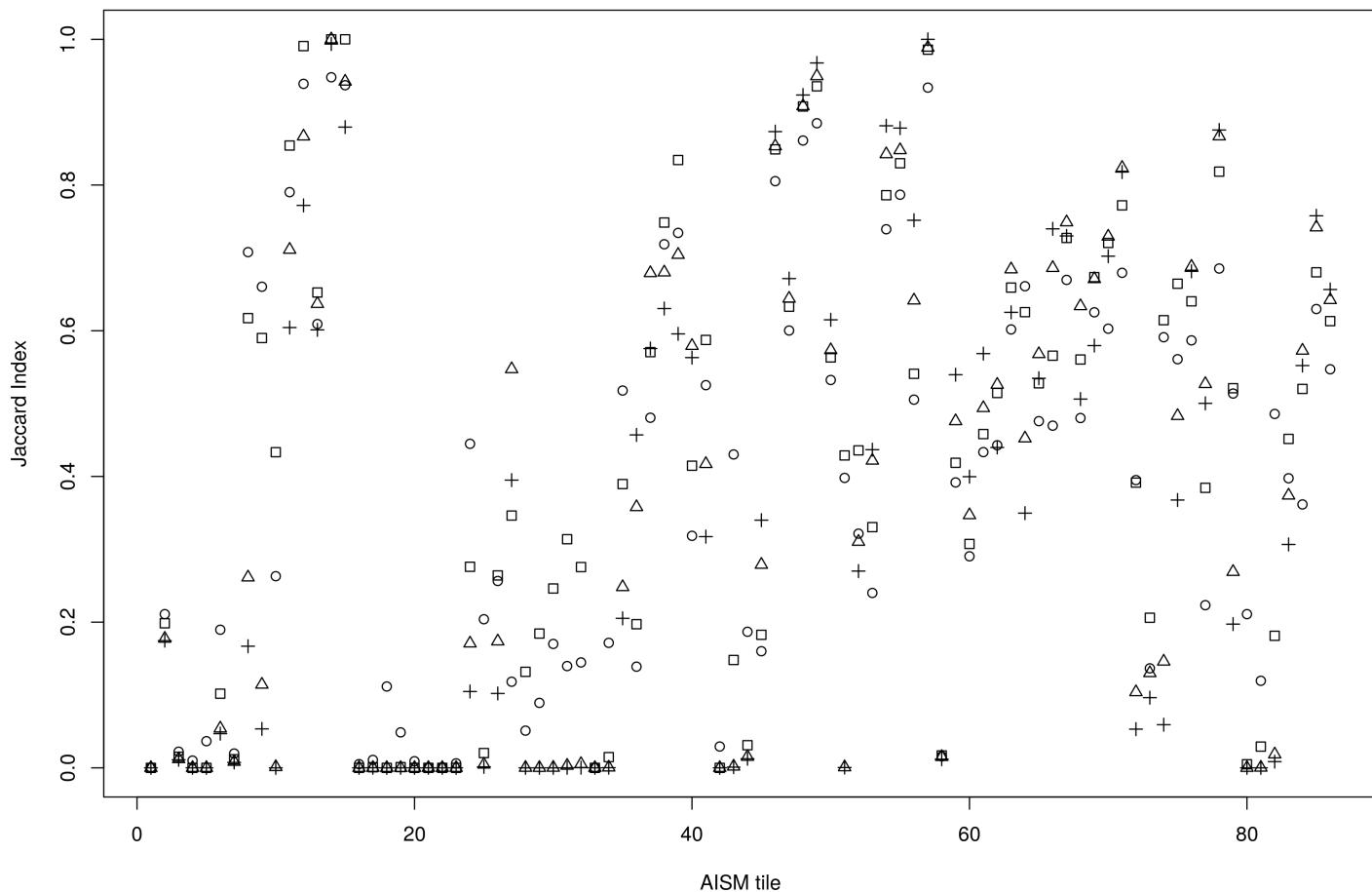
		Latin America				Asia/Australia			
	Cls	$Jl_0$	$Jl_1$	$Jl_2$	$Jl_3$	$Jl_0$	$Jl_1$	$Jl_2$	$Jl_3$
Asia	$Jl_0$	1.	-	-	-	-	-	-	-
	$Jl_1$	-	1.	-	-	-	-	-	-
	$Jl_2$	-	-	1.	-	-	-	-	-
	$Jl_3$	-	-	-	1.	-	-	-	-
Africa	$Jl_0$	.00†	-	-	-	.00†	-	-	-
	$Jl_1$	-	.00†	-	-	-	.00†	-	-
	$Jl_2$	-	-	.00†	-	-	-	.00†	-
	$Jl_3$	-	-	-	.00†	-	-	-	.00†



**Figure B.1. Caption:** Caption



**Figure B.2. Caption:** Caption



**Figure B.3. Caption:** Caption

## C. Proximate deforestation drivers

**Table C.1. Estimates of proximate drivers of deforestation for countries, continents, and tropical zone:**

Region	Cultivated	Forest	Regrowth	Grassland	Shrubland	Water	Artificial	Bareland	Total	Total <sup>†</sup>
Argentina	5733.9 (65.8%)	1380.6 (15.9%)	147.9 (1.7%)	461.8 (5.3%)	718.7 (8.3%)	230.8 (2.7%)	32.8 (0.4%)	2.0 (0.0%)	8708.5	7097.1
Bahamas	0.0 (0.0%)	3.3 (41.2%)	0.1 (1.3%)	0.0 (0.0%)	0.3 (3.8%)	3.0 (37.5%)	1.3 (16.3%)	0.0 (0.0%)	8.0	1.7
Belize	212.2 (24.4%)	113.0 (13.0%)	155.5 (17.9%)	308.9 (35.6%)	54.5 (6.3%)	12.7 (1.5%)	11.1 (1.3%)	0.0 (0.0%)	867.9	742.2
Bolivia	9288.2 (37.2%)	5106.4 (20.5%)	2428.1 (9.7%)	5581.0 (22.4%)	1825.1 (7.3%)	590.8 (2.4%)	108.3 (0.4%)	19.5 (0.1%)	24947.4	19250.2
Brazil*	56443.7 (19.1%)	16433.7 (5.6%)	34746.7 (11.8%)	146802.9 (49.7%)	35954.2 (12.2%)	3832.4 (1.3%)	738.8 (0.3%)	209.0 (0.1%)	295161.4	274895.3
Chile*	-	-	-	-	-	-	-	-	-	-
Colombia	1297.9 (6.2%)	5272.3 (25.2%)	5177.2 (24.7%)	6995.9 (33.4%)	1593.1 (7.6%)	573.6 (2.7%)	26.0 (0.1%)	9.7 (0.0%)	20945.7	15099.8
Costa Rica	307.4 (21.2%)	536.8 (37.1%)	178.5 (12.3%)	397.4 (27.5%)	11.0 (0.8%)	7.0 (0.5%)	9.0 (0.6%)	0.1 (0.0%)	1447.2	903.4
Cuba	233.5 (17.6%)	499.9 (37.7%)	181.2 (13.7%)	330.4 (24.9%)	23.2 (1.8%)	35.2 (2.7%)	6.2 (0.5%)	15.7 (1.2%)	1325.3	790.2
Dominican Republic	81.6 (5.4%)	763.5 (50.5%)	122.2 (8.1%)	510.8 (33.8%)	2.9 (0.2%)	4.2 (0.3%)	13.3 (0.9%)	13.1 (0.9%)	1511.6	743.9
Ecuador	1143.9 (28.2%)	1163.7 (28.7%)	823.3 (20.3%)	735.7 (18.1%)	142.2 (3.5%)	34.7 (0.9%)	8.0 (0.2%)	6.3 (0.2%)	4057.8	2859.4
El Salvador	58.1 (11.6%)	28.2 (5.7%)	6.2 (1.2%)	390.9 (78.3%)	8.5 (1.7%)	1.6 (0.3%)	5.6 (1.1%)	0.0 (0.0%)	499.1	469.3
French Guiana	46.9 (14.7%)	52.8 (16.5%)	100.0 (31.2%)	91.4 (28.6%)	0.0 (0.0%)	7.8 (2.4%)	20.6 (6.4%)	0.5 (0.2%)	320.0	259.4
Guatemala	1843.2 (23.9%)	1172.6 (15.2%)	924.2 (12.0%)	2974.9 (38.6%)	733.1 (9.5%)	21.2 (0.3%)	39.6 (0.5%)	0.0 (0.0%)	7708.8	6515.0
Guyana	22.1 (3.1%)	244.7 (34.3%)	274.7 (38.5%)	134.0 (18.8%)	10.1 (1.4%)	19.9 (2.8%)	8.3 (1.2%)	0.0 (0.0%)	713.8	449.2
Haiti	8.3 (3.9%)	146.5 (69.3%)	13.9 (6.6%)	40.9 (19.3%)	0.1 (0.0%)	0.9 (0.4%)	0.2 (0.1%)	0.6 (0.3%)	211.4	64.0
Honduras	101.2 (2.5%)	1200.1 (30.0%)	487.9 (12.2%)	2050.5 (51.2%)	137.8 (3.4%)	20.4 (0.5%)	8.9 (0.2%)	0.0 (0.0%)	4006.8	2786.3
Jamaica	7.2 (2.8%)	127.7 (48.8%)	25.5 (9.7%)	82.4 (31.5%)	0.0 (0.0%)	1.9 (0.7%)	16.7 (6.4%)	0.3 (0.1%)	261.7	132.1
Mexico*	3267.8 (18.7%)	4857.8 (27.8%)	3763.8 (21.5%)	3950.6 (22.6%)	1169.5 (6.7%)	88.8 (0.5%)	373.4 (2.1%)	11.1 (0.1%)	17482.8	12536.2
Nicaragua	68.9 (0.9%)	3405.0 (45.9%)	1567.8 (21.2%)	2268.8 (30.6%)	78.3 (1.1%)	19.9 (0.3%)	2.5 (0.0%)	0.0 (0.0%)	7411.2	3986.3
Panama	142.2 (6.3%)	619.7 (27.6%)	352.8 (15.7%)	1019.7 (45.5%)	70.8 (3.2%)	22.3 (1.0%)	13.5 (0.6%)	0.4 (0.0%)	2241.4	1599.4
Paraguay*	13718.0 (48.9%)	6383.0 (22.8%)	216.2 (0.8%)	2480.2 (8.8%)	5107.2 (18.2%)	118.8 (0.4%)	24.9 (0.1%)	1.2 (0.0%)	28049.5	21547.7
Peru	780.4 (6.6%)	2333.9 (19.8%)	3394.0 (28.8%)	2710.8 (23.0%)	1230.9 (10.4%)	1213.5 (10.3%)	27.8 (0.2%)	102.2 (0.9%)	11793.5	8246.1
Puerto Rico	4.1 (3.7%)	56.4 (50.5%)	10.2 (9.1%)	30.7 (27.5%)	0.0 (0.0%)	0.4 (0.4%)	8.8 (7.9%)	1.0 (0.9%)	111.6	54.8
Suriname	36.7 (7.4%)	132.2 (26.8%)	167.5 (33.9%)	108.5 (22.0%)	0.0 (0.0%)	40.8 (8.3%)	8.2 (1.7%)	0.0 (0.0%)	493.9	320.9
Venezuela	1082.3 (10.2%)	2758.9 (26.1%)	1644.0 (15.5%)	3382.4 (32.0%)	1388.8 (13.1%)	267.4 (2.5%)	47.7 (0.5%)	12.7 (0.1%)	10584.2	7557.9
Latin America	95929.7 (21.3%)	54792.7 (12.2%)	56909.4 (12.6%)	183841.5 (40.8%)	50260.3 (11.1%)	7170.0 (1.6%)	1561.5 (0.3%)	405.4 (0.1%)	450870.5	388907.8
Algeria*	-	-	-	-	-	-	-	-	-	-
Angola	4818.6 (32.7%)	697.8 (4.7%)	452.5 (3.1%)	8065.9 (54.7%)	371.2 (2.5%)	90.9 (0.6%)	226.5 (1.5%)	10.9 (0.1%)	14734.3	13945.6
Benin	1518.7 (58.2%)	200.6 (7.7%)	45.1 (1.7%)	74.6 (2.9%)	762.5 (29.2%)	0.5 (0.0%)	5.6 (0.2%)	0.0 (0.0%)	2607.6	2406.5
Botswana*	9.4 (17.9%)	2.1 (4.0%)	0.0 (0.0%)	26.4 (50.2%)	3.6 (6.8%)	9.9 (18.8%)	1.2 (2.3%)	0.0 (0.0%)	52.6	40.6
Burkina Faso	510.9 (29.4%)	4.6 (0.3%)	0.0 (0.0%)	1190.0 (68.5%)	31.7 (1.8%)	0.3 (0.0%)	0.1 (0.0%)	0.7 (0.0%)	1738.3	1733.4
Burundi	49.7 (30.6%)	6.3 (3.9%)	3.6 (2.2%)	98.8 (60.9%)	2.0 (1.2%)	0.5 (0.3%)	1.3 (0.8%)	0.0 (0.0%)	162.2	155.4
Cameroon	337.8 (8.5%)	938.2 (23.6%)	717.2 (18.1%)	1838.8 (46.3%)	14.4 (0.4%)	34.6 (0.9%)	68.6 (1.7%)	21.6 (0.5%)	3971.2	2998.4

Region	Cultivated	Forest	Regrowth	Grassland	Shrubland	Water	Artificial	Bareland	Total	Total <sup>†</sup>
C. African Rep.	707.3 (18.3%)	217.1 (5.6%)	328.4 (8.5%)	2470.2 (64.1%)	118.6 (3.1%)	3.3 (0.1%)	9.8 (0.3%)	0.6 (0.0%)	3855.3	3634.9
Chad	1454.7 (54.7%)	1.8 (0.1%)	0.1 (0.0%)	1037.5 (39.0%)	142.2 (5.4%)	15.9 (0.6%)	3.1 (0.1%)	2.4 (0.1%)	2657.7	2640.0
Congo	107.8 (4.9%)	406.0 (18.4%)	591.8 (26.8%)	968.0 (43.8%)	0.3 (0.0%)	109.7 (5.0%)	25.8 (1.2%)	0.1 (0.0%)	2209.5	1693.8
DR Congo	4771.8 (9.4%)	3518.2 (6.9%)	13508.7 (26.6%)	26400.4 (52.0%)	508.2 (1.0%)	1736.9 (3.4%)	314.7 (0.6%)	0.1 (0.0%)	50759.0	45503.9
Djibouti	-	-	-	-	-	-	-	-	-	-
Egypt*	-	-	-	-	-	-	-	-	-	-
Equatorial Guinea	0.0 (0.0%)	77.2 (28.7%)	77.7 (28.9%)	72.6 (27.0%)	1.3 (0.5%)	1.3 (0.5%)	38.5 (14.3%)	0.0 (0.0%)	268.6	190.1
Eritrea	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.1 (100.0%)	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.1	0.1
Ethiopia	356.8 (16.1%)	296.9 (13.4%)	137.2 (6.2%)	1321.6 (59.5%)	92.3 (4.2%)	4.5 (0.2%)	3.9 (0.2%)	6.6 (0.3%)	2219.8	1918.4
Gabon	1.4 (0.1%)	528.7 (32.9%)	664.1 (41.3%)	362.7 (22.5%)	0.0 (0.0%)	22.9 (1.4%)	24.9 (1.5%)	3.9 (0.2%)	1608.6	1057.0
Gambia	35.6 (29.1%)	0.0 (0.0%)	0.0 (0.0%)	63.5 (51.9%)	19.8 (16.2%)	0.1 (0.1%)	3.4 (2.8%)	0.0 (0.0%)	122.4	122.3
Ghana	319.1 (6.8%)	525.3 (11.1%)	1523.7 (32.3%)	1990.8 (42.2%)	235.5 (5.0%)	27.6 (0.6%)	95.0 (2.0%)	0.0 (0.0%)	4717.0	4164.1
Guinea	261.3 (8.2%)	61.2 (1.9%)	118.0 (3.7%)	2296.7 (72.4%)	419.2 (13.2%)	3.3 (0.1%)	11.4 (0.4%)	2.1 (0.1%)	3173.2	3108.7
Guinea Bissau	136.7 (24.5%)	2.3 (0.4%)	15.5 (2.8%)	181.5 (32.5%)	216.6 (38.8%)	1.2 (0.2%)	3.9 (0.7%)	0.0 (0.0%)	557.7	554.2
Ivory Coast	1636.4 (12.5%)	2619.8 (20.0%)	2515.1 (19.2%)	6147.4 (46.9%)	81.7 (0.6%)	34.0 (0.3%)	69.1 (0.5%)	0.7 (0.0%)	13104.2	10450.4
Kenya	1178.9 (48.1%)	118.8 (4.9%)	213.6 (8.7%)	832.7 (34.0%)	85.7 (3.5%)	13.7 (0.6%)	4.2 (0.2%)	1.6 (0.1%)	2449.2	2316.7
Liberia	8.1 (0.2%)	395.4 (12.0%)	1863.5 (56.6%)	986.7 (30.0%)	3.5 (0.1%)	2.2 (0.1%)	30.7 (0.9%)	0.0 (0.0%)	3290.1	2892.5
Libya*	-	-	-	-	-	-	-	-	-	-
Madagascar	197.1 (1.6%)	667.1 (5.3%)	5126.0 (40.6%)	5627.5 (44.6%)	957.9 (7.6%)	43.5 (0.3%)	1.5 (0.0%)	0.1 (0.0%)	12620.7	11910.1
Malawi	457.8 (40.0%)	185.4 (16.2%)	83.6 (7.3%)	398.3 (34.8%)	8.3 (0.7%)	1.9 (0.2%)	4.6 (0.4%)	4.7 (0.4%)	1144.6	957.3
Mali	480.6 (39.4%)	1.0 (0.1%)	0.0 (0.0%)	708.7 (58.1%)	25.6 (2.1%)	0.6 (0.0%)	2.9 (0.2%)	0.3 (0.0%)	1219.7	1218.1
Mauritania	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.2 (33.3%)	0.4 (66.7%)	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.6	0.6
Morocco*	-	-	-	-	-	-	-	-	-	-
Mozambique*	6419.1 (32.7%)	1527.7 (7.8%)	1967.4 (10.0%)	8959.4 (45.6%)	664.8 (3.4%)	49.2 (0.3%)	37.0 (0.2%)	17.8 (0.1%)	19642.4	18065.5
Namibia*	46.1 (38.0%)	6.8 (5.6%)	0.0 (0.0%)	59.9 (49.4%)	5.1 (4.2%)	2.1 (1.7%)	1.0 (0.8%)	0.2 (0.2%)	121.2	112.3
Niger	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.9 (10.3%)	0.0 (0.0%)	7.8 (89.7%)	0.0 (0.0%)	0.0 (0.0%)	8.7	0.9
Nigeria	2392.5 (34.4%)	1462.3 (21.0%)	452.3 (6.5%)	1754.3 (25.2%)	689.8 (9.9%)	24.4 (0.4%)	122.3 (1.8%)	63.6 (0.9%)	6961.5	5474.8
Rwanda	65.9 (42.7%)	18.2 (11.8%)	16.3 (10.6%)	46.3 (30.0%)	5.8 (3.8%)	0.6 (0.4%)	1.2 (0.8%)	0.0 (0.0%)	154.3	135.5
Senegal	221.4 (30.0%)	0.3 (0.0%)	0.1 (0.0%)	401.5 (54.4%)	112.2 (15.2%)	0.7 (0.1%)	1.6 (0.2%)	0.0 (0.0%)	737.8	736.8
Sierra Leone	11.1 (0.7%)	167.3 (10.1%)	455.9 (27.5%)	995.0 (60.1%)	6.1 (0.4%)	3.2 (0.2%)	16.3 (1.0%)	0.0 (0.0%)	1654.9	1484.4
Somalia	7.2 (11.0%)	10.0 (15.3%)	0.1 (0.2%)	0.9 (1.4%)	46.0 (70.6%)	0.9 (1.4%)	0.1 (0.2%)	0.0 (0.0%)	65.2	54.3
Somaliland	0.0 (0.0%)	0.2 (66.7%)	0.0 (0.0%)	0.0 (0.0%)	0.1 (33.3%)	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.3	0.1
South Africa*	47.5 (8.3%)	2.9 (0.5%)	319.2 (56.1%)	190.8 (33.5%)	2.8 (0.5%)	1.6 (0.3%)	4.2 (0.7%)	0.3 (0.1%)	569.3	564.8
South Sudan	291.4 (21.4%)	64.7 (4.7%)	15.1 (1.1%)	792.5 (58.1%)	161.9 (11.9%)	30.5 (2.2%)	8.2 (0.6%)	0.3 (0.0%)	1364.6	1269.4
Sudan	1.2 (3.7%)	0.2 (0.6%)	0.0 (0.0%)	22.5 (69.7%)	7.7 (23.8%)	0.6 (1.9%)	0.0 (0.0%)	0.1 (0.3%)	32.3	31.5
Tanzania	6677.4 (42.1%)	1013.2 (6.4%)	1397.1 (8.8%)	6436.0 (40.6%)	247.6 (1.6%)	38.0 (0.2%)	28.6 (0.2%)	7.2 (0.0%)	15845.1	14793.9

Region	Cultivated	Forest	Regrowth	Grassland	Shrubland	Water	Artificial	Bareland	Total	Total <sup>†</sup>
Togo	345.5 (48.2%)	65.1 (9.1%)	8.4 (1.2%)	23.8 (3.3%)	270.0 (37.7%)	2.1 (0.3%)	1.7 (0.2%)	0.0 (0.0%)	716.6	649.4
Uganda	777.8 (27.7%)	75.7 (2.7%)	153.3 (5.5%)	1772.0 (63.1%)	9.6 (0.3%)	15.0 (0.5%)	4.7 (0.2%)	0.0 (0.0%)	2808.1	2717.4
West Sahara	-	-	-	-	-	-	-	-	-	-
Zambia	5966.6 (53.5%)	1111.0 (10.0%)	110.1 (1.0%)	3614.2 (32.4%)	216.0 (1.9%)	70.0 (0.6%)	54.5 (0.5%)	0.0 (0.0%)	11142.4	9961.4
Zimbabwe	1662.3 (55.6%)	95.7 (3.2%)	100.1 (3.3%)	1069.8 (35.8%)	51.5 (1.7%)	4.0 (0.1%)	6.5 (0.2%)	0.1 (0.0%)	2990.0	2890.3
Africa	44289.5 (22.8%)	17093.1 (8.8%)	32980.8 (17.0%)	89301.4 (46.0%)	6599.5 (3.4%)	2410.0 (1.2%)	1238.6 (0.6%)	146.0 (0.1%)	194058.9	174555.8
Australia*	0.0 (0.0%)	1.0 (17.2%)	0.6 (10.3%)	0.8 (13.8%)	2.9 (50.0%)	0.2 (3.4%)	0.0 (0.0%)	0.3 (5.2%)	5.8	4.6
Bangladesh*	86.5 (19.4%)	216.9 (48.7%)	85.0 (19.1%)	36.9 (8.3%)	8.2 (1.8%)	2.0 (0.4%)	8.7 (2.0%)	1.1 (0.2%)	445.3	226.4
Brunei	3.4 (1.3%)	24.4 (9.0%)	204.2 (75.3%)	17.5 (6.5%)	0.0 (0.0%)	2.9 (1.1%)	18.7 (6.9%)	0.0 (0.0%)	271.1	243.8
Cambodia	3014.8 (34.0%)	2279.7 (25.7%)	2200.1 (24.8%)	1027.9 (11.6%)	0.0 (0.0%)	327.9 (3.7%)	13.3 (0.2%)	0.0 (0.0%)	8863.7	6256.1
China*	3193.3 (15.4%)	3009.9 (14.5%)	10801.1 (52.0%)	2710.0 (13.0%)	940.1 (4.5%)	57.8 (0.3%)	59.8 (0.3%)	0.2 (0.0%)	20772.2	17704.5
East Timor	4.0 (3.0%)	55.8 (41.3%)	43.3 (32.0%)	29.1 (21.5%)	0.0 (0.0%)	1.9 (1.4%)	0.0 (0.0%)	1.1 (0.8%)	135.2	77.5
India*	658.8 (23.4%)	976.9 (34.7%)	716.9 (25.5%)	290.5 (10.3%)	142.4 (5.1%)	8.6 (0.3%)	20.7 (0.7%)	1.1 (0.0%)	2815.9	1830.4
Indonesia	17415.1 (15.4%)	11227.1 (9.9%)	77352.2 (68.6%)	5209.1 (4.6%)	0.0 (0.0%)	1312.1 (1.2%)	303.0 (0.3%)	17.2 (0.0%)	112835.8	100296.6
Laos	1390.7 (14.8%)	1078.2 (11.5%)	5352.5 (57.0%)	1537.3 (16.4%)	9.7 (0.1%)	11.4 (0.1%)	4.3 (0.0%)	6.4 (0.1%)	9390.5	8300.9
Malaysia	2722.1 (7.5%)	3238.3 (8.9%)	29001.4 (79.4%)	963.6 (2.6%)	0.0 (0.0%)	249.8 (0.7%)	359.4 (1.0%)	0.3 (0.0%)	36534.9	33046.8
Myanmar*	2430.4 (24.5%)	2406.7 (24.3%)	3622.2 (36.6%)	904.1 (9.1%)	383.8 (3.9%)	140.0 (1.4%)	21.1 (0.2%)	0.1 (0.0%)	9908.4	7361.7
Pakistan	0.1 (50.0%)	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.1 (50.0%)	0.0 (0.0%)	0.0 (0.0%)	0.2	0.1
Papua New Guinea	177.6 (3.8%)	1264.6 (26.8%)	2630.7 (55.7%)	318.8 (6.8%)	188.1 (4.0%)	115.8 (2.5%)	14.8 (0.3%)	10.9 (0.2%)	4721.3	3340.9
Philippines	550.9 (12.4%)	1247.3 (28.1%)	2178.2 (49.0%)	423.3 (9.5%)	0.1 (0.0%)	37.4 (0.8%)	7.0 (0.2%)	0.2 (0.0%)	4444.4	3159.7
Saudi Arabia*	-	-	-	-	-	-	-	-	-	-
Solomon Island	0.0 (0.0%)	84.6 (36.2%)	101.2 (43.2%)	19.6 (8.4%)	27.4 (11.7%)	1.0 (0.4%)	0.1 (0.0%)	0.1 (0.0%)	234.0	148.4
Sri Lanka	274.4 (42.0%)	213.5 (32.7%)	108.3 (16.6%)	42.4 (6.5%)	0.0 (0.0%)	5.2 (0.8%)	9.3 (1.4%)	0.3 (0.0%)	653.4	434.7
Taiwan*	24.6 (11.2%)	110.4 (50.3%)	37.8 (17.2%)	45.2 (20.6%)	0.0 (0.0%)	0.7 (0.3%)	0.8 (0.4%)	0.0 (0.0%)	219.5	108.4
Thailand	1943.9 (22.6%)	1989.6 (23.1%)	3463.3 (40.2%)	1090.8 (12.7%)	0.0 (0.0%)	99.2 (1.2%)	29.5 (0.3%)	0.0 (0.0%)	8616.3	6527.5
Vanuatu	0.0 (0.0%)	0.1 (100.0%)	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.0 (0.0%)	0.1	0.0
Vietnam	2928.7 (32.8%)	1102.2 (12.3%)	2754.5 (30.8%)	1635.7 (18.3%)	443.5 (5.0%)	57.9 (0.6%)	19.5 (0.2%)	0.2 (0.0%)	8942.2	7782.1
Yemen*	-	-	-	-	-	-	-	-	-	-
Asia/Australia	36819.3 (16.0%)	30527.2 (13.3%)	140653.5 (61.2%)	16302.6 (7.1%)	2146.2 (0.9%)	2431.9 (1.1%)	890.0 (0.4%)	39.5 (0.0%)	229810.2	196851.1
Global	177038.5 (20.2%)	102412.9 (11.7%)	230543.7 (26.4%)	289445.5 (33.1%)	59005.9 (6.7%)	12011.9 (1.4%)	3690.1 (0.4%)	590.9 (0.1%)	874739.6	