Open-access cloud resources contribute to mainstream REDD + : The case of Mozambique

Catarina Lopes[a,1,*], Ana Leite[a,b,1], Maria José Vasconcelos[b]

[a] Remote Sensing, Environment and Technology for Development (RSeT), NGO for International Cooperation, Lisbon, Portugal
[b] University of Lisbon, School of Agriculture, Forest Research Center (CEF), Lisbon, Portugal

A B S T R A C T

The objective of this work is to investigate how Earth Observation data and processing platforms accessible on the cloud can facilitate the implementation of REDD + in developing countries. For that, we explore newly available open-access satellite data, cloud processing, and a ready-made land cover map to assess the extent to which such resources can directly respond to monitoring and measuring, reporting, and verification requisites. Mozambique is one of the 47 countries selected to benefit from the Forest Carbon Partnership Facility to implement REDD + strategies. However, to meet funding agreements, the country needs to periodically produce national land cover and land cover change maps at given resolution and accuracy levels. The work presented here shows that the land cover mapping requisites of REDD + may be quickly and cost-effectively met through the development and use of newly available cloud resources. The study relies on an experimental design that tests the results of image processing approaches with algorithms available or developed in Google Earth Engine against country-wide reference data collected by a team of national experts. The results show that, in addition to pre-processing advantages, which facilitate multi-temporal compositing and mosaicking of very large and heavy data sets, developments in cloud processing and image classification swiftly produce large extent and high-resolution land cover maps, tailored to a specific objective. The comparison of results between the in-house map obtained using Google Earth Engine, and the pan-African map produced by the European Space Agency (2016) at the same spatial resolution, reveals that both maps meet REDD + requirements for a binary Forest/Non-forest legend. However, the in-house map is more accurate and reaches considerably better results if a more complex, six class IPCC legend is required. Nevertheless, this study shows that, given adequate reference data, the need to periodically produce high resolution land cover maps for national forest monitoring purposes is no longer an obstacle for mainstreaming the implementation of REDD +.

1. Introduction

1.1. Context and background

Deforestation is reported as the second largest anthropogenic source of carbon dioxide (CO_2) to the atmosphere, accounting for around 12% of global gross emissions in the 1990s and 2000s (Van der Werf et al., 2009). Although only 14% of this global forest loss is reported to occur in sub-Saharan Africa (Harris et al., 2012), its consequences are particularly relevant here, where agriculture is central to the livelihoods of most of its population and the current rates of deforestation are increasing (Sitoe et al., 2012).

Reducing emissions from deforestation and forest degradation, conserving and enhancing forest carbon stocks, and sustainably managing forests represent crucial opportunities for developing countries. Such a path not only facilitates the engagement in global initiatives for combating climate change, but it also induces the installment of financial flows for sustainable, low carbon, development. Through international financing mechanisms, such as the "Reducing Emissions from Deforestation and Forest Degradation and the role of Conservation, Sustainable Management of Forests and Enhancement of Forest Carbon Stocks in Developing Countries" (REDD +), developing countries can discuss the future of their forests at a higher level. This context facilitates the adoption of actions for mitigating the effects and promoting adaptation to climate change, while inducing sustainable land use management and improving local livelihoods.

REDD + requires that the cause, magnitude, and location of emissions and removals of Greenhouse Gases (GHG) from forests be

---

* Corresponding author at: NGO RSeT, Rua Fialho de Almeida, 14, 2º E, Esc. Q18, 1070-129 Lisbon, Portugal.
E-mail address: catarinagouveialopes@gmail.com (C. Lopes).
[1] These authors contributed equally to this work.

characterized and periodically quantified through credible estimates at national and sub-national level. However, due to major data gaps and to technical and technological complexities, this has not been easy to achieve in tropical Africa. Thus, the establishment of credible national forest reference levels (FRLs) and the development of adequate forest monitoring (M) systems, which can respond to international standards, has been a relevant obstacle for the implementation of REDD + in tropical countries (Romijn et al., 2012; Ochieng et al., 2016) such as Mozambique.

Mozambique's natural resources are being rapidly depleted: 138 000 ha of natural forests (approximately 0,3%) are lost every year, and erosion is pervasive (MITADER, 2017). Thus, the development of measures for conservation of the natural resource base that sustains agriculture and forestry, particularly soils and water, is critical. To address these challenges, the Ministry of Land, Environment and Rural Development (MITADER) adopted several strategic actions. One of these actions is the implementation of REDD +, which requires that key-components, allowing the official accounting of emissions from the land use sector, to be developed. Therefore, with the support of the Forest Carbon Partnership Facility (FCPF) of the World Bank, MITADER is organizing and developing a multi-stakeholder National Forest Monitoring System (NFMS) and an operational Measuring, Reporting and Verification (MRV) System (see www.forestcarbonpartnership.org/mozambique for all the documentation related with the National REDD + Strategy in Mozambique). These key components (M&MRV) will produce the national land cover change analysis and will support forest monitoring activities, such as the National Forest Inventory and the constitution of the FRL.

Our work concentrates on developing, analysing and proposing simple, fast and low-cost solutions to facilitate the compliance of REDD + monitoring requisites adequate for the M&MRV activities in Mozambique. More specifically, the work delivers the results of a preliminary study designed to assist decisions regarding the development of technical and technological configurations for the operational production of wall-to-wall land cover maps.

1.2. The study area

Mozambique is a tropical to sub-tropical country richly endowed with natural resources and home to important biodiversity hotspots. The country´s mainland covers 786 380 km² of a total 799 380 km², with about 13 000 km² of insular area (CIA, 2018) and a coastline of 2 700 km (DSU, 2015). Elevation has a mean of 345 m, reaching the lowest point at Indian Ocean and the highest point at Monte Binga (2 436 m) (World Bank, 2017). Mozambique´s climate includes two marked seasons: a wet season from October to March and a dry season from April to September. The lowest average rainfall occurs in southern regions, which are drier, more so inland than towards the coast (300–800 mm/year). Average temperatures are highest also along the coast (25–27 °C in summer and 20–23 °C in winter), as well as in the southern region (24–26 °C in summer and 20–22 °C in winter). High inland regions have cooler temperatures (GFDRR, 2011).

Mozambique has a varied array of ecosystems and is biologically diverse, encompassing 13 World Wildlife Fund (WWF) Terrestrial Ecoregions (Olson et al., 2001). As illustrated in Fig. 1, the northern areas are predominantly occupied by miombo woodlands and replaced by Zambezian and Mopane woodlands in the western and southern borders. The most widespread vegetation in the north coast is the Zanzibar-Inhambane forest mosaic followed by the African mangroves and the Maputaland forest mosaic in the south-east coast.

Despite Mozambique's relevant quantities of arable land, forests, fisheries, water and mineral resources, which translate into potentially significant economic returns, including from agriculture and forestry (DSU, 2015), complex socio-economic factors and the frequent occurrence of natural hazards, such as droughts, floods and cyclones, have had an impact in shaping the country's poverty and vulnerability
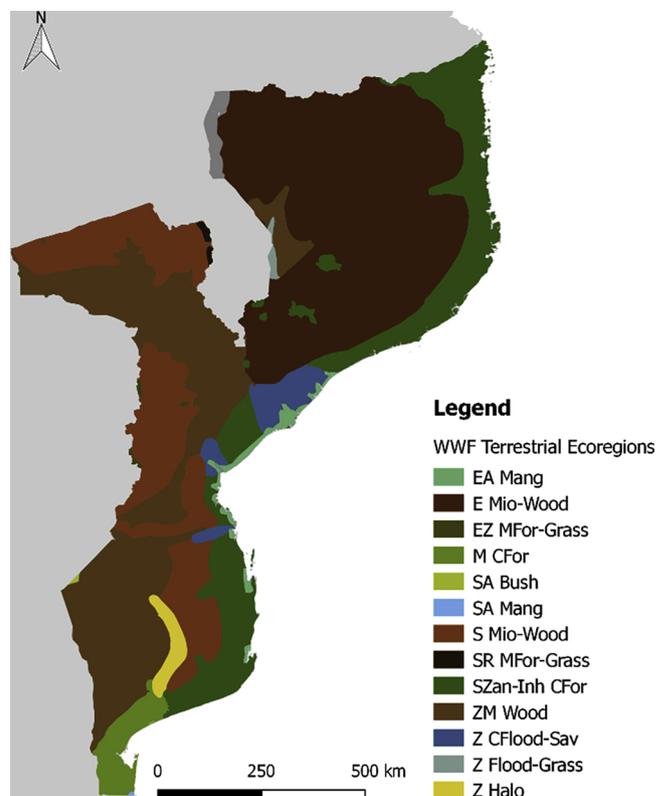

Fig. 1. WFF Terrestrial Ecoregions of Mozambique - EA Mang, East African Mangroves; E Mio-Wood, Eastern Miombo Woodlands; EZ MFor-Grass, Eastern Zimbabwe Montane Forest Grassland mosaic; M CFor, Maputaland Coastal Forest Mosaic; SA Bush, Southern Africa Bushveld; SA Mang, Southern Africa Mangroves; S Mio-Wood, Southern Miombo Woodlands; SR MFor-Grass, Sothern Rift Montane Forest Grassland Mosaic; SZan-Inh CFor, Southern Zanzibar-Inhambane Coastal Forest Mosaic; ZM Wood, Zambezian and Mopane Woodlands; Z CFlood-Sav, Zambezian Flooded Grasslands; Z Flood-Grass, Zambezian Flooded Grasslands; Z Halo, Zambezian Halopytics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

situation (Artur and Hilhorst, 2012). Since subsistence agriculture is mainly rain-fed and highly dependent on natural resources, it is significantly exposed to climate variability and to the effects of climate change. This is particularly relevant given socio-economic factors that also contribute to increase local vulnerability and to decrease the capacity for adaptation. Furthermore, population growth (2,5% per year) also contributes to increase the pressure on natural resources (DSU, 2015).

1.3. Problem description

To clarify and organize the main issues associated with the development of a national participatory M&MRV, the national REDD + strategy (2016) includes a Road Map where the main definitions are presented, and activities are planned (MITADER, 2016). The national M &MRV system should, first and foremost, be able to estimate greenhouse gas emissions from deforestation and forest degradation with internationally accepted transparency, consistency, technical-methodological robustness and credibility. For that, two basic inputs must be collected and analyzed at national and sub-national level: Activity Data (AD), which is the areal extent of deforestation, forest degradation or forest enhancements; and Emission Factors (EF) which correspond to carbon stock change factors.

Of the possible options to assess AD, the MRV Road Map specifies that a spatially explicit tracking of land-use conversions over time

should be employed so that extended applications to the entire Agriculture, Forestry and Other Land Use (AFOLU) sector, following the guidelines of the International Panel on Climate Change (IPCC), can also be achieved based on this data. This option can be fulfilled by implementing two different approaches, both accepted by the FCPF Carbon Fund (CF) Methodological Framework (MF) (FCPF, 2016) and the Verified Carbon Standard (VCS) Jurisdictional and Nested REDD + (JNR) (VCS, 2017). The first is based on systematic spatial sampling. The other is based on wall-to-wall mapping using satellite based spatial tracking.

Regarding the wall-to-wall land cover mapping requisites for Mozambique, a Forest/Non-forest (F/NF) benchmark map of 2016, derived from Sentinel-2 (S-2) imagery and a minimum overall accuracy (OA) of 75%, is required for monitoring purposes. Nevertheless, the disaggregation of the non-forest class (into cropland, grassland, settlements, wetlands and other land) is desirable given that it constitutes a requirement of the 2006 IPCC Guidelines for reporting purposes (MITADER, 2016).

The national definition of forest should be used consistently over time for all REDD + activities. According to the MRV Road Map, forest is defined as follow: "*forest are lands that occupy at least 1 ha with canopy cover greater than 30%, and with trees with potential to reach a height of 3 m at maturity, temporarily cleared forest areas and areas where the continuity of land use would exceed the thresholds of the definition of forest, or trees capable of reaching these limits in situ*".

Satellite-based spatial tracking of land cover is recognized as the most consistent, efficient, comprehensive and cost-effective approach to produce recurrent wall-to-wall AD (deforestation/reforestation/afforestation) (GOFC-GOLD, 2009). However, relying on classification of satellite images for tracking land cover change in a country as large and complex as Mozambique presents many technical and technological challenges. Not only is the handling of vast amounts of digital images a technically heavy-duty endeavour, but it also entails high operational costs (infrastructure, man-power, knowledge, time/effort). Thus, the recent release of Google Earth Engine (GEE) (Google Earth Engine Team, 2015), an open-access online image hub and processing platform, is worth exploring. Several studies from the regional (Johansen et al., 2015; Dong et al., 2016; Huang et al., 2017) to the continental (Midekisa et al., 2017; Xiong et al., 2017) level, point out GEE capabilities as being possibly useful for producing fast and reliable products for the AFOLU sector.

On another hand, the European Space Agency (ESA) launched the Climate Change Initiative (CCI) in order to provide adequate response for long-term satellite-based products for climate that is concerned with addressing explicit needs of UNFCCC. In this context, ESA produced a land cover prototype map at 20 m resolution over Africa based on the classification of a set of Sentinel-2 A (S-2 A) images corresponding to one year of observations (from December 2015 to December 2016). This prototype product was released to collect user's feedback for further improvements. There is a web interface to visualize and interact with the data available at http://2016africalandcover20m.esrin.esa.int/viewer.php (CCI Land Cover team, 2016).

The aim of this study is to provide a base for the fulfilment of AFOLU wall-to-wall monitoring requisites in Mozambique and illustrate the usefulness, cost-effectiveness, and efficacy of open source data and technology in fulfilling the requirements of specific international policies, while contributing to advance the achievement of national development goals. The main objectives are (1) to investigate how Earth Observation data and online processing platforms freely accessible on the cloud can facilitate the gathering of wall-to-wall information (adequate for unleashing the process of generating AD) and to evaluate if the result fulfils REDD + requisites for Mozambique; and (2) to evaluate the extent to which a freely available land cover map – the 20 m resolution ESA CCI prototype map for 2016 – is able to fulfil the REDD + requisites for Mozambique; and (3) to compare the results of objective (1) and (2).

2. Methods

2.1. Data and implementation tools

2.1.1. The reference data set

The MRV Road Map presents three hierarchical levels (level 1: IPCC legend with 6 classes; level 2: national classification legend with 21 classes; and level 3: national classification legend with 42 classes) for the land cover classification system in Mozambique (presented in Table 4 of MITADER (2016) MRV Road Map). The reference data set used in this study was produced based on a classical on-screen delineation of polygons over satellite imagery through a purposive homogeneous sampling of each legend category at level 3, in order to provide a balanced, national scale coverage for all classes. For that, a team of five experienced experts from the official MRV team of Mozambique analyzed Very High-Resolution images and manually digitized 28 317 polygons, ranging between 0,1 ha to 100 ha.

2.1.2. Satellite images

The work presented here relies on two different image collections available through GEE. The first is derived from the MODerate-resolution Imaging Spectroradiometer (MODIS) mounted on both the Terra and Aqua satellites, which deliver complete coverage twice a day. The second results from S-2, which was recently made available by ESA, and delivers complete coverage every 5 to 10 days.

In this study we use the combined 16-day Normalized Difference Vegetation Index (NDVI) product derived from MODIS MCD43A4 (NASA LP DAAC, 2015). This includes a time series of 500 m resolution composite images, each resulting from the processing of daily surface reflectance observations, acquired by Terra and Aqua sensors, within a 16-day period, from which the best representative pixel is selected (Schaaf and Wang, 2015). The daily composites, produced in the sinusoidal projection, minimize clouds and cloud shadows and are used to produce the NDVI time series. NDVI is computed using the near-infrared and red bands of each composite image, ranging between -1,0 and 1,0.

The S-2 sensor is identified as the main source of remotely sensed data for deriving land cover information in the MRV Road Map of Mozambique. The mission was launched on the 23rd June 2015 and provides a global coverage of the Earth's land surface every 10 days with S-2 A satellite and every 5 days with S-2 A and Sentinel-2B (S-2B, launched on the 7th March 2017). This mission offers high-resolution optical imagery with a spatial resolution of 10 m, 20 m and 60 m and multi-spectral data with 13 bands in the visible, near-infrared and short-wave infrared parts of the spectrum. The processing Level-1C of S-2 A, available in GEE, is used in this study to generate the land cover classification. The Level-1C product is composed of $100 \times 100 \ km^2$ tiles (ortho-images in UTM/WGS84 projection) in top-of-atmosphere reflectance.

2.1.3. Feature data sets

The administrative boundaries of Mozambique used in our study to define the spatial extent of the land cover maps are freely available at GADM, 2015. This data set also contains lower level subdivisions such as provinces and districts. Additionally, the WWF Terrestrial Ecoregions boundaries of Mozambique included in this study, were acquired from a vector data set that contains the 825 terrestrial ecoregions of the world nested within two higher-order classifications: biomes (14) and biogeographic realms (8) (Olson et al., 2001).

The 20 m resolution ESA CCI prototype map for 2016, henceforth designated by ESA CCI map, uses a legend that includes 10 classes: "trees cover areas", "shrubs cover areas", "grassland", "cropland", "vegetation aquatic or regularly flooded", "lichen and mosses/ sparse vegetation", "bare areas", "built up areas", "snow and/or ice" and "open water". To produce this map, the CCI team applied two different classification procedures to the S-2 A multi-temporal data set in top-of-canopy reflectance to generate two preliminary maps. The first is a supervised random forests algorithm and the second consists of

unsupervised machine learning algorithms. The two preliminary maps were then combined, using a rule-based procedure, to select the best representation of a land cover class (Ramoino et al., 2016). For a full description of the algorithms used see Ramoino et al. (2016). The ESA CCI map is freely available and can be downloaded with a size of approximate 6GB (CCI Land Cover team, 2016). For this study we used ESA CCI prototype map version 1.0.

2.2. Experimental set-up

2.2.1. Organization of the legend

A fundamental step in our analysis consists on the organization of the legend established in the MRV Road Map for land cover mapping (Annex A - Table A1). The reference data set used in this study was compiled by the MRV team at level 3 of the legend (presented in Table 4 of MITADER (2016) MRV Road Map) and, as a pre-processing step, it was then aggregated by the MRV team and provided to us at level 2. An aggregation of classes into level 1 categories is needed to fulfill three objectives: *i)* To improve the separability of the land cover classes in spectral / feature space; *ii)* To obtain consistency in the spectral dynamics – phenology - within each class, such as observed in same-year multi-date imagery; and *iii)* To ensure compatibility and comparability with the legend of other ready-made land cover maps available from international agencies, such as the ESA CCI map. Nevertheless, since the legend structure is hierarchical, several exercises can be easily attempted at different levels of the legend to address the targeted land management purposes. For REDD + purposes, the legend can be as simple as F/NF.

Since the MRV Road Map states that the non-forest classes should be disaggregated into the IPCC legend (level 1), our validation assessments target two levels: level 1 (IPCC) and level 0 (F/NF). Table A1 illustrates the aggregation applied to the MRV Road Map land cover legend into F/NF and the correspondence of categories with the recently produced ESA CCI map. The codes along the ESA CCI map column (Table A1) represent the original class values of the product. For the purpose of this study, the classes were reclassified into level 1 and level 0 (e.g. code 4 stands for "Cropland" within the original ESA CCI map, but it was reclassified into code 1 - "Cropland" for level 1 and into code 2 - "Non-Forest" for level 0). The level 2 corresponds to the data polygons collected and provided by the MRV team and used here as the reference data set. Those polygons were reclassified into level 1 and level 0 according to the classes presented in Table A1 (e.g. codes 31 – "Grasslands", 32 – "Thicket" and 33 – "Shrubland" in level 2 were reclassified into code 3 - "Grassland" for level 1 and into code 2 – "Non-Forest" for level 0).

2.2.2. Constitution of the training and validation data sets

The training set to build the image classifiers relied entirely on a sample of 7 113 polygons (covering approximately 190 545 pixels)
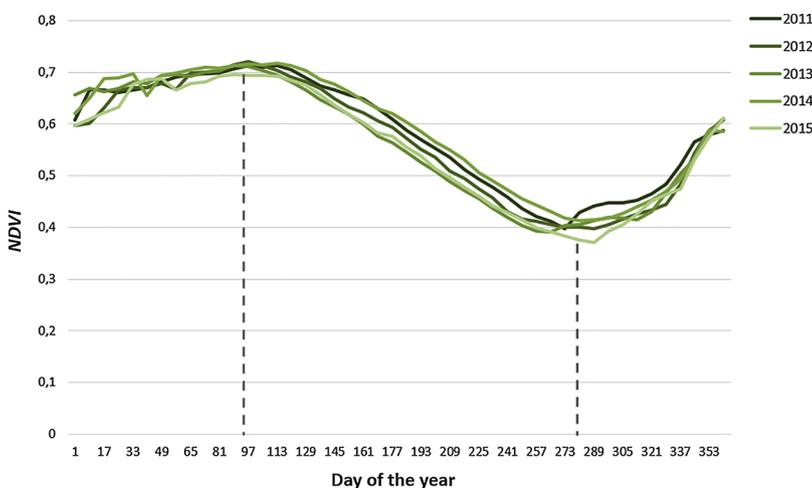
resulting from the application of a stratified random sampling procedure to the reference data set. Computation limits of GEE determine the maximum number of pixels that can be used for training. The validation data set corresponds to a stratified random selection of 20 837 polygons (covering approximately 561 606 pixels) from the remainder set of reference polygons.

To address the two levels of the legend used in this study (level 1 – IPCC and level 0 – F/NF), the polygons selected for training and validation were reorganized following the hierarchy and according to the codes shown in Table A1 and as explained in 2.2.1.

2.3. Map production procedures

2.3.1. Production of a cloud and shadow free 2016 image mosaic of Sentinel-2 images covering the entire country

The production of a radiometrically harmonious image mosaic, which is mostly cloud and shadow free, and does not present inter-frame radiometric differences or frame boundary seams, is very important for country-wide classification purposes. Thus, the production of a good-quality S-2 mosaic that results from the best possible compositing of multi-temporal S-2 images using GEE, is worth exploring.

Compositing techniques have been widely used to produce good quality image mosaics for large extents in tropical regions (Mayaux et al., 1999; Cabral et al., 2003; Stibig et al., 2003; Eva et al., 2004). The process consists of composing a new image by selecting the best possible pixel from a set of same position pixels in a multi-temporal image stack. In tropical regions, we can optimize the set of images to be used in a compositing process by excluding those corresponding to the wet-season. Additionally, profiling *"large units of land containing distinct assemblages of natural communities sharing a large set of species, dynamics, and environmental conditions"* (Olson et al., 2001) separately may reduce radiometric effects or boundary seams caused by phenological differences in the composite images.

To implement the above principles, a set of multi-temporal S-2 A images retrieved between the peak and the end of the growing season were processed in GEE for 2016. However, the length of the growing season varies significantly across the territory of Mozambique, which spans a wide range of ecosystems as illustrated in Fig. 1. Hence, the image compositing outcomes may be optimized if the dates for the peak and the end of the growing season vary spatially according to the ecoregions of the country.

To determine the dates for the peak and end of the growing season of each ecoregion of Mozambique we analyzed the MODIS Combined 16-day NDVI product in GEE. The NDVI time-series profile (from 2011–2015) was computed for each WWF Terrestrial Ecoregion and the period between the peak and end of the growing season was extracted for each region. Fig. 2 represents the peak and end of the growing

Fig. 2. NDVI Trends for Southern Zanzibar-Inhambane coastal forest mosaic WWF terrestrial ecoregion. Dashed gray vertical lines define the period of the year considered for applying the processing algorithm (from the peak of the growing season to the end of the season). The green lines show the daily NDVI behaviour for the years 2011–2015 for the same ecoregion (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

season for one of the ecoregions in Mozambique within 5 years ("Southern Zanzibar-Inhambane coastal forest mosaic WWF terrestrial ecoregion"). This procedure was carried out for all WWF ecoregions that occur in Mozambique to define the period in which the S-2 images would be selected to build the composite image for 2016. Taking the example presented in Fig. 2, S-2 images for that specific ecoregion were selected between day 97 and day 280 (approximately).

S-2 A imagery from the periods determined through the NDVI time-series for each ecoregion was processed using an algorithm that selects the best available cloud-free pixel for the given period. The algorithm is an adaptation of the Landsat-based *Phenology Based Synthesis Classification* using Google Earth Engine (Simonetti et al., 2015), which generates cloud/shadow S-2 masks and excludes non-eligible pixels in a compositing procedure that is based on the computation of the median value among the available dates for each pixel. The procedure is driven by a predefined rule-based reasoning built upon spectral properties, together with morphological filters to fill holes and buffering the edge, usually characterized by thin clouds or haze that causes classification confusions (Simonetti et al., 2015).

After compositing all bands individually at their original resolutions, the blue, green, red and near-infrared bands of S2-A (originally 10 m resolution) were resampled to 20 m using the mean value of the contributing original pixels and, together with the vegetation red edge bands, narrow infra-red, and the short-wave infrared bands, were arranged in a final resulting image. This image at 20 m resolution serves as an input for the classifications that, after post-processing, result in the final land cover map. The processing chain for producing the image mosaic is illustrated in Fig. 3. The GEE script used to build the mosaic over Mozambique is available at: https://code.earthengine.google.com/f3ed7c8622b50202f7fbd376baba6103.

2.3.2. Classification of the Sentinel-2 mosaic

Machine learning (non-parametric) algorithms, such as artificial neural networks, decision trees, support vector machines and ensembles of classifiers, have been emerging as appropriate and efficient alternatives to conventional parametric algorithms (e.g. maximum likelihood), because non-parametric algorithms do not make any assumptions regarding frequency distribution of the remotely sensed data, which rarely have normal distributions. Furthermore, ensemble learning algorithms like random forests, are more accurate and robust to noise than single classifiers (Rodriguez-Galiano et al., 2012; Belgiu and Drăguț, 2016).

In our study we used random forests in a supervised classification procedure. A random forest classifier consists of a combination of tree classifiers where each classifier is generated using a random vector sampled independently from the input vectors. Then, each tree of the ensemble casts a unit vote to classify the input vectors (Pal, 2005) corresponding to each pixel. The most popular class (majority of votes)

is attributed to the pixel. This algorithm can run efficiently on large data bases; can handle thousands of input variables without variable deletion; gives estimates of what variables are important in the classification; it generates an internal unbiased estimate of the generalization error; it computes proximities between pairs of cases that can be used in location outliers; it is relatively robust to outliers and noise; and it is computationally lighter than other tree ensemble methods (Rodriguez-Galiano et al., 2012).

A supervised classification was performed using the GEE random forest classifier "*ee.Classifier.randomForest*" to generate a land cover map corresponding to level 1 of the legend. The number of trees was set to 30, which is the maximum possible in GEE for the number of training pixels used here, and the remainder arguments were left with default values.

2.3.3. Post-processing

In order to obtain a final map compliant with the minimum mapping unit (MMU) of 1 ha, a post-classification operation was applied. This operation identifies raster clusters that are smaller than 1 ha (less than 25 contiguous pixels) and replaces the corresponding pixels with the value of the largest neighbouring cluster. However, GEE does not provide a method to easily implement this operation nor does it have sufficient computation power to perform such neighbour-based operations in a large data set context. The classified maps were exported to a Desktop-PC and the Geospatial Data Abstraction Library (GDAL) "*Sieve*" command available in Quantum GIS (QGIS) was used for the procedure. The result was imported again into GEE. The exact same procedure was applied to the ESA CCI map.

To assess the results for level 0 of the legend, for both the In-house Land Cover map and the ESA CCI map over Mozambique, the level 1 maps were reclassified following the hierarchy presented in Table A1.

2.4. Validation of the land cover maps

The validation of the land cover maps followed the good practices for estimating area and assessing accuracy (Olofsson et al., 2014). As explained in Section 2.2.2, after selecting the training polygons from the reference data set, the validation sample sizes were defined based on the remainder set of polygons available. Both training and validation data sets were selected using a stratified random design. Since the validation sample sizes were already defined (20 837 polygons), the approach was to evaluate if the validation data was sufficient to reach a given standard error. For that, we used the sample estimation formula from Olofsson et al., 2014.

Following the recommendations of Olofsson et al. (2014), the total number of samples needed to obtain an unbiased estimate of map areas was assessed for all maps. Additionally, to define the number of validation plots necessary for each map, a target standard error of 0,01 for
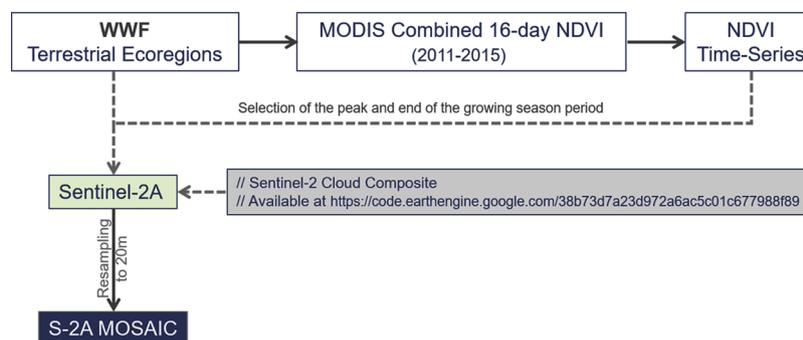

Fig. 3. Mosaic processing chain. The NDVI time-series profile is used to select the peak and end of the growing season period for each WWF Terrestrial Ecoregion that spans Mozambique. Sentinel-2 A imagery is then processed using rules in a pixel-based algorithm built upon spectral properties.

the OA was defined.

The expected value for the user's accuracy (UA) for each strata (class) was defined to be 0,7 for the "Non-Forest" classes ("Cropland", "Grassland", "Wetlands", "Settlements" and "Other Land") and 0,8 for "Forest". The result showed that the total number of validation samples needed for the *In-house Land Cover map at level 1 (Map A)* and for the *In-house Land Cover map at level 0 (Map C)* is 1896 and for the *ESA CCI map at level 1 (Map B –)* and the *ESA CCI map at level 0 (Map D)* is 1920. Considering the results obtained for the stratified random sample design, the validation data set that we are using, which contains 20 837 polygons, is suitable for providing an unbiased accuracy assessment and estimating map areas.

Each map was compared against the validation data set and error matrices were computed using pixel counts by applying the method "ee.errorMatrix" in GEE. Because of computation limitations an error matrix was generated for each province separately and then the individual results were added to obtain a single error matrix for each map.

Error matrices in terms of estimated area proportions were then derived for each map along with parameters such as OA, UA and producer's accuracy (PA). The area for each class was estimated according to the classification based on the validation data. Since the map classes were defined as strata and a stratified random sampling was applied, we used a stratified estimator of OA (see Olofsson et al., 2014: Eqs. (5)) and a stratified estimator of area (see Olofsson et al., 2014: Eq. (9) and (10)) to compute the standard errors. The Margins of Error (ME) for a 95% Confidence Interval (CI) were assessed for the OA and for the area estimates.

In addition, a logistic Geographically Weighted Regression (GWR) was used to analyse spatial variations in the OA for both land cover maps at level 1. The GWR approach uses a moving kernel window to compute local estimates of the regression coefficients and applies a distance weighting to the data. The validation class (dependent variable) was logistically regressed on the classified class (independent variable) using the centroid within each validation polygon and the classified data at that exact point. The logistic geographically weighted regressions were calculated as follows:

$$pr\,(y_i = 1) = \mathrm{logit}(\,b_{0_{(u_i,v_i)}} +\ b_1 x_{1_{(u_i,v_i)}}\,)$$

where $pr\,(y_i = 1)$ is the probability that the validation class is present, x_{1_i} is the independent variable (the presence of the classified class) and the coefficient estimates for the independent variable are assumed to vary across the two-dimensional geographical space defined by the coordinates (u, v). Thus, the coefficient estimates in the logistic GWR are functions of these coordinates. In this case we calculated local models at each location on a 1 km grid. These calculations were performed in R (packages "rgdal", "spgwr" and "GISTools"). More detailed information can be found in Comber et al. (2012) in the reference section.

3. Results and discussion

3.1. Mosaicking

S-2 cloud-free mosaics over the tropical belt will be soon available for download and are currently accessible for visualization at: http://forobs.jrc.ec.europa.eu/recaredd/map/. The later results use equivalent pixel-based algorithms to those used here, but select from a much wider set of images, covering all the S-2 A and S-2B imagery available from October 2015 to October 2017. The MRV Road Map for Mozambique identified the need to produce a benchmark map for 2016, which will complete the historical AD analysis and serve as a starting point for MRV purposes of all REDD+ activities. This requirement reduces significantly the number of cloud-free pixels, candidates in the compositing process (median value of all S-2 A data after cloud masking) and increases the likelihood of having some positions with no data. However, selecting imagery available from the peak to the end of the growing season, assuming similar biological features, excludes the most clouded scenes from the selection domain and reduces boundary seam effects. The mosaic built in this study had a few pixels with no data (corresponding to 0,0016% of the total area), mostly in sand banks, where the algorithm confuses the spectral signature with clouds, an aspect that could represent a challenge for local or regional applications. This approach generates homogeneous single-year clean mosaics of S-2 (Fig. 4), which that can be used for monitoring purposes, without almost any human interaction. The inclusion of S-2B imagery in the selection domain could potentially improve the quality of the mosaic by duplicating the number of selection possibilities and will be explored in the future.

Before the availability of S-2 in GEE, two 2016 bottom-of-atmosphere S-2 A mosaics were produced for the wet and dry season using traditional processing approaches (desktop solutions running proprietary software). However, not only do the resulting mosaics present relevant radiometric problems with sharp differences among S-2 A neighbouring frames, but also still have abundant clouds and shadows. Additionally, the software and computing arrangements needed, as well as the complexity of the processing workflow, decrease the possibility of repetitions.
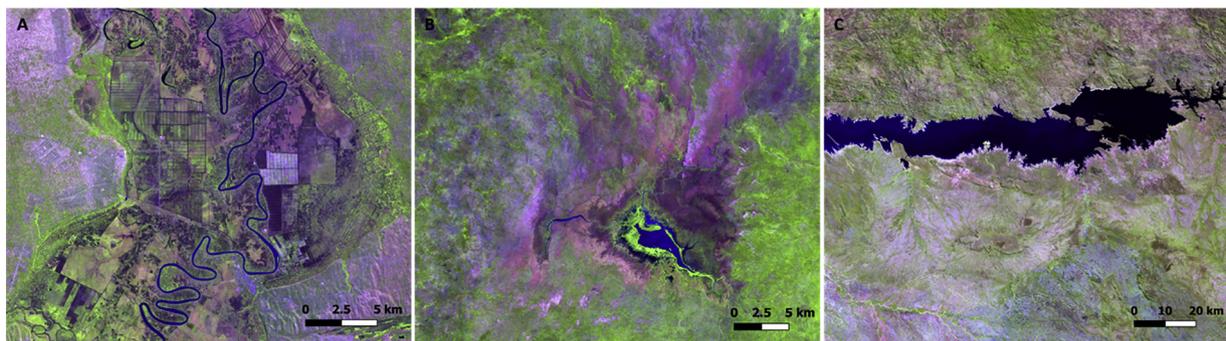

Fig. 4. Mosaic parts displayed with a RGB false color composite using top-of-atmosphere reflectance of S-2 bands B11, B08, and B04 at spatial resolution of 20 m for A) Gaza Province, B) Gorongosa Park and C) Cahora Bassa Lake.

3.2. Classification and map validation

The final maps are presented in Fig. 5 and their respective accuracy assessment is presented in Table 1, where the area estimates per class are presented in proportion to the total area along with PA, UA and OA. The ME was assessed for a 95% CI. The error matrices in terms of area estimates are presented in Annex A - Tables A2 and A3. All maps are publicly available as GEE assets at:

- Map A – In-house Land Cover map at level 1: (https://code. earthengine.google.com/?asset=users/catarinagouveialopes/1_

- Paper/Mozambique/Moz_LandCover_MMU_20m_L1).
- Map B – ESA CCI map at level 1: (https://code.earthengine.google. com/?asset=users/catarinagouveialopes/1_Paper/Mozambique/ Moz_ESA_CCI_MMU_20m_L1).
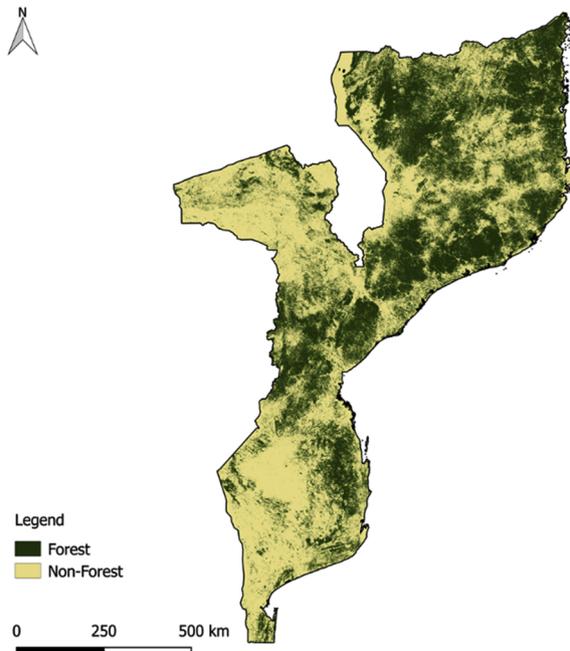- Map C– In-house Land Cover map at level 0: (https://code. earthengine.google.com/?asset=users/catarinagouveialopes/1_ Paper/Mozambique/Moz_LandCover_MMU_20m_L0).
- Map D – ESA CCI map at level 0: (https://code.earthengine.google. com/?asset=users/catarinagouveialopes/1_Paper/Mozambique/ Moz_ESA_CCI_MMU_20m_L0).
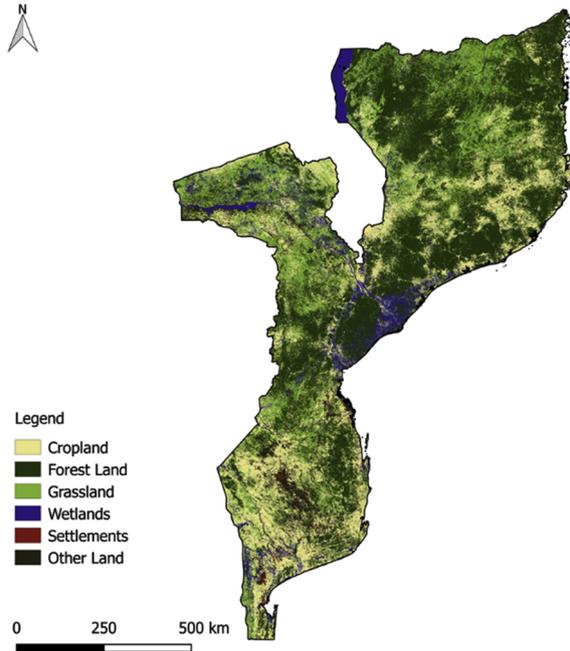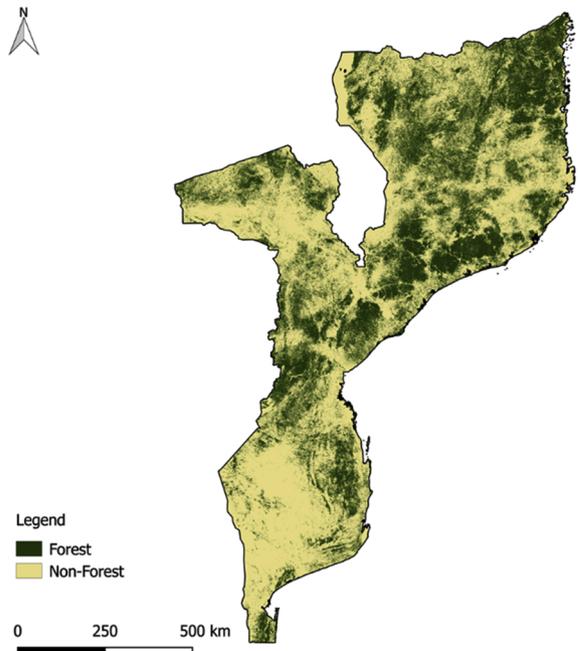

Fig. 5. Land cover maps: A: Level 1 - In-house Land Cover map (20 m); B: Level 1 - ESA CCI map (20 m); C: Level 0 - In-house Land Cover map (20 m); D: Level 0 - ESA CCI map (20 m).

Table 1
Level 1 and Level 0 accuracy assessment for both maps: Estimated area per class in proportion to the total area, overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA) and margin of error (ME) for a 95% confidence interval.

Level 1	A: 20 m In-house Land Cover map				B: 20 m ESA CCI map			
	Estimated area ± ME (%)	PA (%)	UA (%)	OA ± ME (%)	Estimated area ± ME (%)	PA (%)	UA (%)	OA ± ME (%)
Cropland	20,1 ± 2,7	79,8	58,8	**70,7 ± 0,7**	18,1 ± 3,1	31,9	41,9	**45,5 ± 0,6**
Forest	37,8 ± 1,2	88,1	85,2		36,8 ± 1,3	80,7	86,2	
Grassland	15,8 ± 3,7	69,6	55,7		12,0 ± 4,4	74,8	17,9	
Wetlands	11,5 ± 3,7	50,0	72,5		15,3 ± 3,5	55,5	55,5	
Settlements	3,5 ± 7,7	6,2	83,3		3,2 ± 8,3	3,8	90,4	
Other Land	11,4 ± 4,0	39,3	77,0		14,7 ± 3,7	0,3	80,0	
Level 0	C: 20 m In-house Land Cover map				D: 20 m ESA CCI map			
	Estimated area ± ME (%)	PA (%)	UA (%)	OA ± ME (%)	Estimated area ± ME (%)	PA (%)	UA (%)	OA ± ME (%)
Forest	37,0 ± 1,0	90,1	85,2	**90,5 ± 0,4**	36,5 ± 1,2	81,4	86,2	**88,5 ± 0,4**
Non-Forest	63,0 ± 0,6	90,8	94,0		63,5 ± 0,7	92,5	89,6	

As illustrated in Fig. 5, level 0 land cover classes have a similar spatial distribution between the In-house Land Cover map and the ESA CCI map, reaching approximately 90,5 ± 0,4% and 88,5 ± 0,4% in OA, respectively and having very similar estimated areas for the two classes. Both maps fulfill the REDD+ requisites (OA > 75%) for the F/NF classes. However, when comparing level 1 maps, large differences in spatial arrangement and proportion are observed in the non-forest classes. The accuracies achieved in the In-house Land Cover map are significantly higher than those achieved in the ESA CCI map. By looking at the accuracy assessment results of the two maps at level 1 (Table 1), one can see that higher PAs are present in the In-house Land Cover map for "Cropland" (79,8% vs. 31,9%), "Forest" (88,1% vs. 80,7%), "Settlements" (6,2% vs. 3,8%) and "Other Land" (39,3% vs. 0,3%) when compared to the ESA CCI map. Both maps revealed low commission errors, however in the ESA CCI map the class "Cropland" revealed a lower UA (41,9%) and it is largely misclassified by confusion with "Forest" and "Other Land". The class "Settlements" presents a very low PA, both in the In-house Land Cover map (PA = 6,2%) and in the ESA CCI map (PA = 3,8%), which entails that both maps underestimate settlement areas. This difficulty maybe a result of the nature of settlements in most rural areas, with houses built using raw materials (such as adobe walls and straw roofs) and spread over a background (mostly soil) similar to agricultural lands or woodlands and in many instances mixed with trees and other vegetation.

The class "Grassland" reveals a lower PA in the Land Cover map (PA = 69,6%) when compared to the ESA CCI map (PA = 74,8%), which is mainly due to a misclassification of "Grassland" as "Forest". The ESA CCI map also reveals this type of confusion, but with lower percentage levels of error between the two classes. However, the class "Grassland" reveals lower UA (UA = 17,9%) in the ESA CCI map, than in the In-house Land Cover map (UA = 55,7%).

When compared with the In-house Land Cover map, the major issues identified on the ESA CCI map are: (1) the presence of spatial inconsistencies, most likely related to the mosaicking process; (2) the spatial resolution which does not correspond exactly to the reported 20 m (Lesiv et al., 2017) and (3) significant classification errors, which were also reported in Lesiv et al. (2017) for the entire African continent. As for the In-house Land Cover map, the main issues are: (1) smoothing of spectral differences between classes during the compositing process (median value among the available dates for each pixel) may affect the classification results. This aspect is particularly problematic when facing extreme weather conditions. In fact, the rainfall behavior at the end of 2015 (October – December) was significantly below average in almost the entire South and Central region of Mozambique. Also, in the

second period of the 2016 wet-season (January – March), the entire Southern part of the territory, including Sofala, South Manica and North of Niassa, suffered a significant rainfall deficit. Additionally, in 2016, the Southern region of Mozambique was severely affected by drought, which resulted in losses of almost all planted crops in the main growing season (ONU, 2016). This drought in the southern regions may have had implications on the standard spectral differentiation of different features. Such implications may have introduced confusion resulting in lower accuracies in that region, (2) there can be large differences in the reflectance values between the pick and end of the growing season in regions where not enough valid observations were present to compute the median, (3) polygons in the training set may contain mixed spectral signals, and even though we did not tamper with the reference data collected independently by the MRV Unit team, avoiding mixed polygons could have improved the results, (4) the size of the training sample is small in some regions.

Lesiv et al. (2017) reported an OA of approximately 65% for the ESA CCI 20 m prototype map for Africa using two independent data sets, where the OA throughout Mozambique shows lower accuracies in the South and higher accuracies in the North, ranging between 22–65%. Here, a spatially explicit distribution of the OA of both the In-house Land Cover map and the ESA CCI map, computed through GWR, are presented based on results obtained with our validation data set. Table 2 summarizes the OA spatial variation and displays the interquartile range (IQR) obtained in both cases whereas the maps in Fig. 6 illustrate these variations.
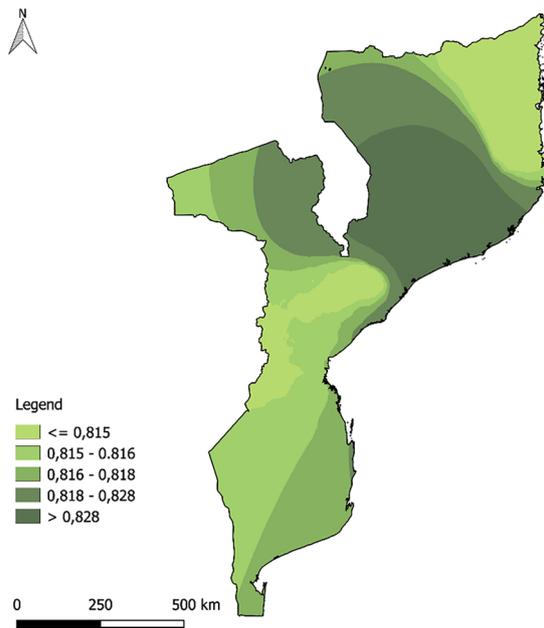
Regarding the response design, Stehman and Wickham (2011) stated that, despite polygons being more prone than pixels (or blocks of pixels) for representing actual earth surface features, it is difficult to design and implement an accuracy assessment using polygons as the spatial assessment units. An alternate response design could be obtained by implementing the sampling of one or more pixels within each validation polygon. However, Zhen et al. (2013) reported a 9% bias in the OA when a simple random selection of pixels from validation polygons, i.e. not used for training, but members of the same reference

Table 2
Summary on the variation of the overall accuracies for both maps assessed through GWR.

Maps	Min.	1stQu.	Median	3rdQu.	Max.	IQR
In-house Land Cover map	0,7985	0,8154	0,8216	0,8217	0,9103	0,0063
ESA CCI map	0,5260	0,5672	0,5710	0,5805	0,7820	0,0133

E: Level 1 – OA distribution of 20m Land Cover map

F: Level 1 – OA distribution of 20m ESA CCI map

Legend
- <= 0,815
- 0,815 - 0.816
- 0,816 - 0.818
- 0,818 - 0.828
- > 0,828

0   250   500 km

Legend
- <= 0,565
- 0,565 - 0,569
- 0,569 - 0,572
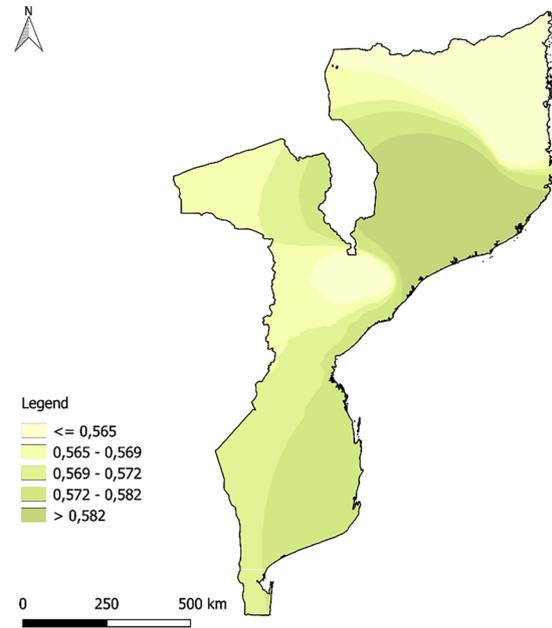- 0,572 - 0,582
- > 0,582

0   250   500 km

Fig. 6. Spatially distributed overall accuracy (OA) of both the In-house Land Cover map (E) and the ESA CCI map (F) at 1 km resolution. Values in the legend represent the probability of the presence of the correct class.

data set, were used to assess accuracy. In the present study, the maps were validated with a random sample for the available polygon data set. Yet, for the GWR implementation only the centroid (pixel value) within each validation polygon was used to derive the spatially explicit OA, so the GWR results contain bias, which can be verified by comparing the OA obtained in the error matrices to the GWR results. Nevertheless, the GWR makes it easier to compare the spatial distribution of error between the two maps and shows which are the regions that need further improvement on the reference data set. The IQR (Table 2) shows that there are greater spatial variations in the relationships between classified and validation data for the ESA CCI map.

Even though OA shows a much wider range of values, and an overall poorer performance, in the case of the ESA CCI map when compared with the In-house Land Cover map, it must be pointed out that this may be a consequence of the fact that, in the latter, both the training and validation data sets (however disjoint and randomly defined) are drawn from the same reference set. This entails, that any existing spatial bias will be mirrored in both de training and validation sets and thus discrepancies in the correspondence between the classified and observed is minimized. Thus, it cannot be excluded that the consistent spatial pattern in the distribution of accuracies is a reflection of a spatial bias in the validation data set.

3.3. Adequacy of GEE for land cover mapping and REDD+

Activity data for land cover changes, like deforestation, are mostly derived using satellite-based techniques. However, this is often costly and requires significant expertise, which can be impeditive for developing countries. The availability of a free cloud computing platform, which stores petabytes of satellite data that can be integrated with tools like Collect Earth and which provides easy to implement pre-prepared scripts, is a stepping stone for developing countries to demonstrate compliance with international agreements on climate change. Countries

are required to establish MRV systems aligned with existing NFMS and provide national and annually estimates of changes in forest carbon stocks and emissions following specific terms of quality assurance and quality control. VCS JNR requires a minimum overall accuracy of 75% for the F/NF maps as well as an uncertainty analysis of the Land Use Land Cover Changes (LULCC) map (AD) (Olofsson et al., 2014). According to FCPF CF MF, the uncertainty associated with the AD should be propagated to estimate the uncertainty of emission reductions using Monte Carlo methods. Additionally, according to both VCS JNR and FCPF CF MF terms, carbon changes associated with forest degradation should also be reported. In this study, forest degradation is not addressed, although Sentinel-1 data available in GEE could represent a valuable asset to explore transitions within the forest classes when there is a loss of carbon sequestration.

Midekisa et al. (2017) presented an approach to process LULCC maps, over a 15-year time period for Africa, using GEE and Landsat imagery along with NDVI and Normalized Difference Water Index (NDWI), showing once again the capabilities of GEE for dealing with big earth observation data. However, the training and validation data were only available for a single year and the pixel median within a three-year window was used to build the annual composites. Also, the total amount of training (5 664 pixels) and validation (1 420 pixels) data is considerably lower when compared to the amount of training data used to derive the In-house Land Cover map and to validate all maps in this study. Here, we demonstrate how to produce a cloud/shadow-free mosaic and a single land cover map for 2016 that fulfils REDD+ requirements, through the GEE cloud platform. The same exact procedure can be applied for upcoming years, by creating new mosaics and updating the reference data set, to generate new F/N and LULCC maps with adequate accuracy levels, indicating that the effort should therefore be employed in obtaining good quality and representative reference data.

Nevertheless, some limitations of GEE were found throughout the

processing within this scale and amount of data. Firstly, the maximum number of training polygons possible (7 113 polygons) and the maximum number of trees for the random forest classifiers (30) were limiting for our case study. If both the training sample and the number of trees used in the random forest classifier could have been increased, it is very likely that the classifier would have performed better, thus increasing the classification accuracy. Secondly, operations that are area influenced by arbitrary distant inputs, such as clustering algorithms, perform poorly in GEE (Gorelick et al., 2017) and this is an important bottle-neck. This is the reason why it was not possible to apply the MMU to the classifications in GEE and parallel software had to be employed, with a reduction of efficacy. To overcome this type of problems when dealing with large extents, the integration of open source geospatial analytic tools (e.g R and QGIS) could facilitate the implementation of such operations. Computational limitations were also found when creating an error matrix using the amount of validation polygons available (20 837 polygons). Despite not being possible to obtain a one-step error matrix, it was possible to obtain an error matrix for each province of Mozambique, and then aggregate each single error matrix into a final one. We concluded that, when dealing with this amount of data, some processes can be performed outside GEE, whenever it is more efficient. Furthermore, the system is entirely responsible for deciding how to run a computation, so the user is unable to access or influence how the system handles computations, which represents some challenges (Gorelick et al., 2017). Nevertheless, Gorelick et al., (2017) addressed that research and experiments are ongoing in order to overcome some of the GEE limitations.

Despite the restrictions, GEE proved to be a reliable and robust asset for achieving considerably good results which can support developing countries in rapidly producing low cost, but valid, activity data. Thus, given the currently established calendar for the Green Climate Fund (GCF) pilot program, countries that did not yet submit their Forest Reference Emission Levels (FREL) together with their National Communication or Biennial Update Reports, could eventually still become ready to access the GCF program and obtain performance payments in the 2020–2022 time window.

Once complete processing chains are developed and established (possibly including high quality pre-processing of dense image time series), allowing the user to manipulate only simple inputs, such as geographic extent, time boundaries, reference data, and output definition, one can envision the possibility of expedite production of high quality maps for land cover monitoring; be it for REDD+, sustainable management in the AFOLU sector, or other purposes. Also, the freely available ESA CCI prototype map at 20 m resolution delivers satisfactory results at the F/NF level for Mozambique. Depending on the frequency with which such maps are made available, they could provide good alternatives to the production of in-house maps and activity data, at almost no cost.

4. Conclusions

Producing land cover maps in compliance with the established accuracy levels for countries as large and complex as Mozambique is challenging and has often led to large investments and poor results.

Annex A

Here, we illustrate how to use the GEE cloud platform and free imagery can deliver high resolution and wall-to-wall land cover maps for Mozambique, which are fully compliant with the national and REDD+ requisites. Additionally, we investigate the extent to which a freely available map - the 20 m resolution 2016 ESA CCI prototype map - can also comply with those requisites.

This study is a demonstration that, despite some current limitations, a cloud computing solution such as GEE, operating over freely available satellite image repositories, can directly overcome an identified technical and technological bottle neck in REDD+. Countries can now rely extensively on cloud and open source satellite image data bases to deliver national and annual estimates of forest area and their change, according to the requirements of REDD+/MRV systems.

Even though this study only presents one case for producing land cover maps, very similar procedures could be employed for mapping forest change if adequate reference data sets were available. However, we conclude that given the possibilities opened by cloud computing and big data analysis, which can take many technological and technical difficulties away from operational production, more effort should be employed in the collection of highly representative and low uncertainty reference data sets. Moreover, high speed internet access and continued support to the implementation of the correct technical procedures at several levels can accelerate access to performance-based payments for development.

Given these conclusions, it is apparent that investing in the long-term technical coaching of key institutions and staff in developing countries may be a better strategy than providing information technology or delivering short and very specialized technical capacity for image processing. The cloud is universal and using it can be done from anywhere without requirements other than an adequate internet access. Thus, important advances can now be obtained, and major steps be taken to facilitate and accelerate the access of more developing countries to REDD+ and the GCF.

Table A1

Aggregation applied to the land cover legend established in the MRV Road Map.

Level 0 – F/NF		Level 1 - IPCC		Level 2 - National			ESA CCI map	
Code	Class	Code	Class	MRV label	Code	Class	Code	Class
2	Non-forest	1	Cropland	**1TCF**	**11**	Tree crops	**4**	Cropland
				1FC	**12**	Field crops		
				1CXF	**13**	Shifting cultivation with open to closed forested areas		
1	Forest	2	Forest Land	**1TCW**	**21**	Forest Plantation	**1**	Trees cover areas
				2FXC	**22**	Forest with shifting cultivation		
				2FE	**23**	Broadleaved (Semi-) evergreen closed forest		
				2FD	**24**	Broadleaved (Semi-) deciduous closed forest		
				2WE	**25**	Broadleaved (Semi-) evergreen open forest		
				2WD	**26**	Broadleaved (Semi-) deciduous open forest		
2	Non-forest	3	Grassland	**2GL**	**31**	Grasslands	**3;6**	Grasslands; Lichen Mosses/Sparse vegetation
				2T	**32**	Thicket	–	–
				2S	**33**	Shrubland	**2**	Shrubs cover areas
2	Non-forest	4	Wetlands	**4SF**	**41**	Aquatic or regularly flooded shrublands	**5**	Vegetation aquatic or regularly flooded
				4HF	**42**	Aquatic or regularly flooded herbaceous vegetation		
				7WB	**43**	Artificial waterbodies	**10**	Open water
				8WB	**44**	Natural water bodies		
				17	**45**	Salt Lake		
2	Non-forest	5	Settlements	**5**	**51**	Settlements	**8**	Built up areas
2	Non-forest	6	Other Land	**6BS**	**61**	Bare soils	**7**	Bare areas
				6BR	**62**	Bare rocks		
				6SS	**63**	Dunes		

Table A2

Level 1 accuracy assessment for both maps: the error matrices in terms of estimated area proportions, overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA), mapped area and estimated area. The Margins of error (ME) were calculated for a 95% confidence interval.

A: Level 1 - In-house Land Cover map (20 m)

OA (%)	ME (%)	# Reference							
70,7	± 0,7	Cropland	Forest	Grassland	Wetlands	Settlements	Other Land	Total	UA (%)
# Mapped	Cropland	**0,16010**	0,01358	0,01936	0,01574	0,02570	0,03788	**0,27235**	58,8
	Forest	0,01346	**0,33317**	0,02034	0,01781	0,00011	0,00619	**0,39108**	85,2
	Grassland	0,01816	0,02475	**0,10980**	0,02189	0,00062	0,02176	**0,19697**	55,7
	Wetlands	0,00727	0,00495	0,00618	**0,05736**	0,00060	0,00272	**0,07908**	72,5
	Settlements	0,00001	0,00000	0,00000	0,00001	**0,00217**	0,00041	**0,00260**	83,3
	Other Land	0,00159	0,00186	0,00211	0,00192	0,00583	**0,04461**	**0,05792**	77,0
	Total	**0,20059**	**0,37831**	**0,15778**	**0,11472**	**0,03502**	**0,11358**	**1**	
PA (%)		79,8	88,1	69,6	50,0	6,2	39,3	**Total**	
Mapped Area (ha)		21 504 523	30 879 579	15 552 590	6 243 776	205 380	4 573 139	**78 958 986**	
Estimated Area (ha)		**15 838 012**	**29 870 742**	**12 458 503**	**9 058 167**	**2 765 159**	**8 968 403**		
ME (ha)		± 420 546	± 365 768	± 462 887	± 333 687	± 212 503	± 356 435		

B: Level 1 - ESA CCI map (20 m)

OA (%)	ME (%)	# Reference							
45,5	± 0,6	Cropland	Forest	Grassland	Wetlands	Settlements	Other Land	Total	UA (%)
# Mapped	Cropland	0,05774	0,01638	0,01372	0,01087	0,01382	0,02522	**0,13774**	41,9
	Forest	0,01345	0,29685	0,01625	0,01588	0,00014	0,00181	**0,34438**	86,2
	Grassland	0,10980	0,05305	0,08943	0,11703	0,01616	0,11326	**0,49873**	17,9
	Wetlands	0,00027	0,00137	0,00010	0,00960	0,00017	0,00579	**0,01731**	55,5
	Settlements	0,000004	0,000004	0	0,00001	0,00121	0,00011	**0,00133**	90,4
	Other Land	0	0	0,00001	0,00008	0,00001	0,00040	**0,00050**	80,0
	Total	**0,181261**	**0,367658**	**0,119511**	**0,153467**	**0,031505**	**0,146598**	**1**	
PA (%)		31,9	80,7	74,8	55,5	3,8	0,3	**Total**	
Mapped Area (ha)		10 876 056	27 191 726	39 379 535	1 366 505	105 317	39 847	**78 958 986**	
Estimated Area (ha)		**14 312 200**	**29 029 896**	**9 436 438**	**12 117 567**	**2 487 621**	**11 575 265**		
ME (ha)		± 448 213	± 367 090	± 411 598	± 429 935	± 205 560	± 427 047		

Table A3
Level 0 accuracy assessment for both maps: the error matrices in terms of estimated area proportions, overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA), mapped area and estimated area. The Margins of error (ME) were calculated for a 95% confidence interval.

C: Level 0 - In-house Land Cover map (20 m)

OA (%)	ME (%)	# Reference			
90,5	± 0,4	Forest	Non-Forest	Total	UA (%)
# Mapped	Forest	0,333	0,058	**0.391**	85,2
	Non-Forest	0,037	0,572	**0.609**	94,0
	Total	0,370	0,630	**1**	
PA (%)		90,1	90,8	**Total**	
Mapped Area (ha)		30 879 579	48 079 407	**78 958 986**	
Estimated Area (ha)		29 201 677	49 757 310		
ME (ha)		± 304 737	± 304 737		

D: Level 0 - ESA CCI map (20 m)

OA (%)	ME (%)	# Reference			
88,5	± 0,4	Forest	Non-Forest	Total	UA (%)
# Mapped	Forest	0,297	0,048	**0,344**	86,2
	Non-Forest	0,068	0,588	**0,656**	89,6
	Total	**0,365**	**0,635**	**1**	
PA (%)		81,4	92,5	**Total**	
Mapped Area (ha)		27 191 726	51 767 260	**78 958 986**	
Estimated Area (ha)		**28 804 815**	**50 154 171**		
ME (ha)		± 349 995	± 349 995		

References

Artur, L., Hilhorst, D., 2012. Everyday realities of climate change adaptation in Mozambique. Glob. Environ. Change 22, 529–536.

Belgiu, M., Drăguț, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.

Cabral, A., De Vasconcelos, M.J.P., Pereira, J.M., Bartholomé, É., Mayaux, P., 2003. Multi-temporal compositing approaches for SPOT-4 VEGETATION. Int. J. Remote Sens. 24, 3343–3350.

CCI Land Cover team, 2016. CCI Land Cover - S2 Prototype Land Cover 20m Map of Africa 2016. http://2016africalandcover20m.esrin.esa.int/download.php.

CIA, 2018. The World Factbook. https://www.cia.gov/library/publications/the-world-factbook/geos/mz.html.

Comber, A., Fisher, P., Brunsdon, C., Khmag, A., 2012. Spatial analysis of remote sensing image classification accuracy. Remote Sens. Environ. 127, 237–246.

Dong, J., Xiao, X., Menarguez, M.A., Zhang, G., Qin, Y., Thau, D., Biradar, C., Moore III, B., 2016. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. Remote Sens. Environ. 185, 142–154.

DSU, 2015. Climate Change Profile MOZAMBIQUE. Netherlands Commission for Environmental Assessment.

Eva, H.D., Belward, A.S., De Miranda, E.E., Di Bella, C.M., Gond, V., Huber, O., Jones, S., Sgrenzaroli, M., Fritz, S., 2004. A land cover map of South America. Glob. Change Biol. Bioenergy 10, 731–744.

FCPF, 2016. Carbon Fund Methodological Framework. https://www.forestcarbonpartnership.org/carbon-fund-methodological-framework.

GADM, 2015. Global Administrative Areas Database Version 2.5. https://gadm.org/.

GFDRR, 2011. Vulnerability, Risk Reduction, and Adaptation to Climate Change Mozambique. The World Bank Group.

GOFC-GOLD, 2009. A Sourcebook of Methods and Procedures for Monitoring and Reporting Anthropogenic Greenhouse Gas Emissions and Removals Associated With Deforestation, Gains and Losses of Carbon Stocks in Forests Remaining Forests, and Forestation. GOFC-GOLD Project Office, Natural Resources, Alberta, Canada.

Google Earth Engine Team, 2015. Google Earth Engine: A Planetary-scale Geospatial Analysis Platform. https://earthengine.google.com/.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Harris, N.L., Brown, S., Hagen, S.C., Saatchi, S.S., Petrova, S., Salas, W., Hansen, M.C., Potapov, P.V., Lotsch, A., 2012. Baseline map of carbon emissions from deforestation in tropical regions. Science 336, 1573–1576.

Huang, H., Chen, Y., Clinton, N., Wang, J., Wang, X., Liu, C., Gong, P., Yang, J., Bai, Y., Zheng, Y., 2017. Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. Remote Sens. Environ. 202, 166–176.

Johansen, K., Phinn, S., Taylor, M., 2015. Mapping woody vegetation clearing in Queensland, Australia from Landsat imagery using the Google Earth Engine. Remote Sens. Appl. Soc. Environ. 1, 36–49.

Lesiv, M., Fritz, S., McCallum, I., Tsendbazar, N., Herold, M., Pekel, J.-F., Buchhorn, M., Smets, B., Van De Kerchove, R., 2017. Evaluation of ESA CCI prototype land cover map at 20m. Working Paper WP-17-021. Int. Inst. Appl. Syst. Anal.

Mayaux, P., Richards, T., Janodet, E., 1999. A vegetation map of Central Africa derived from satellite imagery. J. Biogeogr. 26, 353–366.

Midekisa, A., Holl, F., Savory, D.J., Andrade-Pacheco, R., Gething, P.W., Bennett, A., Sturrock, H.J., 2017. Mapping land cover change over continental Africa using Landsat and Google Earth Engine cloud computing. PLoS One 12, e0184926.

MITADER, 2016. MRV Road Map Moçambique. http://www.redd.org.mz/uploads/SaibaMais/ConsultasPublicas/MRV%20Road.pdf.

MITADER, 2017. R-Package Multi-stakeholder Self-Assessment of REDD+ Readiness in Mozambique. Fundo Nac. Desenvolv. Sustentável.

NASA LP DAAC, 2015. Nadir BRDF-Adjusted Reflectance 16-Day L3 Global 500m. Version 5.

Ochieng, R.M., Visseren-Hamakers, I.J., Arts, B., Brockhaus, M., Herold, M., 2016. Institutional effectiveness of REDD+ MRV: countries progress in implementing technical guidelines and good governance requirements. Environ. Sci. Policy 61, 42–52.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., 2001. Terrestrial ecoregions of the world: a new map of life on earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience 51 (11), 933–938. https://www.worldwildlife.org/.

ONU, 2016. Mozambique: Drought. Office of the Resident Coordinator, Mozambique Emergency Situation Report No. 4.

Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26, 217–222.

Ramoino, F., Pera, F., Arino, O., 2016. S2 Prototype LC Map at 20m of Africa 2016 Users Feedback Compendium. Eur. Space Agency.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93–104.

Romijn, E., Herold, M., Kooistra, L., Murdiyarso, D., Verchot, L., 2012. Assessing capacities of non-Annex I countries for national forest monitoring in the context of REDD+. Environ. Sci. Policy 19, 33–48.

Schaaf, C., Wang, Z., 2015. MCD43A4 MODIS/Terra + Aqua BRDF/Albedo Nadir BRDF Adjusted Ref Daily L3 Global - 500m V006 [Data set]. NASA EOSDIS Land Processes DAAC.

Simonetti, D., Simonetti, E., Szantoi, Z., Lupi, A., Eva, H.D., 2015. First results from the phenology-based synthesis classifier using Landsat 8 imagery. IEEE Geosci. Remote Sens. Lett. 12, 1496–1500.

Sitoe, A., Salomão, A., Wertz-Kanounnikoff, S., 2012. The Context of REDD+ in Mozambique: Drivers, Agents and Institutions. Occasional Paper 79. CIFOR, Bogor, Indonesia.

Stehman, S.V., Wickham, J.D., 2011. Pixels, blocks of pixels, and polygons: choosing a spatial unit for thematic accuracy assessment. Remote Sens. Environ. 115, 3044–3055.

Stibig, H.J., Beuchle, R., Achard, F., 2003. Mapping of the tropical forest cover of insular Southeast Asia from SPOT4-Vegetation images. Int. J. Remote Sens. 24, 3651–3662.

Van der Werf, G.R., Morton, D.C., DeFries, R.S., Olivier, J.G., Kasibhatla, P.S., Jackson,

R.B., Collatz, G.J., Randerson, J.T., 2009. CO2 emissions from forest loss. Nat. Geosci. 2, 737.

VCS, 2017. Jurisdictional and Nested REDD+ (JNR). http://www.v-c-s.org/project/jurisdictional-and-nested-redd-framework/.

World Bank, 2017. Mozambique. https://data.worldbank.org/country/mozambique.

Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnelt, J., Congalton, R.G., Yadav, K., Thau, D., 2017. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. ISPRS J. Photogramm. Remote Sens. 126, 225–244.

Zhen, Z., Quackenbush, L.J., Stehman, S.V., Zhang, L., 2013. Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification. Int. J. Remote Sens. 34, 6914–6930.