Forestry Department
Food and Agriculture Organization
of the United Nations

STANDARD OPERATING PROCEDURES FOR AREA ESTIMATION

Purpose: These Standard Operating Procedures (SPO) Notes acts as a set of general Cheat-sheets, which in conjunction with a more general posterior white paper, aim to be the main support backbone documents for national staff working in area reporting for REDD+. The SPO will be modified to fit the particular protocols of each country and ensure that they are repeatable.

Important Note: Once project documentation has been handed over and signed off by both parties, any further revisions will be the responsibility of Operations and / or Application Services. These revisions will be included in Annex Documentation to ensure the original documentation remains unchanged and as delivered.

Application Name	
Application Portfolio number	
Project Name	Zambézia Integrated Landscapes Management Program – Emissions Reduction Program (ZILMP)/Mozambique
Project Number	<< DTC Number >>
Project Manager	
	(Print name) (signature) (date)
Project Description	Designed at jurisdictional scale, the ZILMP is located in Zambézia province, of which it covers 9 districts: Alto Molócue, Gilé, Gúruè, Ile, Maganja da Costa, Mocuba, Mocubela, Mulevala and Pebane. Its ambition is to reduce emissions due to deforestation in the accounting area.

IMPORTANT NOTES FOR COMPLETING THIS DOCUMENT

Under most headings in this template are instructions for completing each section; they are marked in this way: << instructions for completing the section >>. Please read through the instructions for details on minimum requirements.

Important note: No sections are to be deleted from this document.

Text contained within << >> provides information on how to complete that section and can be deleted once the section has been completed.

TABLE OF CONTENTS

STANDARD OPERATING PROCEDURE 0 (SOP0): MAP PRODUCTION **5**
- SECTION 1 OVERVIEW 5
- SECTION 2 PROCEDURE 6
- SUMMARY OF APPROACH 6
- IMAGE SELECTION 6
- IMAGE PREPROCESSING 8
- CLASS DEFINITIONS 8
- REFERENCE IMAGERY 8
- TRAINING AREAS 9
- FEATURE GENERATION 11
- FEATURE SELECTION 13
- TRAINING A CLASSIFIER 13
- OPTIMISING THE CLASSIFIER 14
- OUTPUTS 14
- SPATIAL FILTERING 16
- BUFFERING 17
- GAP FILLING 17
- MANUAL EDITING 17
- CLASSIFIED OUTPUT 18
- SECTION 3 QUALITY MANAGEMENT 19
- QUANTITATIVE ASSESSMENT 22
- VISUAL ASSESSMENT 23
- DATA FLOW DIAGRAM 24
- REFERENCES 24

STANDARD OPERATING PROCEDURE 1 (SOP1): SPATIALLY REFERENCED SAMPLE DESIGNS FOR AREA ESTIMATION **26**
- SECTION 1 OVERVIEW 26
- SECTION 2 PROCEDURE 26
- SECTION 3 QUALITY MANAGEMENT 29

STANDARD OPERATING PROCEDURE 2 (SOP2): RESPONSE DESIGN **32**
- SECTION 1 OVERVIEW 32
- SECTION 2 PROCEDURE 32

STANDARD OPERATING PROCEDURE 3 (SOP 3): DATA COLLECTION 48
- SECTION 1 OVERVIEW 48
- SECTION 2 PROCEDURE 48
- SECTION 3 QUALITY MANAGEMENT 51

STANDARD OPERATING PROCEDURE 4 (SOP4): SAMPLE-BASED AREA ESTIMATION ANALYSIS 57
- SECTION 1 OVERVIEW 57
- SECTION 2 PROCEDURE 57
- SECTION 3 QUALITY MANAGEMENT 60

SOP 0: MAP PRODUCTION

STANDARD OPERATING PROCEDURE 0 (SOP0): MAP PRODUCTION

Section 1 Overview	
Purpose	Produce a map of forest cover change for each year of the monitoring period to support stratified random sampling (SOP1).
Scope	Wall-to-wall mapping of the project area of forest cover change for each year of monitoring.
Responsibilities	• Data collection manager • Technicians in charge of the visual image interpretation for collecting training data used for map production, referred to as operator.
Prerequisites	• Region or area or interest identified. • Land and/or land use change definitions, applicable to the area of interest. • Mapping may not begin before May 1st of the current year
Requirements	• Computer hardware system ○ Windows 7, Windows 10 or later ○ 2.5 GHz or faster processor ○ 4 or more gigabytes (GB) of RAM • Software ○ Google Chrome or Mozilla Firefox • Stable internet connection • Human resources ○ Data collection manager ○ Team of specialized and trained image interpreters • Access to the FNDS Google Earth Engine mapping interface. • Auxiliary data (i.e. high resolution imagery, historical land use and land cover maps) **Important Note:** The mapping approach described here will develop over time, given experience and availability of new data.

Section 2 Procedure	
Overview	**Summary of approach** The method outlined here is for production of deforestation maps to support robust provincial-scale estimates of deforested area in Mozambique on an annual basis. The mapping approach is based on a supervised classification of annual composite images and time-series metrics from Sentinel-2 for two periods separated by one year (the 'reference' and 'current' period). The resulting map is post-processed to optimise for stratified random sampling to generate estimates of deforested area. The method is implemented on Google Earth Engine, an integrated cloud processing platform for computation and visualisation of large volumes of satellite data. The workflow used to produce estimates of annual deforestation for ZILMP is summarised as follows: i. Build Sentinel-2 image composites for the monitoring area using images from the wet season (January - May) for the reference period and current periods. ii. Collect training data through visual interpretation, identifying 3 classes, namely, stable forest, stable non-forest and deforestation. The interpretation is based on comparing the pair of composite images for the reference and current periods. In addition, an NDVI change image between both periods is used to direct the operator towards locations of potential forest cover change. iii. Generate a set of image features from wet season imagery which are used to identify changes in forest cover, including composite images, changes in vegetation indices (e.g. NDVI, EVI), the phenological state of surrounding forest, and measures of multi-temporal variation in NDVI to highlight locations of forest cover change. iv. Combine the training data with feature images to calibrate a Random Forest classifier, used to produce a map of deforestation. v. Post-processing the map of deforestation, creating multiple deforestation classes using buffers and a filter for spatial area to capture as much deforestation as possible in a deforestation class to minimise errors of omission. This document describes each of these processing steps in detail.
Input selection	**Image selection** Images are drawn from the Sentinel-2 archive from the wet season (defined as January 1st – May 1st) for each of the 'reference' and 'current' years. Cloudy images that are unlikely to contain usable data are removed by excluding images where metadata reports a high cloud cover (>70 %). For monitoring to be possible, there must be a minimum of one cloud free observation for each pixel in the study area each year, which in some cases may require the search period to be extended into the early wet season (i.e. November – December) or early dry season (i.e. May - June) as required. Where additional images are required, the operator should be mindful of the dynamics of vegetation seasonality and deforestation in the area of interest:

- In images drawn from the dry season, contrast between forest and non-forest vegetation is low in many vegetation types (e.g. open savanna), and sharp changes in reflectance from frequent fires can be confused with deforestation.
- Forest cover changes associated with small-scale agriculture are the most common activity associated with forest cover change in Mozambique, with clearance activities most confined to the late dry season. Images included the late dry may result in missed or mis-allocated change between monitoring periods.

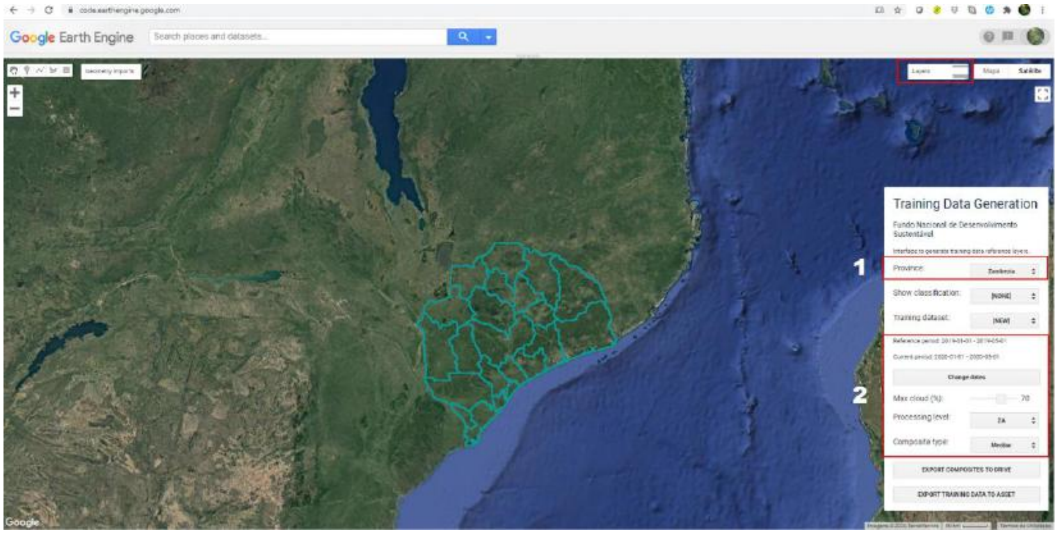
In cases where available images are still insufficient (not yet encountered in Mozambique), Sentinel-2 images may be supplemented by images from Landsat 8.

Images are processed at 20 m resolution throughout.

The images selection is performed in the FNDS mapping interfaces, specifically the section "Training Data Generation", as shown in Figure 1. For a selection of images, the operator should:

- Run the script provided by the section;
- Load the boundary polygon for the area of interest (AOI) on the "Province" menu;
- Define the date ranges from which to draw images for "reference" and "current" periods on the "Change dates" menu;
- Fix the maximum cloud cover to include input images on the "Max cloud (%)" menu;
- Select the desired Sentinel-2 processing level on the "Processing level" menu; and
- Select the method to produce the composite images on the "Composite type" menu.

The composite images are automatically listed on the Layer Manager.



Training Data Generation

Fundo Nacional de Desenvolvimento Sustentável

Province: Zambezia

Show classification: [NONE]

Training dataset: [NEW]

Max cloud (%): 70

Processing level: 2A

Composite type: Median

Figure 1 Selection of images on the section " Training Data Generation" from the FNDS mapping interfaces on the Google Earth Engine Cloud Computing Platform.

Image preprocessing

Either Sentinel-2 level 1C (top-of-atmosphere) or level 2A (surface reflectance) may be used, with a preference of level 2A where available given its radiometric consistency and improved cloud mask.

Cloud and cloud shadow are masked in each image, based on QA bands distributed with Sentinel-2 images and an additional simple threshold-based mask to capture residual cloud cover.

In some cases, there may remain a small number of pixels without any valid observations. Where these locations are associated with persistent errors in cloud masks where the land surface is bright in cloud-free images (sand or bare land), they can be ignored. Where there is a possibility of missed deforestation due to persistent cloud cover, additional images should be included.

Production of training data

To satisfy the requirement for a high-accuracy classification in a region with heterogeneous vegetation types subject to substantial inter- and intra- annual variation, a separate classification is performed for each area of interest in each year using regionally specific training data.

Class definitions

The classifier for forest cover change is calibrated using operator-identified locations of stable forest, stable non-forest and deforestation ('training data'). The three training data classes are defined in Table 1. For the purposes of classification forest cover is defined at the pixel-scale, with no minimum area requirement.

Table 1 Training data class definitions.

Label	Class	Definition
Stable forest	1	Forest cover in reference year that remains forest in current year.
Deforestation	2	Forest cover in the reference year that through land use change is no longer forest in the current year.
Stable non-forest	3	Non-forest cover in the reference year that remains non-forest in the current year.

Reference imagery

Training data production is conducted either online or offline, using a pair of composite map products for the 'reference' and 'current' years. These composites may be generated using any applicable method to obtain a cloud-free image (median composite, maximum NDVI composite etc.), with the requirement that the images forming the composite are derived from the same image inputs as will be used in the classification stage.

Operators may be further guided by a map of NDVI change, which highlights locations of likely deforestation.

Auxiliary data such as high-resolution basemaps (e.g. Google satellite, Bing maps) and historical land use and land cover maps may be used to assist in visual image

<table>
<tr>
<td></td>
<td>
interpretation, noting that these images may not be collected from the same time period as the reference images.

Example reference data for training data production are shown in Figure *2*.

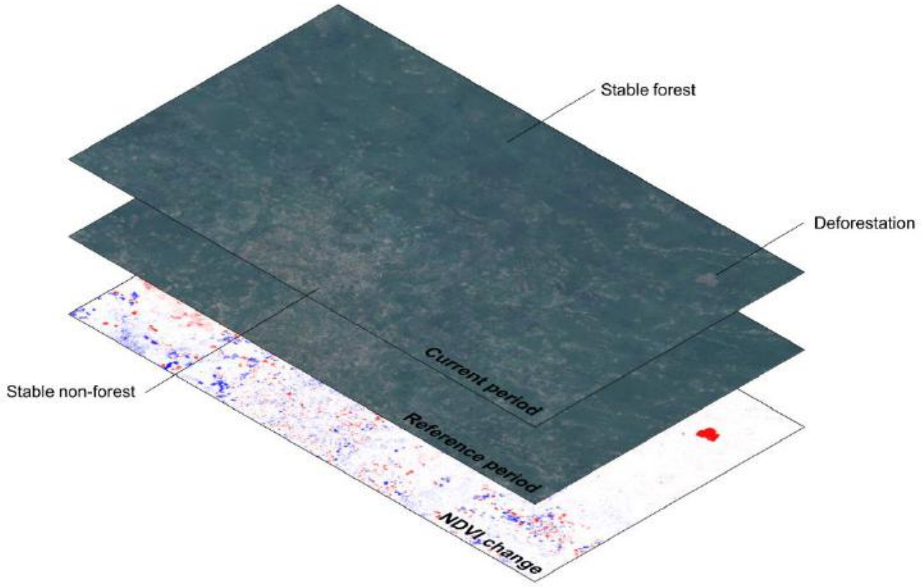


Figure 2 Reference imagery is used for production of training data delineated locations of stable forest, stable non-forest, and deforestation.

Training areas

Training data consists of polygons delineating each of the map classes (forest, non-forest, deforestation). Polygons are preferred to points for efficiency in collection of training data.

A typical training dataset should consist of in the order of 100 polygons for each of the three classes spread over the area of interest, with a total of ~20,000 pixels split across the three classes. Part of an example training dataset is shown in Figure *3*.
</td>
</tr>
</table>

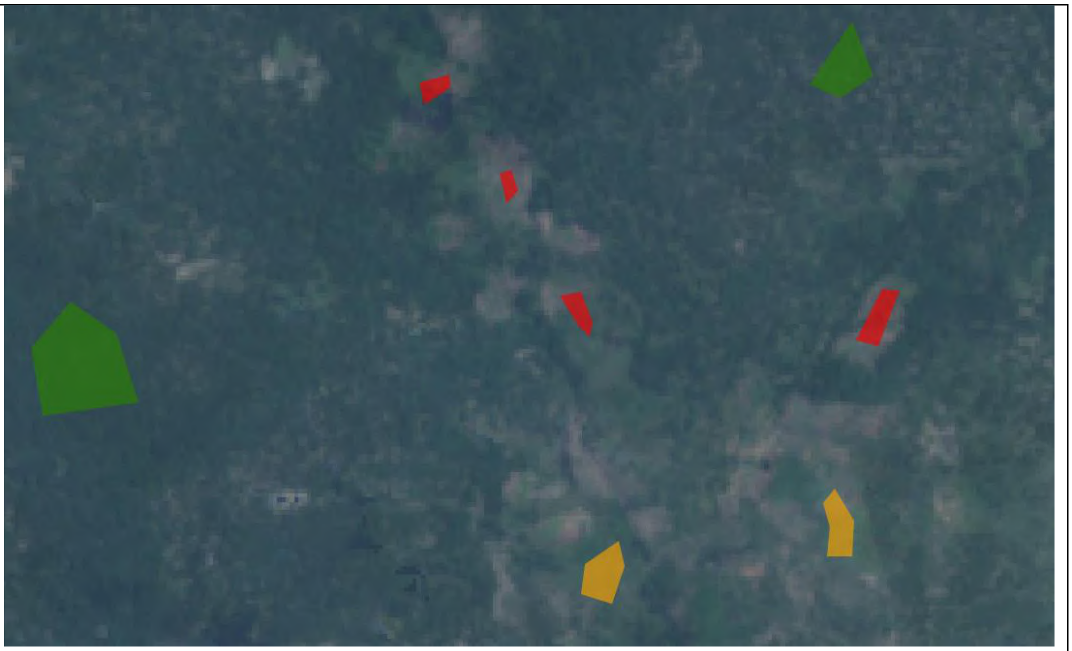


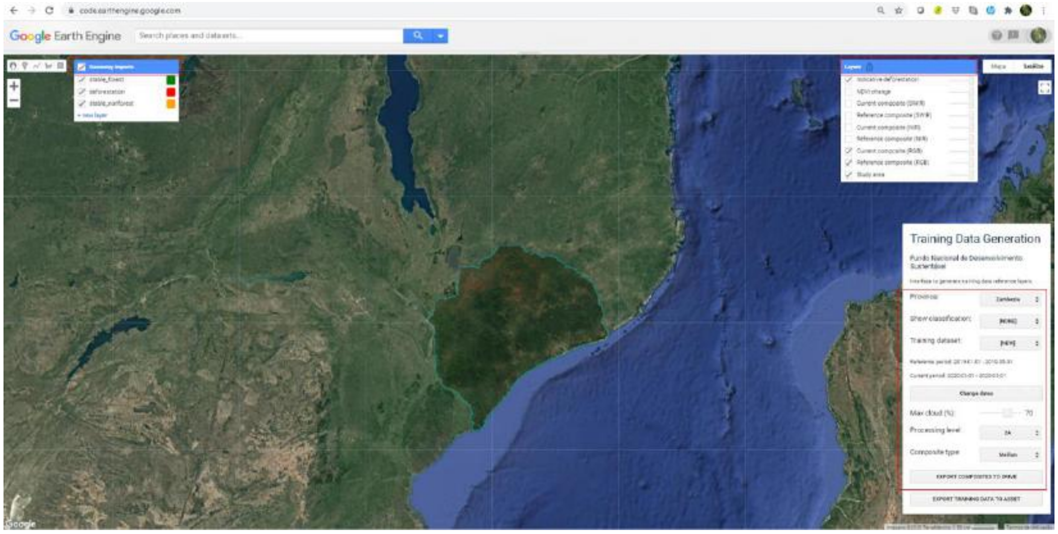
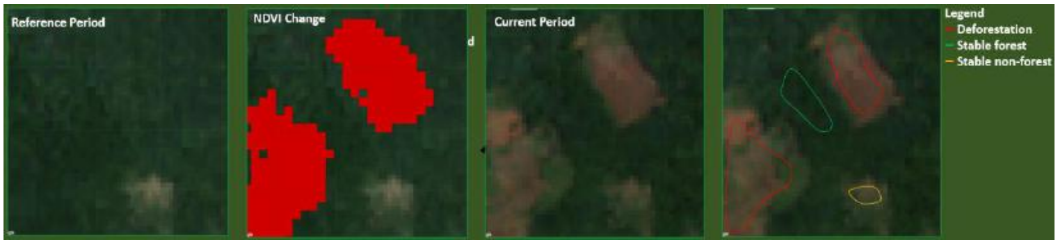
Figure 3 Example training data from Zambezia province, delineating stable forest (green), stable non-forest (orange), and deforestation (red).

See Section 3 Quality management for a detailed description of how training data should be collected.

Training datasets may be later supplemented with further training data following initial classifications as required.

The training data collection is performed in the FNDS Google Earth Engine mapping interface, specifically the section "Training Data Generation", as shown in Figure *4*. For generating an image features, the operator should:

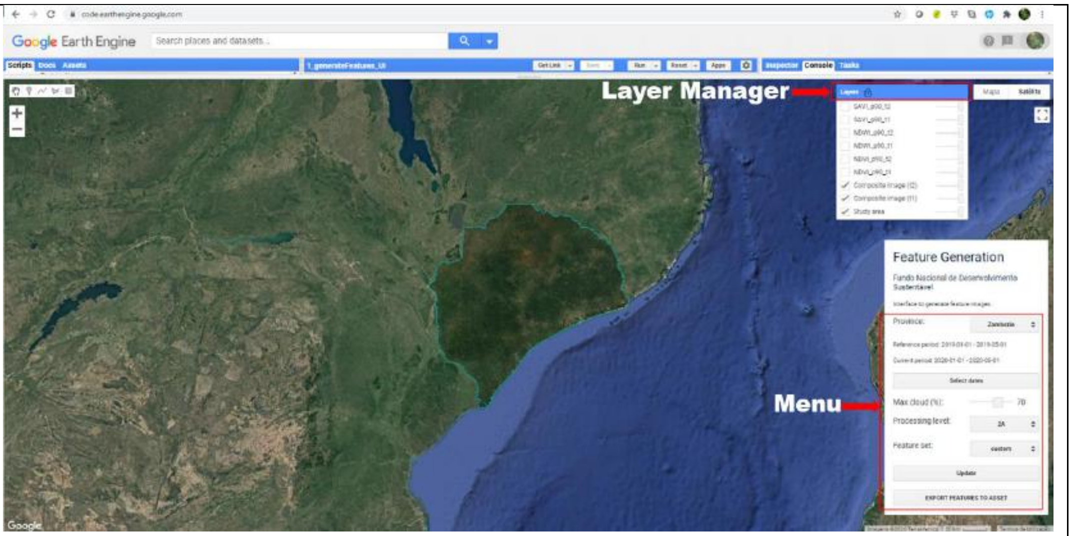
- Run the script provided by the section;
- Load the boundary polygon for the area of interest (AOI) on the "Province" menu;
- Set the date ranges defined at the subsection "Input selection" on the "Changes dates" menu;
- Fix the maximum cloud cover to include input images on the "Max cloud (%)" menu;
- Select the desired Sentinel-2 processing level on the "Processing level" menu;
- Select feature inputs on the "Feature set" menu;
- Select the method to produce the composite images on the "Composite type" menu;
- Select a previously classified map for the AOI on the "Show classification" menu, if available for the respective date ranges defined on the "Changes dates" menu, and necessary to identify and correct their classification errors by adding additional training data to improve the map;
- Create a new training dataset or edit an existing dataset for the respective date ranges defined on the "Training dataset" menu; and
- Click on the button "EXPORT COMPOSITES TO DRIVE" to save the composite images, for offline data generation only if necessary.

	The Layer Manager lists automatically the composite images and their respective NDVI change image, allowing operator to perform the comparison between the both periods as shown in Figure *5*. The Geometry imports lists automatically the training data classes and their respective number of polygons delineated on screen.  Figure 4 The section "Training Data Generation" from the FNDS mapping interfaces on the Google Earth Engine Cloud Computing Platform.  Figure 5 Training data collection
Generation of image features	**Feature generation** Image stacks for the 'reference' and 'current' years are reduced to a series of metrics (or features) that together describe the state and temporal variability of vegetation cover in each year, and highlight differences between years. Feature inputs are formulated to be robust to inter-annual variations in the extent and timing of green-up, heterogeneity of forest cover properties, and confusion with vegetation changes associated with agricultural cycles. Available features are: • **Composite image bands:** Images from Sentinel-2 in the reference and current periods are separately combined into cloud-free composite images. Composites are based on a median of non-masked pixels. Included bands are: 2, 3, 4, 8, 11, 12.

- **Vegetation indices:** A series of vegetation indices are calculated for each input image: NDVI, EVI, SAVI, NBR, and NDWI. These are summarised into a single statistic based on the 90th percentile of non-masked pixels, a value that (i) normalises index values to the peak of the wet season, and (ii) removes residual cloud. The exception is NBR where the median is used in place of the 90th percentile, which frequently includes residual cloud.
- **Spatial normalisation:** A key barrier to reliably detecting changes in savanna and dry forest vegetation is inter-annual variation in tree and grass phenology, which complicates comparison between two time periods. Additional features for NDVI and NDVI change are normalised using the approach of Hamunyela *et al*. (2016), in addition to a related approach based on local negative outliers in NDVI change using Z-scores.
- **Time series metrics:** Step-changes to the land surface are not only associated with deforestation, but also other land use activities such as agriculture. The proportion of pixels above a threshold of NDVI (e.g. 0.6) and the coefficient of variation of NDVI through the wet season separates short-lived agricultural cycles from more stable forest cover.
- **Other features:** A series of more experimental features are available, including ‘spectral distance’ and ‘spectral angle’ between two composite images, image texture from a standard deviation filter of NDVI in each year, and the ‘hue’ of an image as a simplified representation of its colour.

The generation of image features is performed in the FNDS mapping interfaces, specifically the section "Feature Generation", as shown in Figure 6. For generating an image features, the operator should:

- Run the script provided by the section;
- Load the boundary polygon for the area of interest (AOI) on the "Province" menu;
- Set the date ranges defined at the subsection "Input selection" on the "Select dates" menu;
- Fix the maximum cloud cover to include input images on the "Max cloud (%)" menu;
- Select the desired Sentinel-2 processing level on the "Processing level" menu;
- Select feature inputs on the "Feature set" menu;
- Click on the button "Update" to display the selected features (are automatically listed on the Layer Manager); and
- Click on the button "EXPORT FEATURES TO ASSET" to save the image features.



Province: Zambezia

Max cloud (%): 70

Processing level: 2A

Feature set: custom

Figure 6 The section "Feature generation" from the FNDS mapping interfaces on the Google Earth Engine Cloud Computing Platform.

Feature selection

The optimal model for mapping deforestation should only include the minimum number of inputs. This serves two purposes: (i) optimisation of model performance by lowering the risk of over-fitting, and (ii) reduction of computational overheads from processing and storage.

The operator should select features based domain knowledge and past experience of what has been effective. Feature selection can be guided by quality statistics output from the classifier including accuracy metrics and importance values (see Section 3 Quality management). The inclusion or removal of features will likely be an iterative process. The number of features to be included is not prescribed in this protocol as it may differ between years and locations but as a general rule should be the fewest required while still maintaining a high-quality output.

Note that some input features will be strongly correlated and retaining only one of multiple correlated inputs may improve classifier performance.

Classification

Training a classifier

Feature inputs are combined with training regions to construct a classifier to map deforestation. The classifier used is Random Forest, a widely used machine learning approach, due to the following advantages for its application in remote sensing (Breiman, 2001):

- Its accuracy is good and sometimes better than other ensemble methods.
- It's relatively robust to outliers as well as noise.
- It can handle large numbers of input variables and provides useful internal estimates of what variables are important in the classification.
- Its computational complexity is faster, lighter and easier as compared to other ensemble methods.

Training pixels for the classifier are sourced from pixels drawn from each of the stable forest, stable non-forest and deforestation classes of training data. A random sample of 30% of input pixels are held back as (quasi)-independent validation data to test the predictive skill of the classifier.

Optimising the classifier

The number of trees used by the classifier should be tuned by increasing trees to the point that reported accuracy is no longer observed to increase with the addition of more trees. However, the standard number is 64.

The classifier may be further improved through feature selection and addition of further training data focussed on locations that are problematic. See Section 3 Quality management.

Outputs

The Random Forest classifier can produce two outputs: either a categorical 3 class output (Figure *7*) or a continuous probability of a pixel belonging to the deforestation class (Figure *8*). The categorical prediction is recommended where a map product is the key output, whereas the probabilistic output is reserved for stratified area estimation.

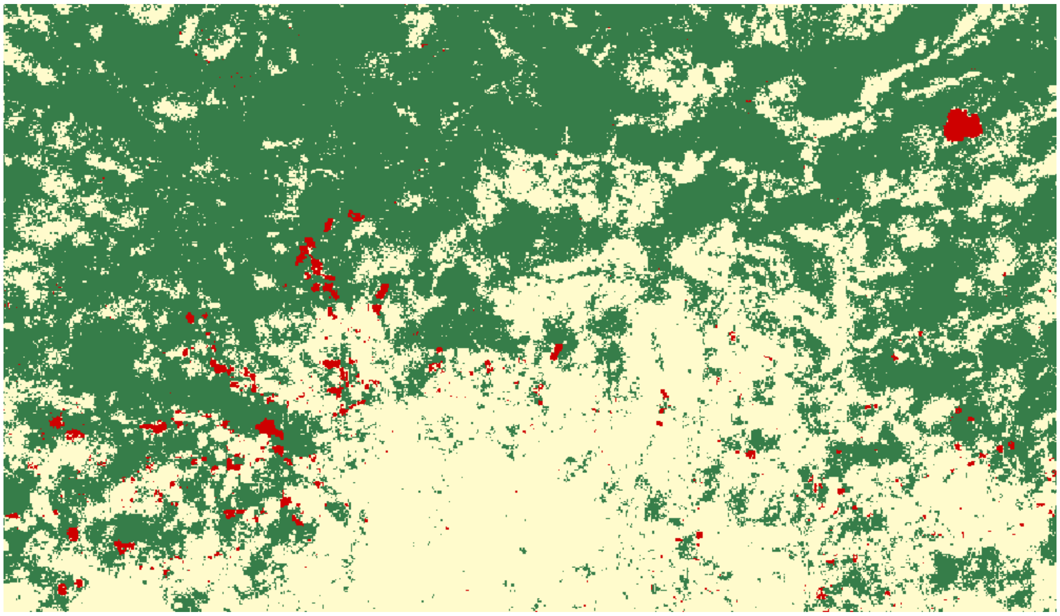


Figure 7 Example classified map output from Zambezia province, showing stable forest (green), stable non-forest (yellow) and deforestation (red).

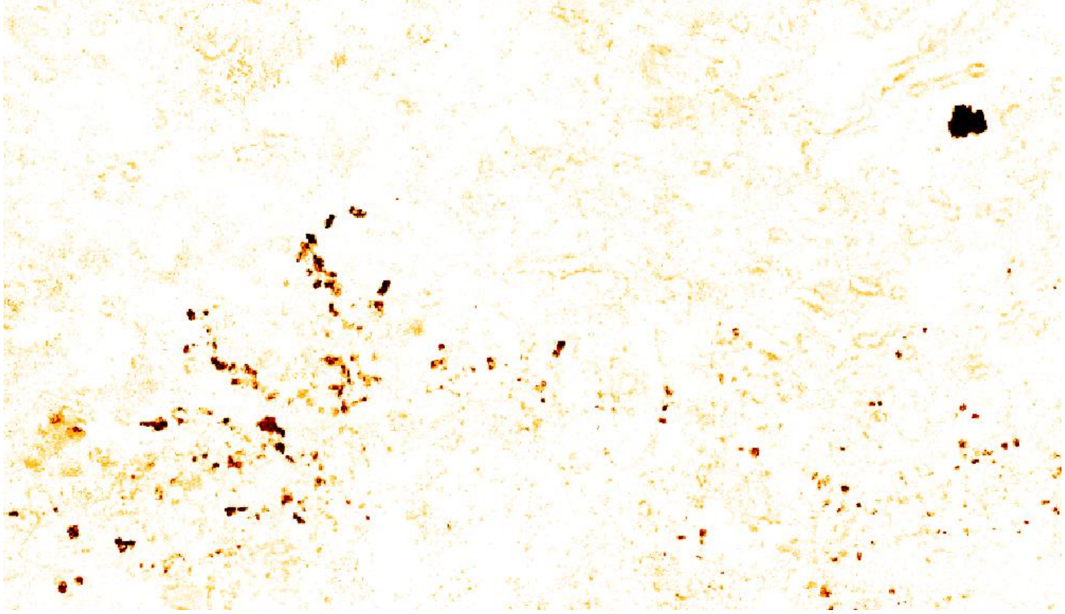


Figure 8 Example map of probability of deforestation from Zambezia province, showing areas of high probability deforestation in darker colours.

Where the classifier is set to output the probability of deforestation, the map is reduced to a binary classification of deforestation/stable using a probability threshold. This threshold takes a default value of 70%, which can be modified by the user. Lower thresholds will result in a more conservatively extensive estimate of deforested area, which may be desirable for stratified area estimation where errors of omission in the deforestation class are a concern. This map is extended to a three class map by replacing non-deforestation pixels with the class from Mozambique’s 2016 land cover map.

The categorical output is standard choice for stratified area estimation, but it can be switched to a manually tuned probability output in case where errors of omission are a serious problem.

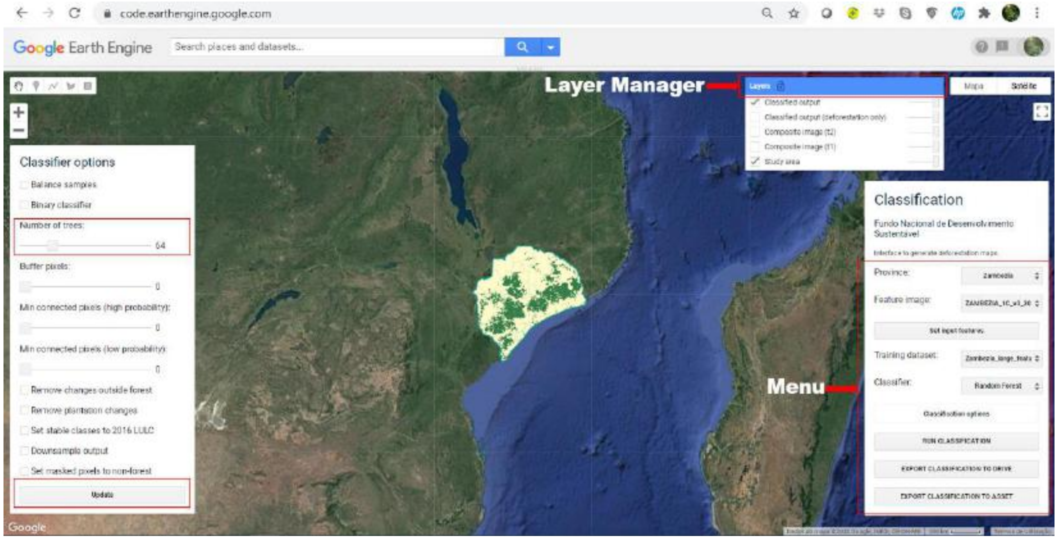
The production of map products is an iterative process, requiring repeated visual and quantitative quality checks. Where a map product is not deemed to meet quality requirements, the map should be regenerated with more training data, improved input features, or different ranges of input imagery. See Section 3 Quality management for more details.

The classification is performed in the FNDS mapping interfaces, specifically the section "Classification", as shown in Figure 9. For generating a classified image, the operator should:

- Run the script provided by the section;
- Load the boundary polygon for the area of interest (AOI) on the "Province" menu;
- Load the image features for the AOI on the "Feature image" menu;
- Click on the "Set input features" menu to remove an input feature from the image features, if necessary, by unchecking the checkbox input;

- Load the training data for the respective image features on the "Training dataset" menu;
- Select the desired classification algorithm on the "Classifier" menu;
- Click on the "Classifier options" menu;
- Set the desired number of trees at the option "Number of trees";
- Click on the button "Update" to update the classifier options;
- Click on the button "RUN CLASSIFICATION" to display the classified image; and
- Click on the button "EXPORT CLASSIFICATION TO ASSET" or "EXPORT CLASSIFICATION TO DRIVE" to save the classified image.

The Layer Manager lists automatically the classified image and their respective composite images.



Classifier options
☐ Balance samples
☐ Binary classifier
Number of trees: 64
Buffer pixels: 0
Min connected pixels (high probability): 0
Min connected pixels (low probability): 0
☐ Remove changes outside forest
☐ Remove plantation changes
☐ Set stable classes to 2016 LULC
☐ Downsample output
☐ Set masked pixels to non-forest
Update

✓ Classified output
☐ Classified output (deforestation only)
☐ Composite image (t2)
☐ Composite image (t1)
✓ Study area

Classification
Fundo Nacional de Desenvolvimento Sustentável
Interface to generate deforestation maps
Province: Zambezia
Feature image: ZAMBEZIA_10_v3_30
Set input features
Training dataset: Zambezia_large_feats
Classifier: Random Forest
Classification options
RUN CLASSIFICATION
EXPORT CLASSIFICATION TO DRIVE
EXPORT CLASSIFICATION TO ASSET

Figure 9 The section "Classification" from the FNDS mapping interfaces on the Google Earth Engine Cloud Computing Platform.

Post-processing

A number of post-classification alterations are performed on classified images to improve their accuracy and optimise their use in stratified area estimation.

Errors of omission are a particular problem for stratified area estimation with small class areas (see Olofsson *et al*. 2020). While elimination of errors of omission is not realistic, their incidence can be minimised through further stratification of the map. The following steps result in an output with multiple probability classes of deforestation, driven by the need to eliminate errors of omission of deforestation.

Spatial filtering

Deforestation tends to occur over spatially contiguous areas, such as the area of a new field or high-intensity charcoal production site. Where an isolated group of pixels of deforestation exists over a small area, it is more likely that this represents a false positive, and even if a real change it is less likely to meet monitoring criteria for deforestation (the percentage of remaining forest cover in 1 ha is lower than 30% after a disturbance).

This property is used to aid stratified area estimation by removing or downgrading the probability of isolated pixels. Pixel isolation is judged by the number of connected pixels of deforestation, with any area of deforestation of < 6 pixels (0.24 ha) reclassified as non-forest, and assign any area of 6 – 10 pixels (0.24 – 0.4 ha) allocated to the low probability deforestation class. Events of >10 pixels (0.4 ha) are retained in the high probability deforestation class.

Buffering

A common technique is to use buffers around areas of deforestation to capture missed deforestation (e.g. Olofsson *et al*. 2020). The theory of using buffers is that mis-classified deforestation events are likely to occur in the vicinity of other deforestation, because (i) poor geo-location accuracy of imagery or reference data can lead to a spatial offset between the two, which will be captured by a buffer region, and (ii) deforestation is a process that spreads spatially, thus errors of omission are highly likely to occur in the vicinity of other deforestation events.

A buffer of 40 m in all directions is applied around all high probability deforestation events. The buffer area is assigned to a separate map class, which supersedes all map classes except high-probability deforestation.

Gap filling

In some cases, no valid observations will be included from Sentinel-2 over the wet season. In most cases these will be small areas, which have been persistently mis-classified as cloud cover. This can occur on some land cover types in Sentinel-2 masks, particularly bright sand along coasts and rivers.

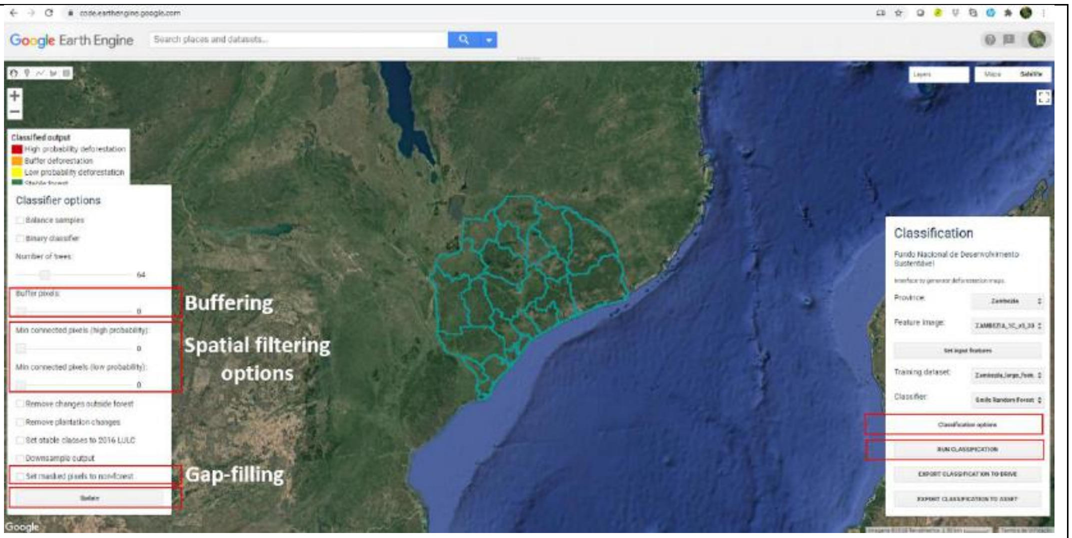
Where small gaps exist in images, provided the areal coverage is small and unlikely to hide deforestation events, these are grouped into the non-forest class.

Manual editing

Manual re-classification of clear false positives in the deforestation class is encouraged, provided there is good evidence that they do not represent true deforestation. This step must be conducted in advance and independently of stratified area estimation.

The steps described above, except the last, are performed at the section "Classification" from the FNDS mapping interfaces, as shown in Figure 10. For applying the post-processing, the operator should:

- Click on the "Classifier options" menu to see the post-processing options;
- Fix at 2 the number of pixels at the option "Buffer pixels";
- Fix at 10 the number of pixels at the option "Min connected pixels (high probability)";
- Fix at 6 the number of pixels at the option " Min connected pixels (low probability)";
- Check the checkbox at the option "set masked pixels to non-forest ";
- Click on the button "Update" to update the classifier options; and
- Click on the button "RUN CLASSICATION" to display the classified image.



Classifier options

☐ Balance samples

☐ Binary classifier

Number of trees: 64

Buffer pixels: 0

Min connected pixels (high probability): 0

Min connected pixels (low probability): 0

☐ Remove changes outside forest

☐ Remove plantation changes

☐ Set stable classes to 2016 LULC

☐ Downsample output

☐ Set masked pixels to nonforest

Classification

Fundo Nacional de Desenvolvimento Sustentável

Interface to generate deforestation maps.

Province: Zambezia

Feature image: ZAMBEZIA_SC_v1_30

Training dataset: Zambezia_large_fore...

Classifier: Smile Random Forest

Figure 10 Post-processing in the section "Classification" from the FNDS mapping interfaces on the Google Earth Engine Cloud Computing Platform.

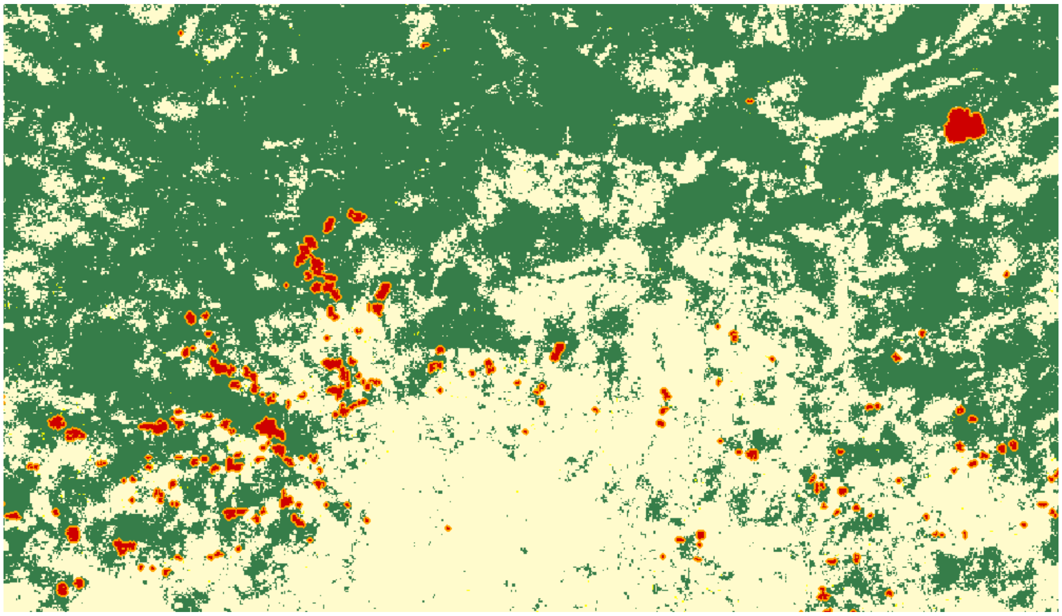
Output

Classified output

The above processing steps will result in an image optimised for stratified area estimation (see SOP1). Image classes are shown in Table 2 and an example map output in Figure 11.

Table 2 Post-processed map classes, for use in stratified area estimation.

Label	Class
No data	0
High probability deforestation	1
Buffer deforestation	2
Low probability deforestation	3
Stable forest	4
Stable non-forest	5

	 Figure 11 Example post-processed map output, showing stable forest (green), stable non-forest (light-yellow), and the three deforestation classes (high probability deforestation in red, buffer deforestation in orange, and low-probability deforestation in yellow).
Section 3 Quality management	
Input selection	For monitoring to be effective, there must be a minimum possible of the undesirable presence of the atmospheric interferences, such as clouds, hazes and shadows, etc. Image composites should be manually inspected to ensure that these impacts are minimised, and additional images included where necessary. The atmospheric interferences are minimized by increasing or decreasing percentage of cloud cover on the "Max cloud (%)" menu. The additional images are included, extending the period of months into the early wet season (i.e. November – December) or early dry season (i.e. May - June) on the "Change dates" menu.
Production of training data	Training data should be representative of all land cover and transition types. Sample locations should be spread broadly across the study region, taking into account the variation in vegetation and its dynamics across the landscape. Care should be taken that sample areas cover only a single class. This may, for instance, require sampling only the central portion of a deforestation event. A relatively balanced sample is preferred; map producers are recommended to define one polygon of stable forest and stable non-forest for each polygon of deforestation. To ensure representative coverage of the more common classes, the size of polygons in stable forest and non-forest classes should be comparable to the scale of deforestation (see Figure 3). Given the range of land cover varieties encompassed by the stable non-forest class, additional polygons may be considered in this class. The locations of poor quality data in the composite satellite images should be avoided as it's hard to make a clear determination about its class. For example, where pixels flagged

as ‘no data’, ‘saturated’, ‘moisture content’, ‘dark features/shadows’, ‘black stripes’ and ‘fog/haze/cloud covering’, as shown in Figure 6, obscure the feature of interest. For example, where pixels flagged as ‘no data’, ‘saturated’, ‘moisture content’, ‘dark features/shadows’, ‘black stripes’ and ‘fog/haze/cloud covering’, as shown in Figure 12, obscure the feature of interest or modify their spectral response as detected by the sensor.

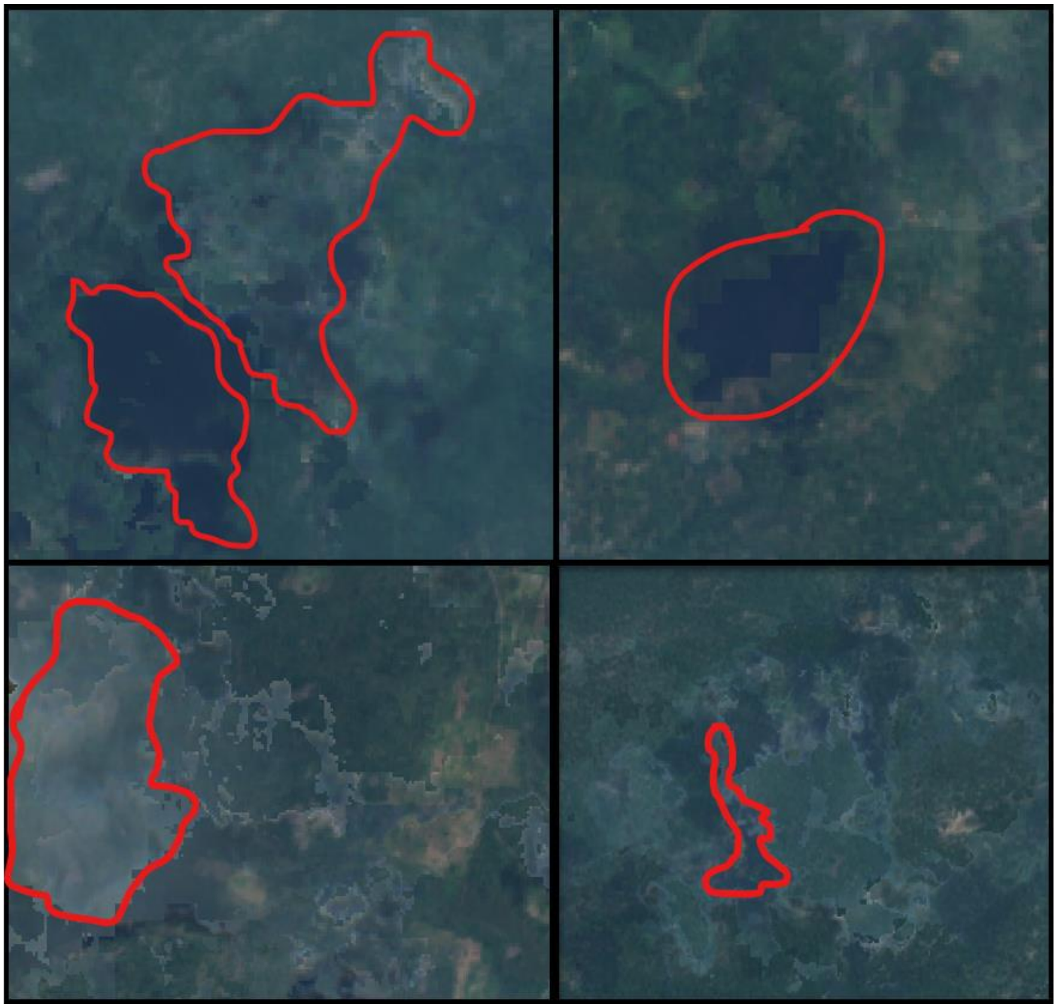


Figure 12 Some factors in satellite imagery that might bias the sample.

The data collection manager performs checks on the training data, during and after data collection. The quality checks involve the confirmation that:

- the sample polygon represents exactly their respective class, as shown in Figure 13;

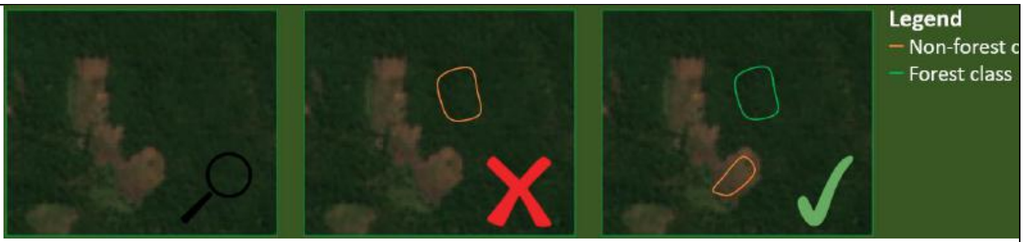
<table>
<tr>
<td></td>
<td>


Figure 13 Quality check of training data collection: Example 1

- the sample polygon covers exactly one class, as shown in Figure 14;

Figure 14 Quality check of training data collection: Example 2

- the sample is/was collected more towards the inside or central portion of the class (useful and indispensable for deforestation events), as shown in Figure 15; and

Figure 15 Quality check of training data collection: Example 3

- Each class contains at least 100 sample polygons representatively spread across the entire area of interest.

Incorrectly collected sample is discussed directly with the operator and corrected to as a continuing calibration exercise.
</td>
</tr>
<tr>
<td>Quality assessment</td>
<td>The mapping workflow provides the operator with quantitative and qualitative measures of map quality. Given the challenges associated with monitoring small areas of deforestation to be measured on annual time-scales, it is important that maps are of high-quality. In particular it is necessary to ensure that errors of omission in the deforestation class are rare, without errors of commission requiring a high sampling effort to attain a well constrained estimate of the area of deforestation.</td>
</tr>
</table>

Quantitative assessment

The mapping workflow provides the following estimates of uncertainty when training the classifier:

- **Out of bag error:** An estimate of overall (in)accuracy of the classifier
- **Re-substitution accuracy assessment:** Reports overall accuracy and User's and Producer's accuracy of the deforestation class by comparing predictions against the classifier training data. These figures should be viewed with caution, as even a poor classifier may reproduce its training set exactly.
- **Validation accuracy assessment:** Reports overall accuracy and User's and Producer's accuracy of the deforestation class by comparing predictions against quasi-independent data held back from the classifier. These figures will be better representative of the external validity of the classifier.

The most relevant of these statistics are the User's (1 – commission error) and Producer's (1 – omission error) accuracy reported for the deforestation class using independent validation data. There should be a particular focus on maximising the Producer's accuracy in order to reduce the probability of errors of omission in the deforestation class, a complicating factor for stratified area estimation (Olofsson *et al*. 2020).

The reported accuracy realistically obtainable will vary depending on the area of interest, with a general guide being an aim for at least 80% User's and Producer's accuracy reported for the deforestation class. As area estimates derive from reference data in place of pixel-counts it is acceptable for this accuracy to be lower, however the operator should be aware that lower accuracies will increase required sampling intensity so all reasonable efforts should be made to maximise map quality.

Random Forest classifiers also report 'importance values', which measure the relative influence of each of the classifier inputs on predictions. These may be used to identify which image features might be removed and retained to improve the model (Figure 16).

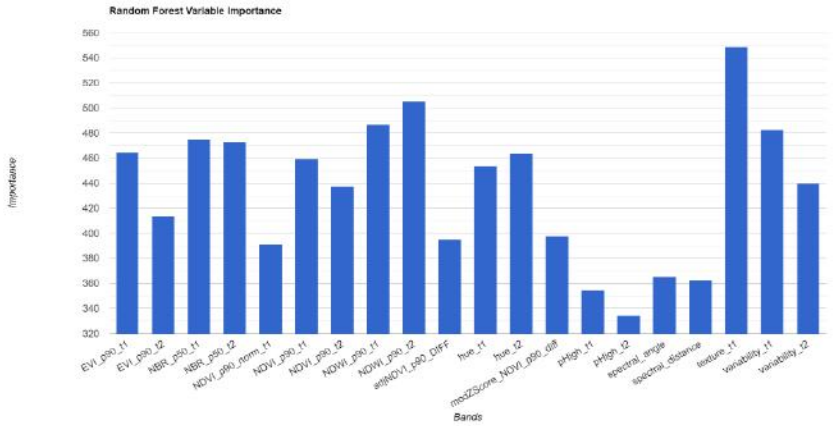
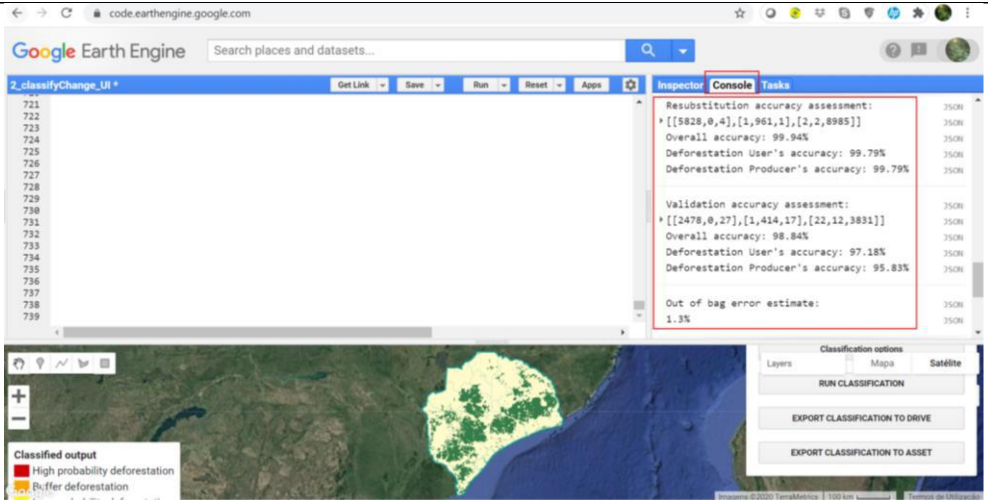


Figure 16 Example Random Forest 'importance values', showing the relative contribution of input features to classifier predictions.

The estimates of uncertainty described above are displayed at the section "Classification" from the FNDS mapping interfaces, specifically on the "Console Tab", as shown in Figure 17. The "Console Tab" displays the estimates after run the script provided by the section.

```
Resubstitution accuracy assessment:
[[5828,0,4],[1,961,1],[2,2,8985]]
Overall accuracy: 99.94%
Deforestation User's accuracy: 99.79%
Deforestation Producer's accuracy: 99.79%

Validation accuracy assessment:
[[2478,0,27],[1,414,17],[22,12,3831]]
Overall accuracy: 98.84%
Deforestation User's accuracy: 97.18%
Deforestation Producer's accuracy: 95.83%

Out of bag error estimate:
1.3%
```

Figure 17 The estimates of uncertainty displayed when training the classifier in the section "Classification" from the FNDS mapping interfaces on the Google Earth Engine Cloud Computing Platform.

Visual assessment

As important as quantitative measures of map quality is a visual inspection of the map. This can be performed by comparing map outputs against the reference imagery used for training data collection. Like with quantitative measures, this should focus on the deforestation class.

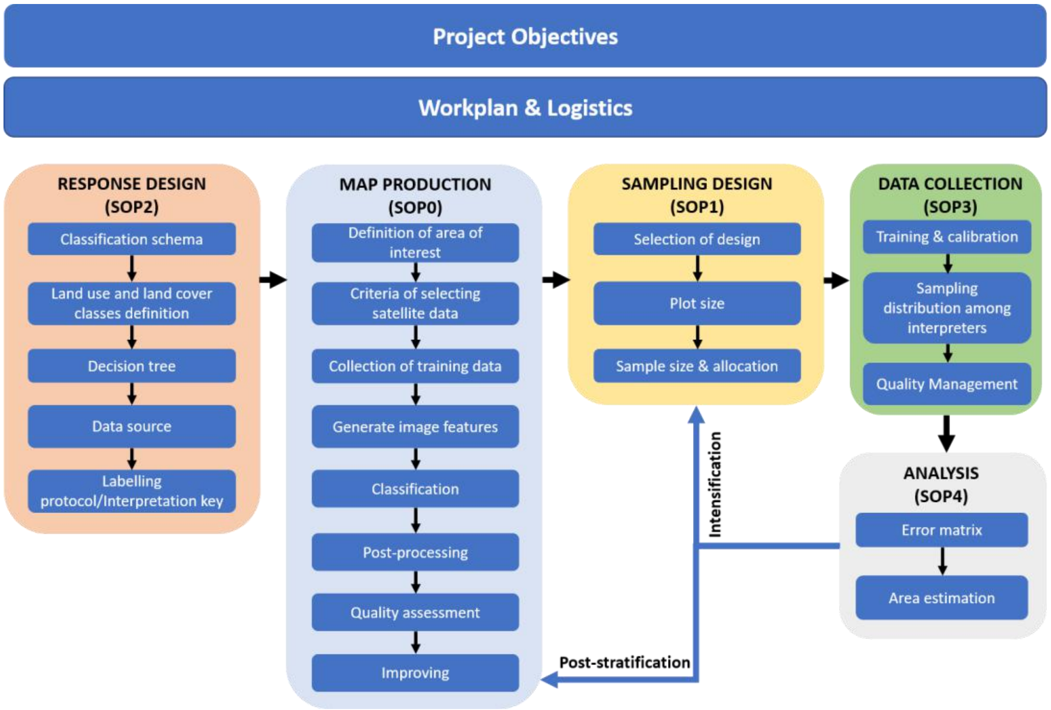
Given the importance of reducing errors of omission in the deforestation class, if there is any strong indication of errors of omission of deforestation from the high/low or buffer deforestation class, the operator must improve the map before continuing.

Improving the map

The output map is shared and submitted to data manager for performing the visual assessment. Where the map does not meet the quality thresholds set out (e.g., gross errors of omission and commission easily noticed and confirmed), the operator has the following options available to improve the map:

- **Parameter tuning** – Increase the number of trees used in the Random Forest classifier.
- **Feature selection –** Addition or removal of image features can improve model performance. This can be guided by improvement of error statistics, visual assessment, and importance values reported by the classifier.
- **Addition of further training data** – New data should be focussed on regions and land covers where the existing map quality is poor.
- **Manual editing** – Editing of errors in map outputs is encouraged, though this can be time consuming so is considered a last resort. It is important that all edits are made separately and independently of reference data sampling (SOP1).

These improvement options should be performed before and after the visual assessment conducted by data collection manager.

Post stratification	In some cases, errors of omission in the deforestation class may be discovered after the finalisation of map production. This could be either through future visual assessment, or after encountering an error of omission during stratified area estimation. In these cases, this mapping procedure may be re-run with improvements applied. The three deforestation classes are combined into a single class (class 6), supporting a new map stratum of deforestation pixels that were previously included in the stable class of the original map. Where post-stratification is driven by an error of omission, it is important that map edits should be performed independently of area estimation. It is not acceptable to tweak the map until the error of omission is captured, so the method to be employed to improve the map must be decided in advance of map production.
Data Flow Diagram	
References	Breiman, L. 2001. Random forests. *Machine Learning, 45*, pp. 5-32. Hamunyela, E., Verbesselt, J. & Herold, M. 2016. Using spatial context to improve early detection of deforestation from Landsat time series. *Remote Sensing of Environment, 172*, pp. 126-138. Olofsson, P., Arévalo, P., Espejo, A., Green, C., Lindquist, E., McRoberts, R. & Sanz, M. 2020. Mitigating the effects of omission errors on area and area change estimates. *Remote Sensing of Environment, 236:* 111492.

SOP 1: SPATIALLY REFERENCED SAMPLE DESIGNS

STANDARD OPERATING PROCEDURE 1 (SOP1): SPATIALLY REFERENCED SAMPLE DESIGNS FOR AREA ESTIMATION

Section 1 Overview

Purpose	Establish a spatially-referenced, probability-based and geographically balanced sampling design for the estimation of areas in land surveys.
Scope	The reporting requirements will identify the sampling and target populations.
Responsibilities	Remote-sensing/GIS and statistics technical officers, as well as policy- and decision makers in REDD+ countries and international agencies, multilateral and bilateral programmes, in particular for the estimation of activity data (i.e., land use/land cover change) for emissions/removals in REDD+ activities.
Prerequisites	• Project objectives • Land and/or land use change definitions • Region or area of interest defined • Response design, aligned with land/land use change definitions • Map production (SOP0)
Requirements	• Spatially explicit representation (i.e., satellite imagery) of land or else geospatially referenced field ground data • Maximum allowable error (d), given as an area fraction, $E \in (0,1)$. This is the maximum half-width of the confidence interval we aim towards in our estimate. • Confidence level for the estimation of uncertainties • Access to the FNDS mapping interfaces on the Google Earth Engine Cloud Computing Platform.

Section 2 Procedure

Sampling designs and sampling units

Sampling units

The sampling unit is centered on the 20m pixel. See response design (SOP2) for more details. All pixels are derived from an annual deforestation map and classified as stable forest, stable non-forest, or one of several deforestation probability classes. Map classes are as follows:

Table 3 Map classes

Label	Class
High probability deforestation	1

Buffer deforestation	2
Low probability deforestation	3
Stable forest	4
Stable non-forest	5
(Post-stratified deforestation)	6

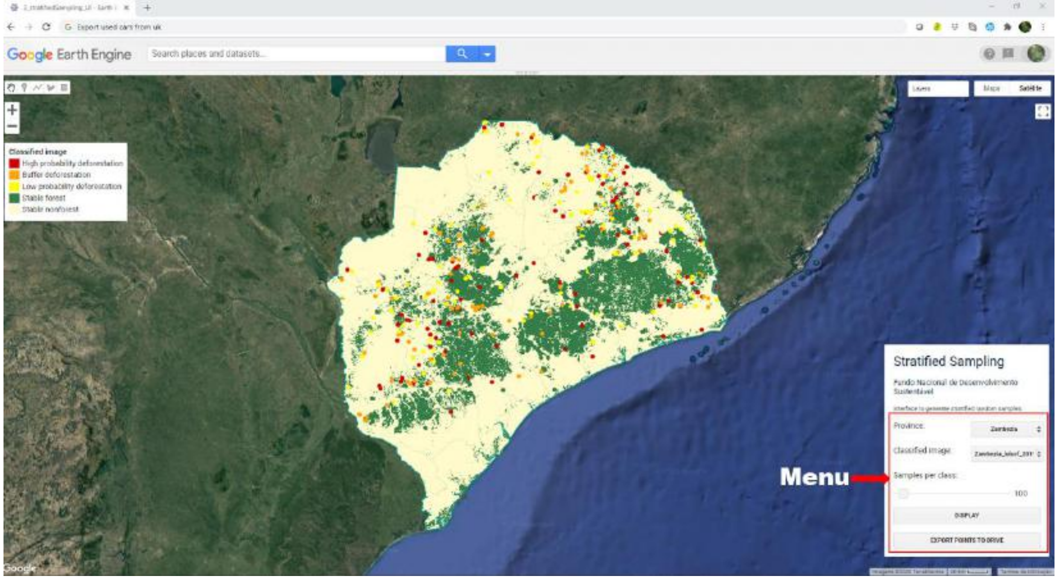
Classes 1-5 will always be present, whereas class 6 is specific to the case where post-stratification has been performed to improve map quality due to the omission of deforestation events. See Map production – SOP0 for more details.

Sampling design

According to Olofsson *et al*. (2014) and FAO (2016), stratification is recommended to ensure a sufficient representation of rare classes (e.g. that only represent a small proportion of the area of interest – forest loss) and improve the precision of the accuracy and area estimates by increasing the sampling density in the change classes. For this reason, the sampling design is Stratified Random Sampling (STR), with the strata taken from the annual deforestation map (see Map production – SOP0). Sample points should be centered on map pixels and all have unique locations (i.e. no double-counting).

The sampling design is performed in the FNDS mapping interfaces, specifically the section "Stratified Sampling", as shown in Figure 18. This section was designed to apply a fixed number of samples defined by the user for each map class. For generating samples, the operator should:

- Run the script provided by the section;
- Load the boundary polygon for the area of interest on the "Province" menu;
- Load the annual deforestation map on the "Classified image" menu;
- Set the desired number of samples on the "Samples per class" menu;
- Display the samples by clicking on the button "DISPLAY"; and
- Export the generated samples in "CSV format" by clicking on the button "EXPORT POINTS TO DRIVE".



Stratified Sampling

Fundo Nacional de Desenvolvimento Sustentável

Interface to generate stratified random samples

Province: Zambezia

Samples per class: 100

Figure 18 *The section "Stratified Sampling" from the FNDS mapping interfaces on the Google Earth Engine Cloud Computing Platform, for stratified sampling of deforestation map.*

Estimating sampling size

The sample size *n* is determined from the equation (Cochran, 1977):

$$n = \frac{(\sum W_i S_i)^2}{\left[S(\hat{O})\right]^2 + (\frac{1}{N})\sum W_i S_i} \approx \left(\frac{\sum W_i S_i}{S(\hat{O})}\right)^2$$

Equation 1

Where:

N is the number of units in the region of interest,

S(Ô) is the standard error of the estimated overall accuracy that we would like to achieve,

W_i is the mapped proportion of area of class *i*; and

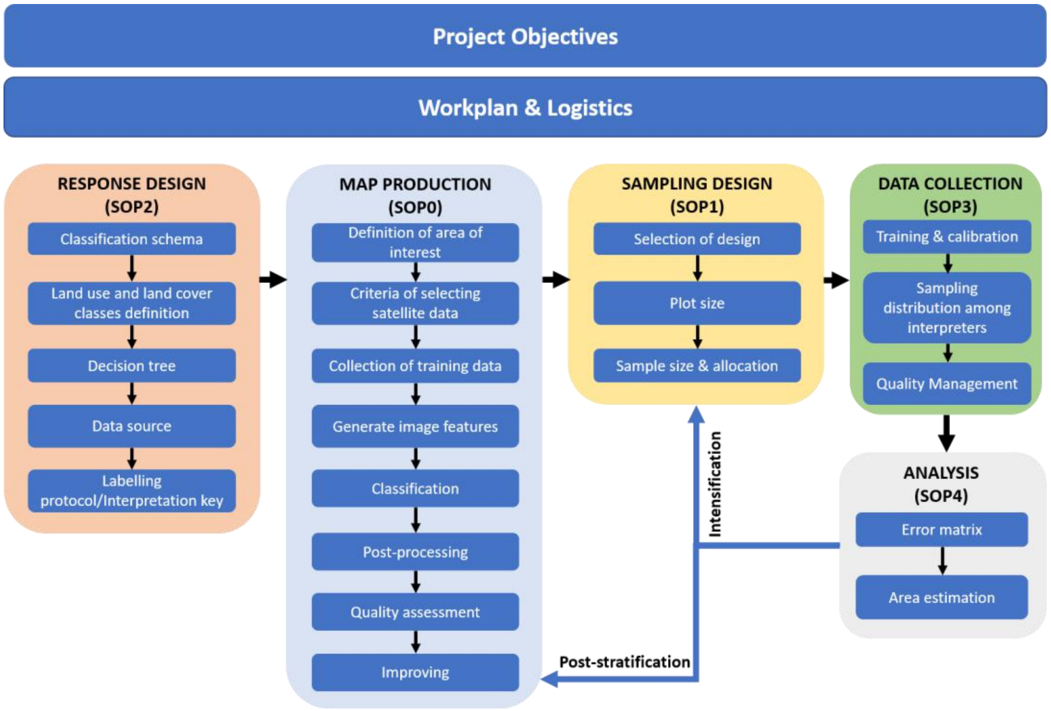
S_i is the standard deviation of stratum *i*.

In order to obtain approximate values of proportion of deforestation in each stratum (p_i), a pilot sampling is conducted. According to Congalton and Green (2008) and Olofsson *et al* (2014), the minimum sample size should be at the least 20 to 100 samples per stratum. For this reason, the pilot sampling uses 100 sample points for each map stratum.

After the pilot sampling, samples need to be added to each stratum, in order to reach the desired relative error. It was decided to use the Optimum (Neyman) allocation (Neyman, 1934), where the stratum standard deviation $S_h = \sqrt{U_h \cdot (1 - U_h)}$ increases the number of plots (ensuring larger numbers of plots in rare classes or strata) and sampling unit costs are constant:

$$n_h = n \frac{w_h \cdot S_h}{\sum_{h=1}^{H} w_h \cdot S_h}$$

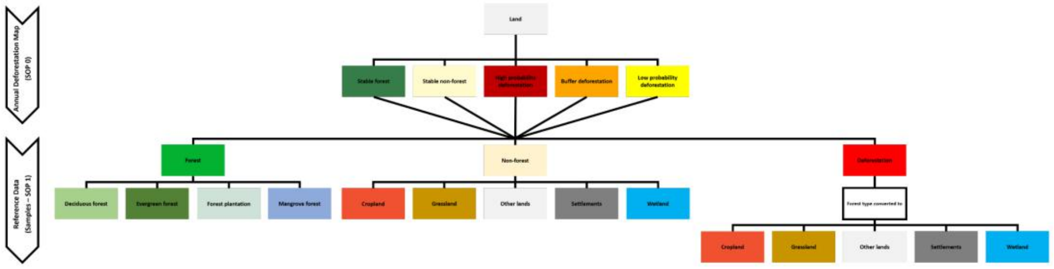
Equation 2

	There should be a minimum of 300 sample units in each of the stable classes. The reason behind this minimum is that if no deforestation events are found in the 100 sample units of each stable stratum, then p_i will be 0, and we would require no further sampling of these strata. This would mean that the sample size for the stable strata would be much smaller than for the change strata, even though the stable classes make up the majority of map area.
Section 3 Quality management	
Sampling designs and sampling units	Each sample unit should have a unique location, without repeat-measurements at the same location. It should be checked that no sample units overlap between the pilot sampling phase and adding of new strata. This can be ensured by generating a large number of random sample points at the outside (~1000 per class) and drawing new sample units from these. The same procedure also should be ensured in the initial sampling so that there is no a replacement of a sample from the pilot sampling with adding additional samples (if necessary), once the additional samples will be extracted from the same source.
Sampling size	The initial sampling size is 100 for all strata, in order to have a first estimate of proportion of deforestation in each stratum. This allows the determination of total sample size required for desired error. Sample size for stable classes, in the case of no omission errors of deforestation, is set to a minimum of 300. This is to reduce the probability of not finding omission errors “by chance”.
Data Flow Diagram	

References	Cochran, W. G. 1977. Sampling techniques. John Wiley & Sons. Congalton, R. G. & Green, K. 2008. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Second Edition. CRC Press FAO. 2016. Map Accuracy Assessment and Area Estimation: A Practical Guide. National forest monitoring assessment working paper No.46/E, 60p. Neyman, J. 1934. On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. Journal of the Royal Statistical Society, Vol. 97, 558-606. 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 Sensing of Environment, 148:42–57.

SOP 2: RESPONSE DESIGN

STANDARD OPERATING PROCEDURE 2 (SOP2): RESPONSE DESIGN

Section 1 Overview	
Purpose	Assign to a spatial point a land cover / land use class following an objective protocol that reduces interpreter bias
Scope	The reporting requirements
Responsibilities	Technicians in charge of the visual interpretation of samples
Prerequisites	Sampling design available (SOP 1)
Requirements	Define an overall classification scheme Document class definitions Formalize the decision tree Anticipate specific cases and exceptions **Important Note:** …………………
Section 2 Procedure	
Classification scheme	
Definitions	**1. Stable forest:** Forest land that remains forest land based on annual deforestation map (See SOP 0 for further information). **2. Stable non-forest**: Non-forest land that remains non-forest land based on annual deforestation map (See SOP 0 for further information). **3. High / low probability deforestation**: Forest land that converted to non-forest land based on annual deforestation map (See SOP 0 for further information). **4. Buffer deforestation**: Buffer around the high probability deforestation (See SOP 0 for further information). **5. Forest** 5.1. **Forest plantation:** Land with a minimum of 1 ha, containing more than 30% cover of exotic tree species (pine, eucalyptus, etc.) with the potential of reaching at least 3m in height. **5.2. Deciduous forest**

5.3. **Semi-deciduous closed forest:** Land with a minimum area of 1ha and containing native semi-deciduous tree species with at least 3m in height and canopy cover of ≥ 65%.

5.4. **Semi-deciduous open forest:** Land with a minimum area of 1ha and containing native semi-deciduous tree species with at least 3m in height and canopy cover of ≥ 30 and ≤ 65%.

5.5. Evergreen forest

5.5.1. **Evergreen closed forest:** Land with a minimum area of 1ha, containing native evergreen tree species with at least 3m in height and canopy cover of ≥ 65%.

5.5.2. **Mountain forest:** Land with a minimum area of 1ha and altitude above 1000m, or with a steep decline, containing native evergreen tree species with at least 3m in height and canopy cover of ≥ 30%.

5.5.3. **Evergreen open forest:** Land with a minimum area of 1ha, containing native evergreen tree species with at least 3m in height and canopy cover of ≥ 30 and ≤ 65%.

5.6. **Mangrove forest:** Transition land between terrestrial and marine ecosystems with a minimum area of 1 ha, containing native mangrove tree species of at least 3m in height and canopy cover of ≥ 30%.

6. Non-forest

6.1. Cropland

6.1.1. **Tree crops:** Land with a minimum area of 1 ha, with ≥ 20% cover of tree crop species (coconut, mango, macadamia, cashew, etc.).

6.1.2. **Field crops:** Land with a minimum of 1 ha, with ≥ 20% cover of field crops (corn, cassava, beans, manioc, soy, etc.).

6.2. Grassland

6.2.1. **Grassland:** Land with a minimum area of 1 ha, dominated by grass species and with a tree and shrub cover of ≤ 10%.

6.2.2. **Shrubland/Wooded grassland/thicket:** Land with a minimum area of 1 ha, with ≥ 20% cover of woody species below 3m in height and tree cover of <30%.

6.3. Wetlands

6.3.1. **Flooded herbaceous vegetation:** Land with a minimum area of 1ha, dominated by grass species and with a tree and shrub cover of ≤ 10% and permanently or temporarily flooded.

6.3.2. **Flooded shrub vegetation:** Land with a minimum area of 1ha, dominated by woody species and with a tree and shrub cover of ≥ 10% and permanently or temporarily flooded.

6.3.3. **Water bodies:** Land with a minimum of 1ha, and ≥ 20% water cover.

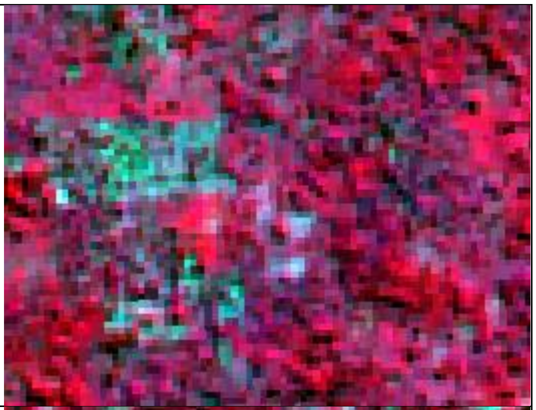
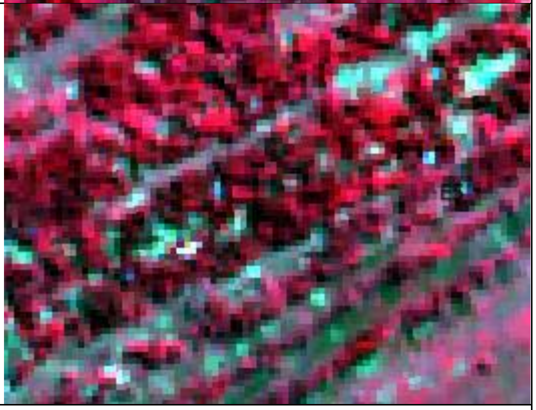
6.4. **Settlements:** Land with a minimum area of 1ha, and ≥ 20% cover of infra-structure (houses, roads, railways, buildings, etc.).

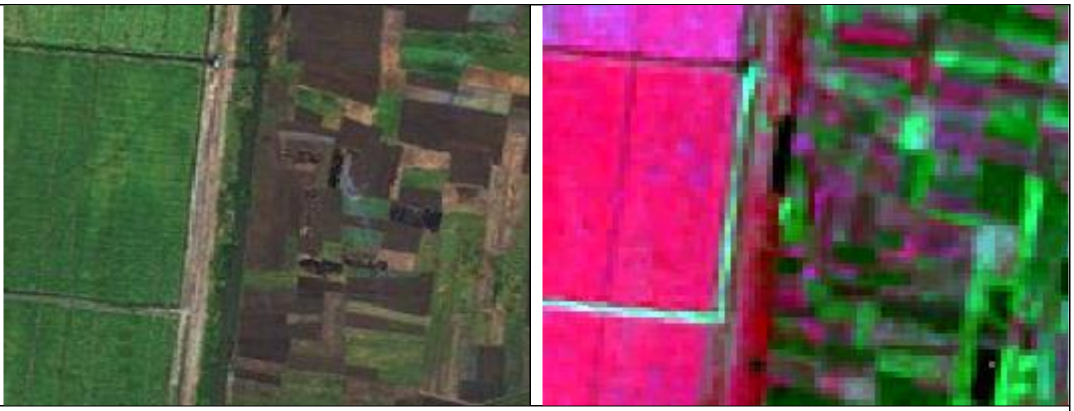
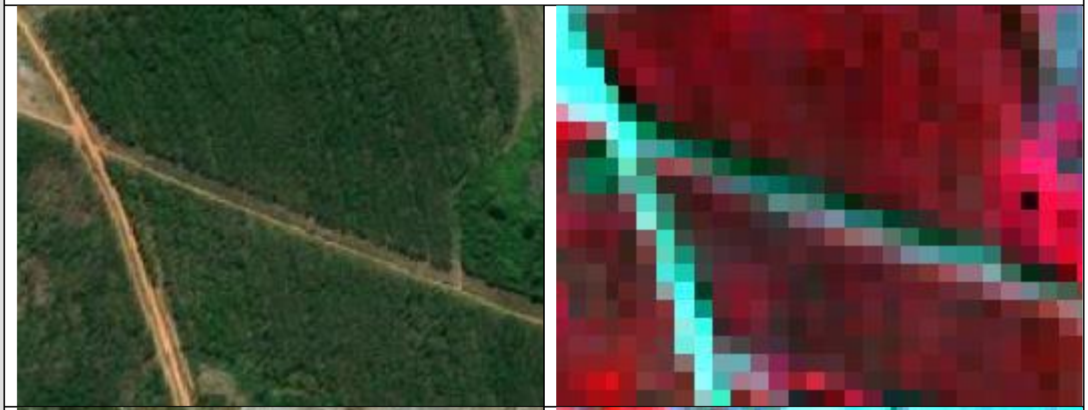
6.5. **Other lands**

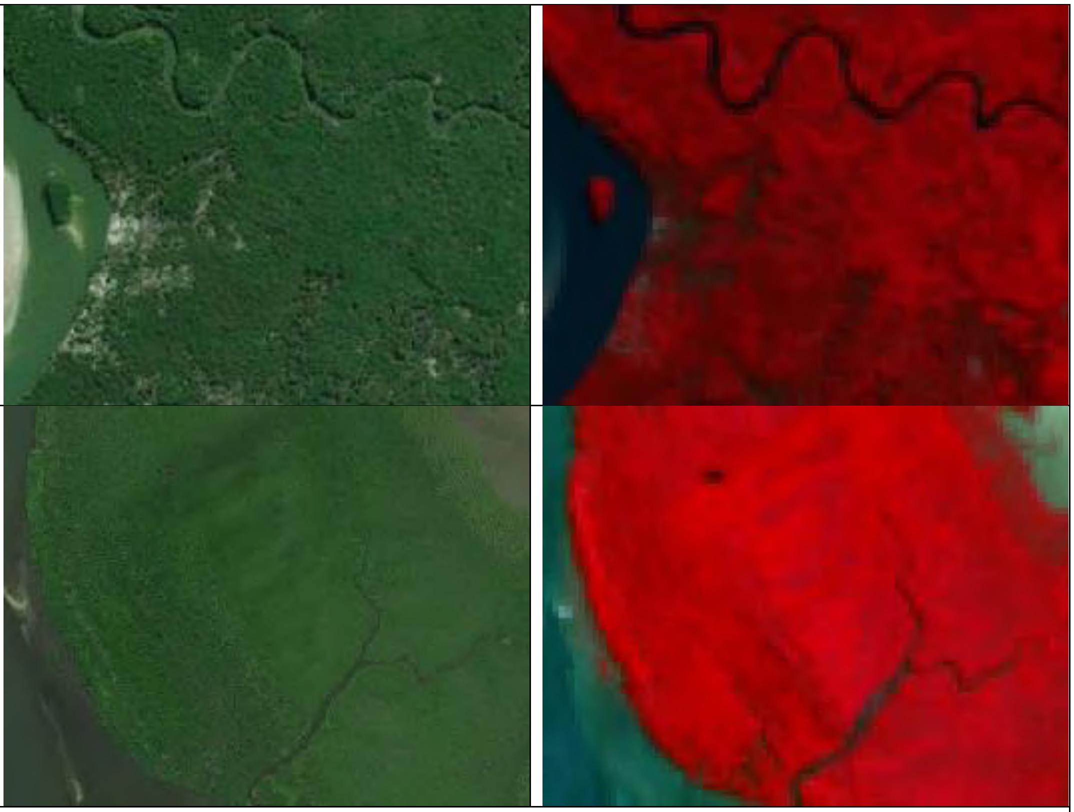
6.5.1. **Bare soil:** Land with a minimum area of 1ha and ≥ 20% cover of soil with no vegetation.

6.5.2. **Bare rock:** Land with a minimum area of 1ha and ≥ 20% cover of rocks with no vegetation.

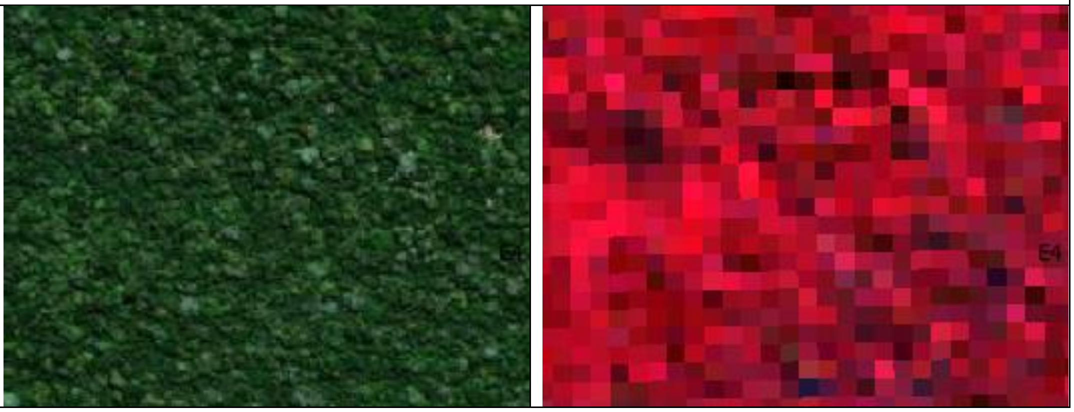
Interpretation key	**Bing Maps Imagery**	**False color composites of Sentinel-2A/Landsat Image**
	Tree crops	

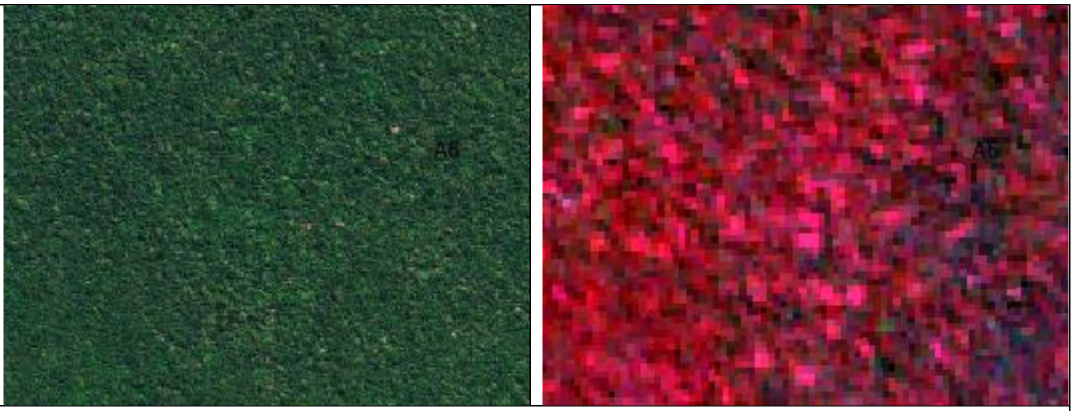
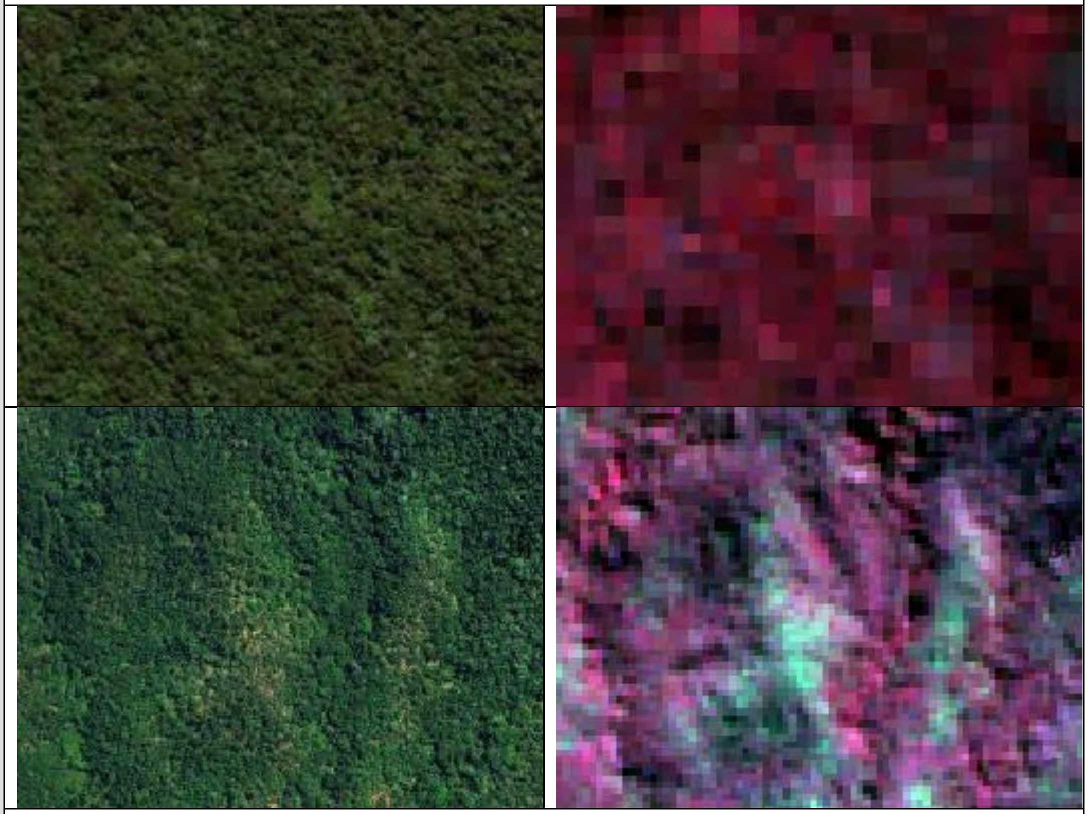
		
		
	Field crops	
		

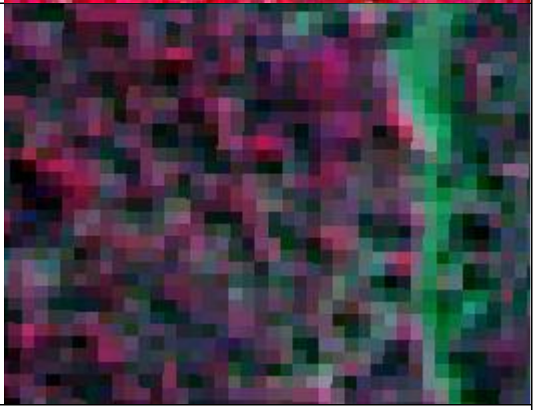
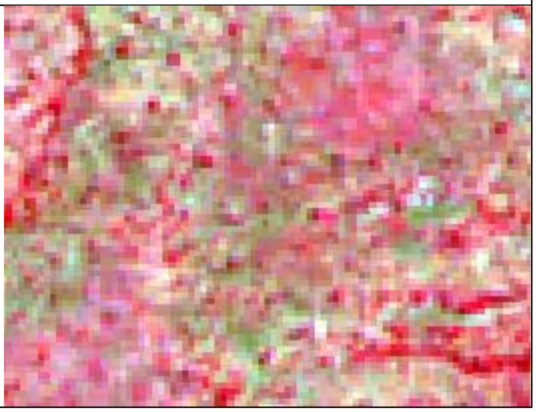
	
	Forest plantation
	
	
	Mangrove forest

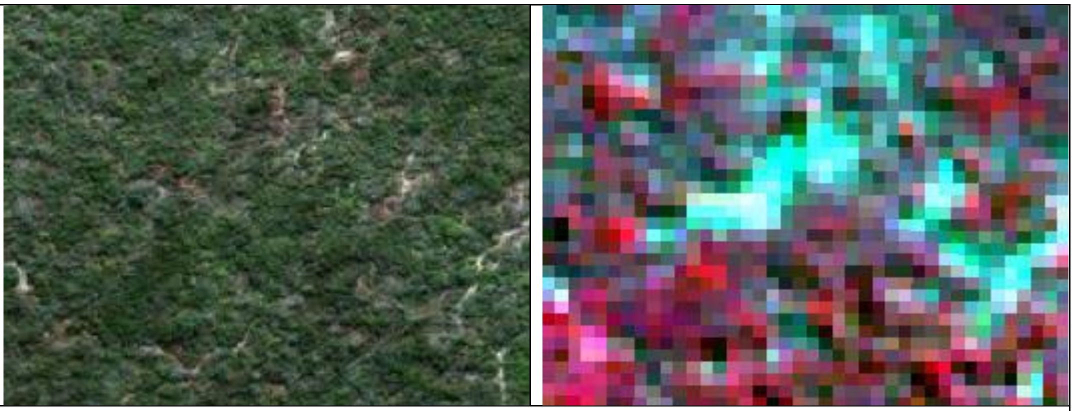


Evergreen closed forest

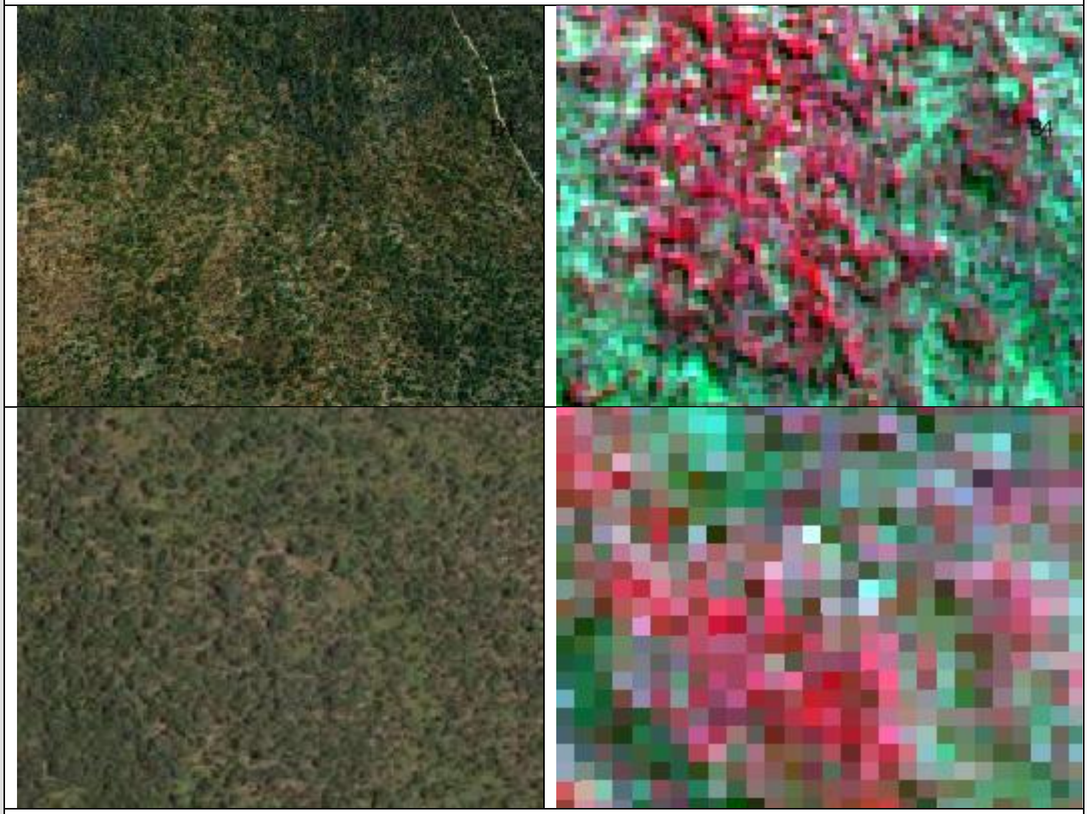


	
	Mountain forest
	
	Semi-deciduous closed forest

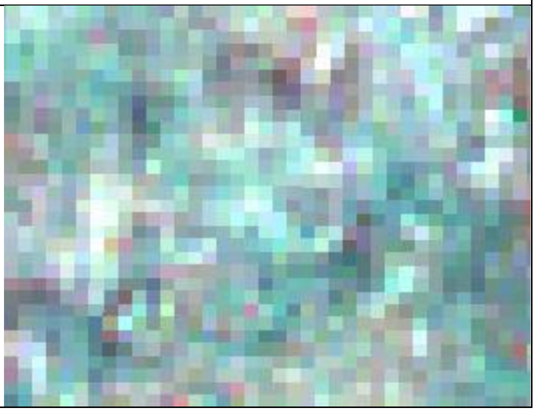
		
		
	Evergreen open forest	
		

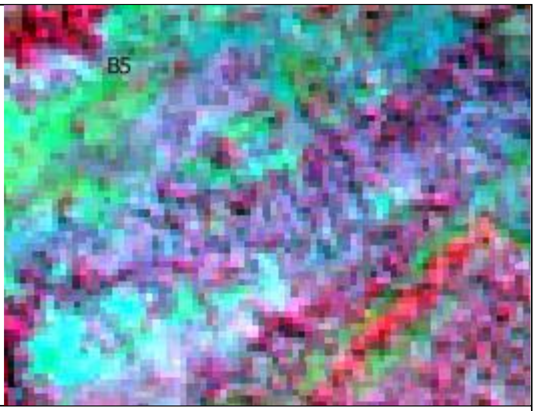
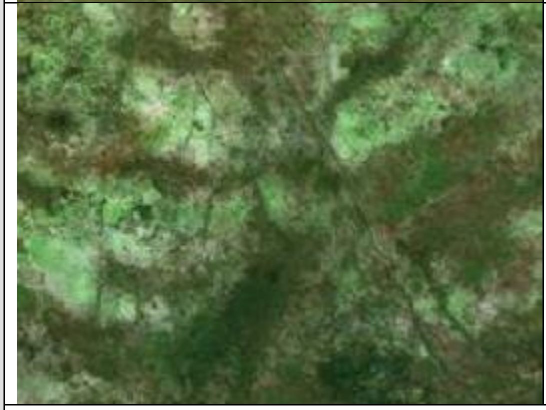
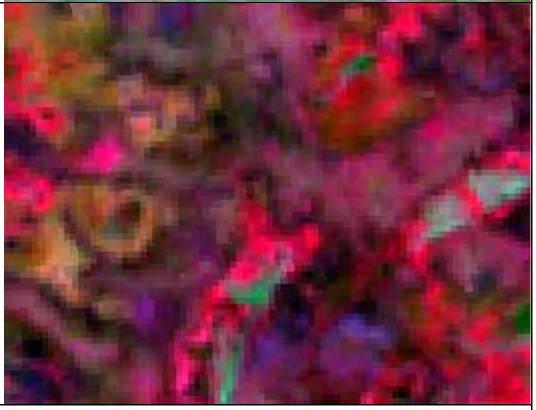


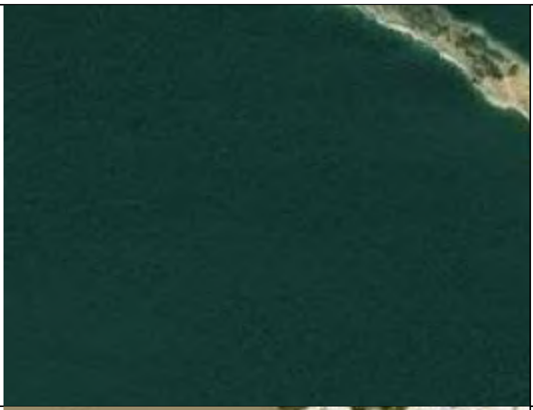
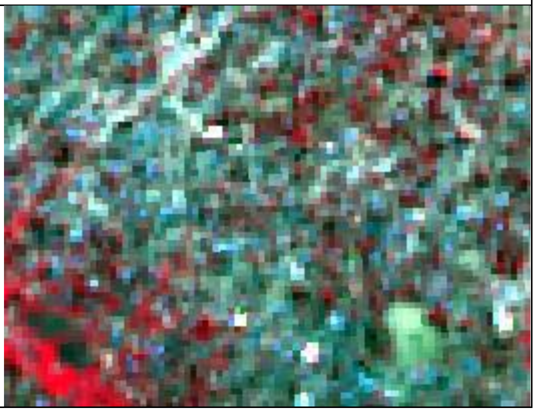
Semi-deciduous open forest

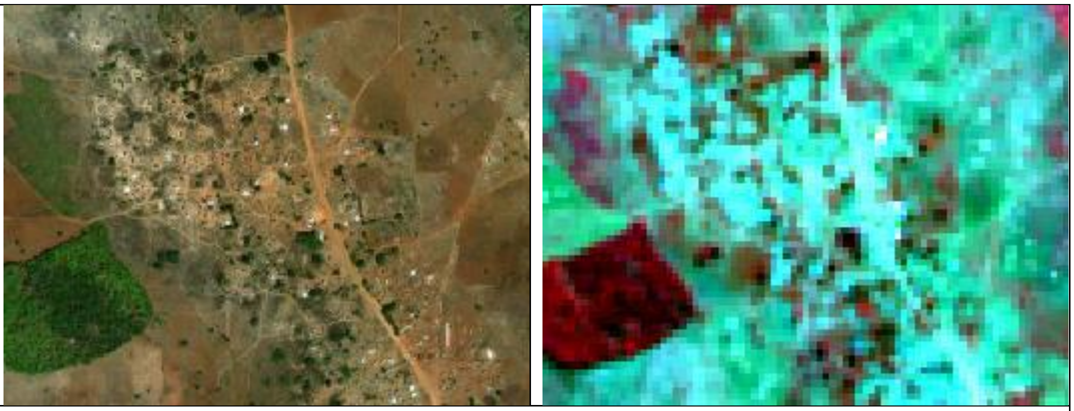
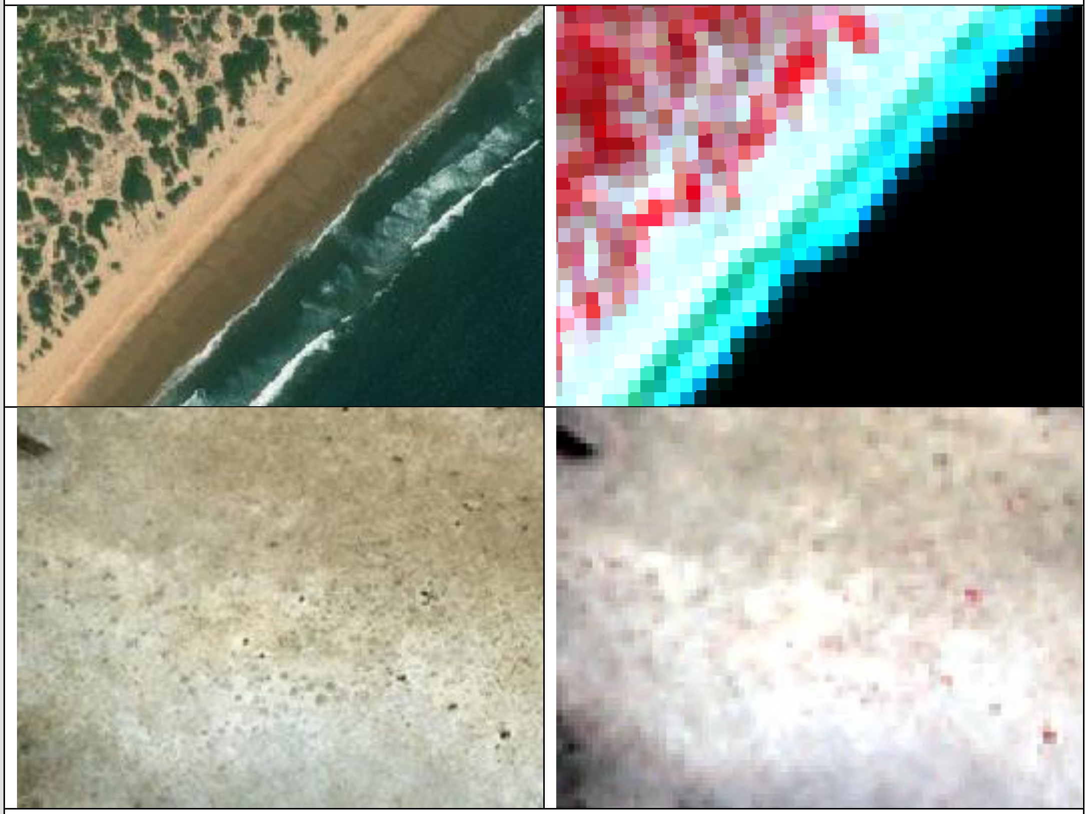


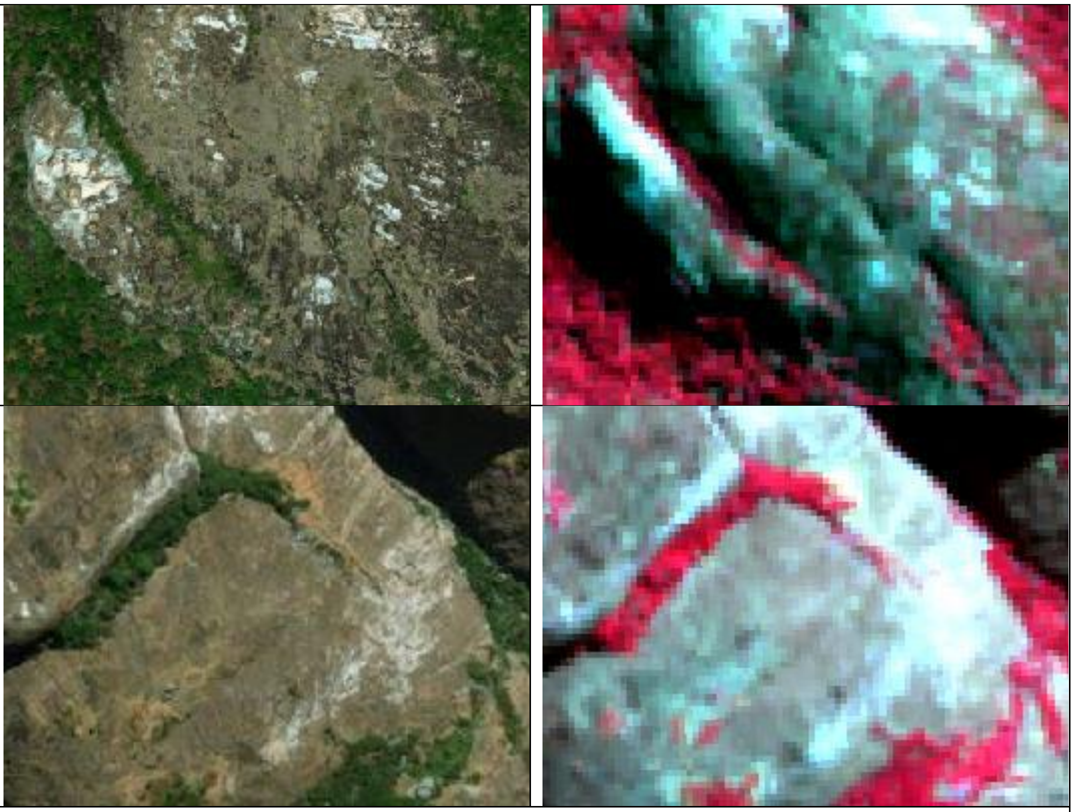
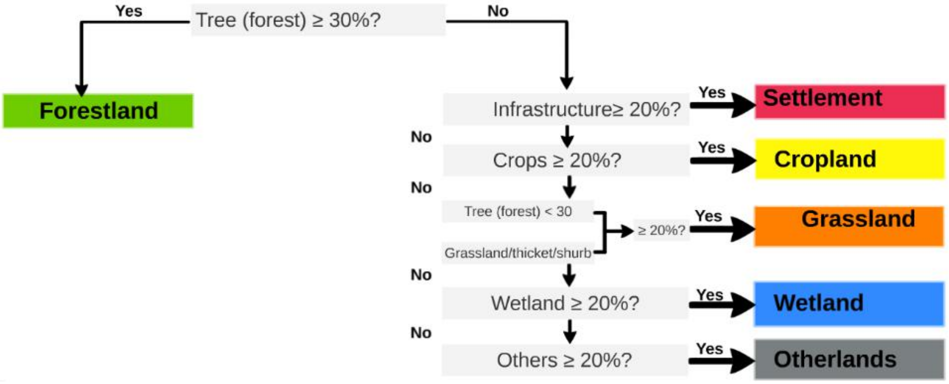
Grassland

		
		
	Shrubland	
		

		
	Flooded herbaceous vegetation	
		
		
	Water bodies	

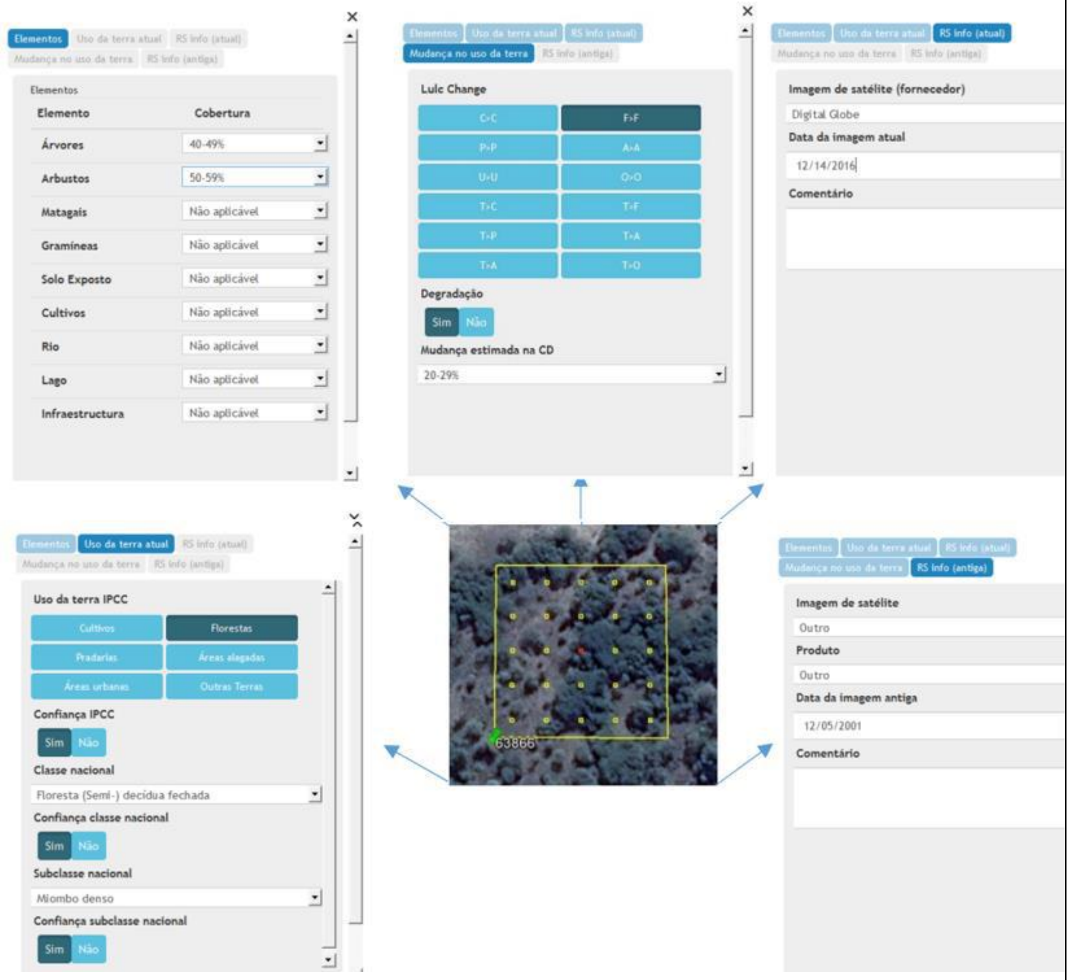
		
		
	Settlements	
		

	
	Bare soil
	
	Bare rock

	
Decision tree	The decision tree is a set of rules that help the interpreter assign an overall land use class when the sample is composed of mixed land cover features. For instance, if the scheme is composed of the 6 IPCC Land representation categories with a set of land cover features, the decision tree can be illustrated like this:  *Figure 19 Decision Tree* So any sampling unit that has 30% of tree canopy cover is considered a forest even if it has more than 20% of settlements, crops or other land use, the forest has priority. In the

<table>
<tr>
<td></td>
<td>case the sampling unit is classified as forest land and different forest types are present in the sampling unit, a majority rule is used, i.e. the largest forest class is the winner.</td>
</tr>
<tr>
<td>Confidence</td>
<td>An indicator of the confidence of the interpretation is available for all samples, with two modalities (yes / no).</td>
</tr>
<tr>
<td>Data sources</td>
<td>For the land cover and land use change interpretation is allowed to use the data from Google, Bing, Planet, Sentinel-2, Landsat, MODIS.</td>
</tr>
<tr>
<td>Working environment</td>
<td>

The samples are assessed using Collect Earth Desktop. This tool enables access to high-resolution images in conjunction with Google Earth, Bing Maps and Planet Labs, as well as a medium resolution image repository available through Google Earth Engine Explorer and Code Editor (Landsat and Sentinel-2). The tool enables to display the digital forms designed to collect the Land-Use Land Cover Change and Forestry (LULCCF) information on the sampling points (Figure 20). The technical document "Passo a Passo para o Levantamento e Estimativa de Emissões do Sector de Uso da Terra, Mudanças do Uso da Terra e Florestas" provides further description details of the forms and how fill them.



Elementos

Elemento	Cobertura
Árvores	40-49%
Arbustos	50-59%
Matagais	Não aplicável
Gramíneas	Não aplicável
Solo Exposto	Não aplicável
Cultivos	Não aplicável
Rio	Não aplicável
Lago	Não aplicável
Infraestructura	Não aplicável

Uso da terra IPCC
Florestas
Confiança IPCC
Sim
Classe nacional
Floresta (Semi-) decidua fechada
Confiança classe nacional
Sim
Subclasse nacional
Miombo denso
Confiança subclasse nacional
Sim

Lulc Change
F›F
Degradação
Sim
Mudança estimada na CD
20-29%

Imagem de satélite (fornecedor)
Digital Globe
Data da imagem atual
12/14/2016
Comentário

Imagem de satélite
Outro
Produto
Outro
Data da imagem antiga
12/05/2001
Comentário

Figure 20 The digital forms designed to collect the LULCCF information

</td>
</tr>
</table>

Moreover, the tool displays the spatial sampling unit from each sample defined as a 100m x 100m plot (1ha). The spatial sampling unit contains an internal grid of 5 x 5 points (20m x 20m grid). Each point weights 4% (Figure 21).

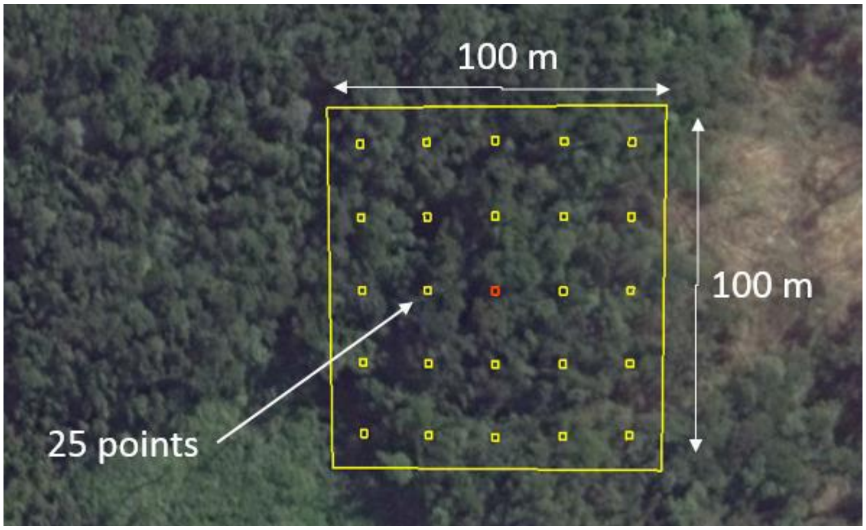
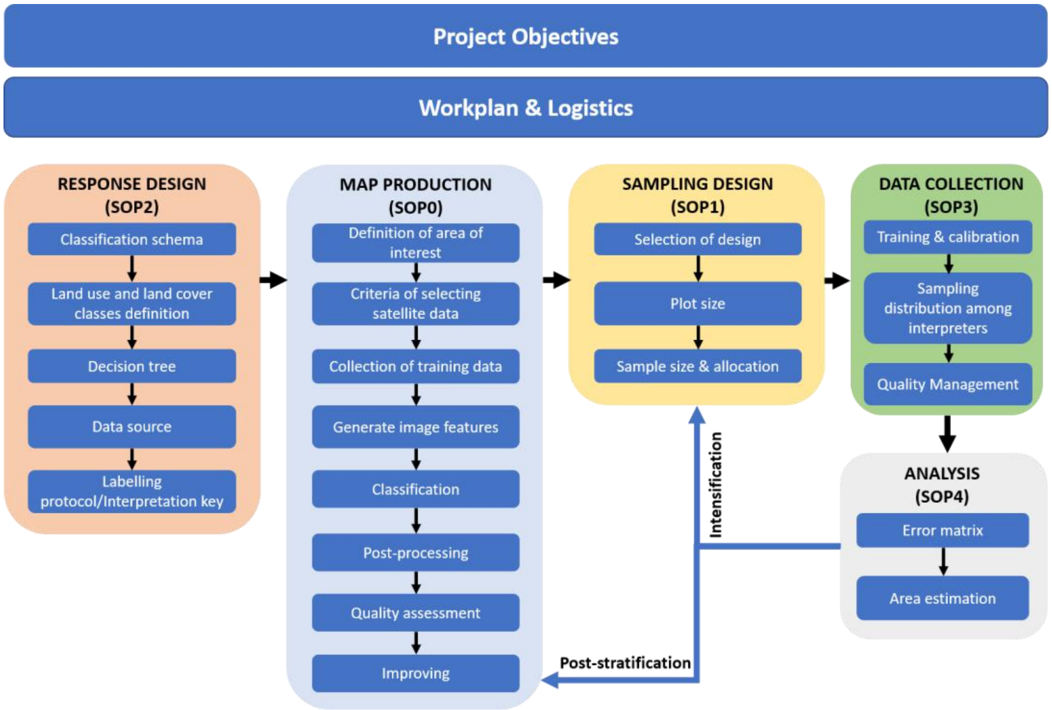


Figure 21 Spatial sampling unit

Data Flow Diagram



References

SOP 3: DATA COLLECTION

STANDARD OPERATING PROCEDURE 3 (SOP 3): DATA COLLECTION

Section 1 Overview	
Purpose	Provide a framework for sample data collection and quality management
Scope	Technical decision and policy makers in REDD+ countries and international agencies, multilateral and bilateral programmes, in particular for the estimation of activity data (i.e., land use/land cover change) for emissions/removals in REDD+ activities.
Responsibilities	• Data collection manager • Technicians in charge of the visual interpretation of samples, referred to as interpreters or supervisors.
Prerequisites	• Sample design (SOP 1) • Response design (SOP 2)
Requirements	• Hardware and software • Internet connection • Human resources • Adequate time for data collection and quality management • Training and calibration of interpreters before and during data collection • Iterative process for data collection to improve data quality • Quality management implemented before, during and after data collection **Important Note:** …………………
Section 2 Procedure	
Data collection requirements	The sample-based data collection procedure provides the framework for collecting an unbiased and accurate dataset that can be used to quantify land use and land use change over an area of interest. This standard operating procedure focuses on using primarily remotely sensed data for collecting sample information. Quality management is essential in order to achieve the highest possible standard of quality for the data collected and it requires: • Sampling design (SOP1) • Response design (SOP2) • Computer hardware system ○ Windows 7, Windows 10 or later ○ 2.5 GHz or faster processor ○ 4 or more GB of RAM • Stable and fast internet connection speed • Software

	○ Open Foris Collect Earth ○ Google Earth ○ Google Chrome or Mozilla Firefox ○ Microsoft Office Excel • Batches of sample data • Free access to medium and high-resolution satellite imagery sources • Auxiliary data • Database storage • Human resources ○ Data collection manager ○ Team of specialized and trained image interpreters ○ Supervisors well-qualified for the data collection • Daily data collection target – 50 samples per image interpreter
Training and calibration	Sample based data collection should be consistent both across and within interpreters, as much as possible. This means: i) given the same sample, two different interpreters will return similar results ii) given similar samples over time, the same interpreter will return similar results (Patterson and Mccallum, 2020). Training of interpreters covers the theoretical domains of the technical document "Passo a Passo para o Levantamento e Estimativa de Emissões do Sector de Uso da Terra, Mudanças do Uso da Terra e Florestas", a detailed discussion and review of the classification system, response design, and image interpretation key (Patterson and Mccallum, 2020). The image interpretation key is described and classes clearly-defined in the response design document (SOP 2). During the training session, example samples are reviewed as a group and a discussion is facilitated to identify any changes needed in the definitions to reflect what can be identified in the available imagery, and consequently updating and expanding the image interpretation key to ensure the integrity cover new target areas. Additional example samples are distributed to each of the interpreters and they individually interpret the samples. The interpretations are compared and the agreement among the interpreters is determined. Differences in interpretations are discussed and a consensus is reached for each of the samples to calibrate the interpreters and better achieve similar results between interpreters.
Sample distribution	To commence the data collection, sample units are randomly split into batches of 100 samples or smaller. This aims to increase team productivity, organization, and time management. It is recommended that the samples are distributed randomly, to avoid introducing bias from a particular interpreter. The order of samples should be randomised to avoid introduction of systematic errors. Each batch of samples should ensure the representativeness of each one change map strata described in the SOP0 and SOP1.

<table>
<tr>
<td></td>
<td>For each batch, 10% of samples are randomly selected and duplicated, by applying a Proportionate Stratified Random Sampling based on deforestation map strata on the batch, and then assigned to different interpreters.

After the interpreter finishes the collection of a batch, the batch database is labelled with the interpreter's name, batch number, area of interest, time period of interest and date of its creation. The results of the sample data collection of the respective batch are submitted to the data manager for quality management. The data manager also labels the batch database with their name, area of interest, time period of interest and date of its creation.</td>
</tr>
<tr>
<td>Use of data sources</td>
<td>During the data collection, interpreters need to use all the data available, according the data sources defined in the response design (see SOP 2), in order to assist in the sample assessment. The coverage of sub-meter very high-resolution imagery from freely available sources such as Google and Bing Maps have fragmented spatial and temporal availability and are not sufficient alone for unbiased visual interpretation (Schepaschenko <i>et al</i>., 2019). Other sources of data that are temporal and spatially complete, such as the archive of Sentinel-2 imagery, should be used together with very high-resolution imagery to assign labels to the samples. Temporal mismatch between the period that is being assessed and the data sources available can cause errors in the sample data classification. The interpreter must ensure that sample data are collected for the time period of interest.

Data sources that can bias the interpretation of the sample should be avoided. Particularly, in the context of accuracy assessment, the change map data should not be known to the interpreter.

If the interpreter still needs any clarification or further information, the visual image interpretation can be performed in combination with auxiliary data (e.g., forest inventory data, historical mapping of land use and land cover, etc.).</td>
</tr>
<tr>
<td>Documentation</td>
<td>The data collected must be carefully documented and archived, such that the reporting process remains transparent.

The technical document ("Passo a Passo para o Levantamento e Estimativa de Emissões do Sector de Uso da Terra, Mudanças do Uso da Terra e Florestas" provides further details of the data collection process such as, the tools used for data collection, data sources, etc. Challenges and limitations should be recorded, as well as potential sources of bias during the data manager collection, using a productivity control electronic spreadsheet (see the structure in the Table 4) shared by the data manager via Google Sheets and/or the "Commentaries" field displayed in the digital forms (See SOP2). Areas for improvement could be suggested.

Table 4 Table structure of the Productivity Control Spreadsheet

<table>
<tr><th colspan="8">Productivity Control Spreadsheet</th></tr>
<tr><td colspan="8">AOI:
Year of monitoring:</td></tr>
<tr><th rowspan="2">Batch No</th><th rowspan="2">Number of samples</th><th rowspan="2">Interpreter Name</th><th colspan="3">Daily Number of Samples Assessed</th><th rowspan="2">Observations and comments</th><th rowspan="2">Progress status (%)</th></tr>
<tr><th>dd-mm-yyyy</th><th>dd-mm-yyyy</th><th>...</th></tr>
<tr><td>1</td><td>100</td><td>First name + Last name</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>
</td>
</tr>
</table>

...	...	...					
Total							

Documentation of the interpreters or institutions involved in the data collection is important for logistics of data collection. It can be used to improve communication with the data collection team by sharing contact information and accountability for the quality of the data collected. A contact sheet (Table 5) should be stored and can be referred to for future data collection activities, to have a consistent group of interpreters over time.

Table 5 Table structure of the Contact sheet

No	Name	Institution	Position	Contact
1	First name + Last name	National Sustainable Development Fund (FNDS)	GIS analyst	email@domain.com
x	xx	xx	xx	xx

Section 3 Quality management

Quality management and response design

Quality management is essential to ensure the integrity of the data collection procedure. For data collection, quality management procedures are established before the data collection and are integrated from the planning phases to the preliminary analysis of the data collected. The quality of analysis hinges on the consistency of the data collected. Various checks need to be built into the data collection procedure to ensure unbiased and high-quality data is being produced, and that data is consistent between different interpreters and from the same interpreter over time.

A detailed and comprehensive interpretation key informs the interpreters what they are identifying in the imagery, as detailed in the Response Design (SOP 2). The interpretation key should cover all the classes in the classification scheme and the heterogeneity of the landscape within the area being assessed. Training is imperative to ensure all interpreters understand the classes in the classification scheme and how to identify them. During the training the interpretation key can be augmented to include additional examples that were not previously covered. The amount of time dedicated to the training depends on the complexity of the classification scheme and the experience of the interpreters.

The response design is formulated to facilitate quality management during the data collection. Information beyond the classification scheme can be built into the response design and logged during the data collection through the registration digital form. Contextual information about how the data was collected, such as the person who interpreted the sample, landscape elements, the date the data was collected, the data sources used and the date of those sources, user confidence of their sample classification, and additional comments for further context, can assist in checking the quality of the data.

Quality control

	Quality management during data collection involves the data collection manager collecting results of the batches database from interpreters. For each batch database, the data collection manager selects all samples interpreted as deforestation, and 20% of the remaining samples unidentified as deforestation applying a proportional stratified random sampling based on deforestation map strata on the batch. These selected samples are submitted to two supervisors well-qualified for the data collection, who perform independent checks on the data set, taking into account the following parameters: i. The percentage of coverage for each element within the sample plot ii. The current land cover/land use class (levels 1 and 2) iii. The land cover/land use change class iv. The former land cover/land use class (levels 1 and 2) v. The date of occurrence of land cover/land use change, or evidence date of remaining land cover/land use. After the independent checks, both supervisors compare the data checks and consensually compile a single comment for each sample. If there is record for at least 20% incorrectly classified samples from the 20% of samples unidentified as deforestation, the interpreter should review the entire batch, otherwise the interpreter review only the incorrectly classified samples, according to supervisory data quality checks. On the other hand, in relation to the samples interpreted as deforestation, the interpreter review only the incorrectly classified samples, according to supervisory data quality checks. In case of no consensus between the supervisors or these with the interpreter, the data manager checks the sample with the group to take a final decision. The supervisory data quality check is cyclical until the interpreter achieves a percentage of incorrectly classified samples less than 20% to as a continuing calibration exercise. Incorrect samples are corrected to provide the highest quality interpretations during the data collection. The interpretation is augmented with new examples reflecting various landscapes and difficult samples to interpret.
Training and calibration	The training procedure requires all interpreters to be present and participating in the exercises. The training is led by the data collection manager, someone with extensive experience in image interpretation and familiarity of how the classification scheme looks on the ground and in satellite imagery. The data collection manager must ensure adequate time spent reviewing and discussing classification system, response design, and image interpretation key. Review or update of the selected software and hardware for the visual interpretation might be necessary to ensure all interpreters all comfortable with technology. If there is a range of skill levels in the group and there are sufficient resources, interpreters can work in groups of two, pairing experienced image interpreters or someone with extensive field experience with novice staff. During the training, it is important to perform example interpretation exercises together with the group and individually. Discussion is facilitated to explain how to access and assess the varying types of data and how to record information for each sample on the digital form. Data collection can strive to achieve consistency by documenting the training procedure and consistently repeating the same procedure in future data collection exercises with the same interpreters. Sample based photo-interpretation of land use and land use change can be a highly subjective exercise. Highly skilled image interpreters may give different responses and

	inconsistent availability of very high resolution imagery can lead to bias in the results. Training for a data collection procedure is a way to calibrate the interpreters to the rules of classification, assessment of the available data sources and range of conditions they might observe over the area of interest.
Data checks	There are different types of quality checks during the data collection that can be used for quality management. Patterson and Mccallum (2020) describe different types of checks that can be used for quality management. The following are used: a) **Self-reviewing** results consists of interpreters quickly reviewing their assigned samples during the data collection exercise. This allows interpreters to control for the quality of their samples and correct any obvious errors they find. After assessing all the samples, interpreters can also self-review samples from the beginning of the assessment to make sure they are consistent with how they classified samples at the end of the assessment, and are in line with the established procedures. This type check is performed using the SAIKU extension of Collect Earth tool and the procedures are described in the technical document "Passo a Passo para o Levantamento e Estimativa de Emissões do Sector de Uso da Terra, Mudanças do Uso da Terra e Florestas". b) **Duplication of interpretation** allocates the 10% of samples of each batch initially mentioned to more than one interpreter. This can allow for the calculation of concordance of interpretations, as well as indicate if there are particular landscape types that consistently produce variable interpretations. c) **The cross-validation process** tests an interpreter results against another set of data that is considered accurate and reliable, using the 10% of samples of each batch initially mentioned. This allows for an external standard that interpreters can rely on to maintain a consistent approach. During the training, interpreters can work on samples that have already been classified or interpreted by an expert (with ground-truthing). These examples might have been part of another related project, created as part of key development, or made specifically for cross-validation. d) **Hot checks** are samples that are flagged as low confidence. These marked plots should be further reviewed by other interpreters, the supervising analyst, and discussed during review meetings. Once reviewed, labels that are deemed to be incorrect on these plots should be adjusted by the interpreter. e) **Cold checks** are all samples interpreted as deforestation and 20% of the remaining samples unidentified as deforestation from the results of the batches database produced by interpreters. The decisions made by the interpreters are reviewed by two independent supervisors or/and a data collection manager or group of interpreters meeting together. If there are incorrectly classified samples by the interpreter, they are submitted to the interpreter for the correction. f) **Logical data checks** use the data collected for each sample to ensure the classification for the plot is logical and possible. Logical checks and limitations can be already built into the response design. Samples that have illogical combinations of sub-sample and sample classifications are rechecked. This type check is performed using the SAIKU extension of Collect Earth tool and the procedures are described in the technical document "Passo a Passo para o Levantamento e

Estimativa de Emissões do Sector de Uso da Terra, Mudanças do Uso da Terra e Florestas”.

g) **Self-checks** are 10% of samples of a batch of an interpreter initially mentioned, which is reviewed and interpreted again by the same interpreter. These will typically be samples from previously completed data collection and are considered blind checks.

h) **Cross-checks** are 10% of samples of a batch database of an interpreter initially mentioned, which is reviewed and interpreted by a different interpreter. These will typically be assigned concurrently to several interpreters and are considered blind checks.

For examples of these data checks and further descriptions see Patterson and Mccallum (2020).

Quality management analysis

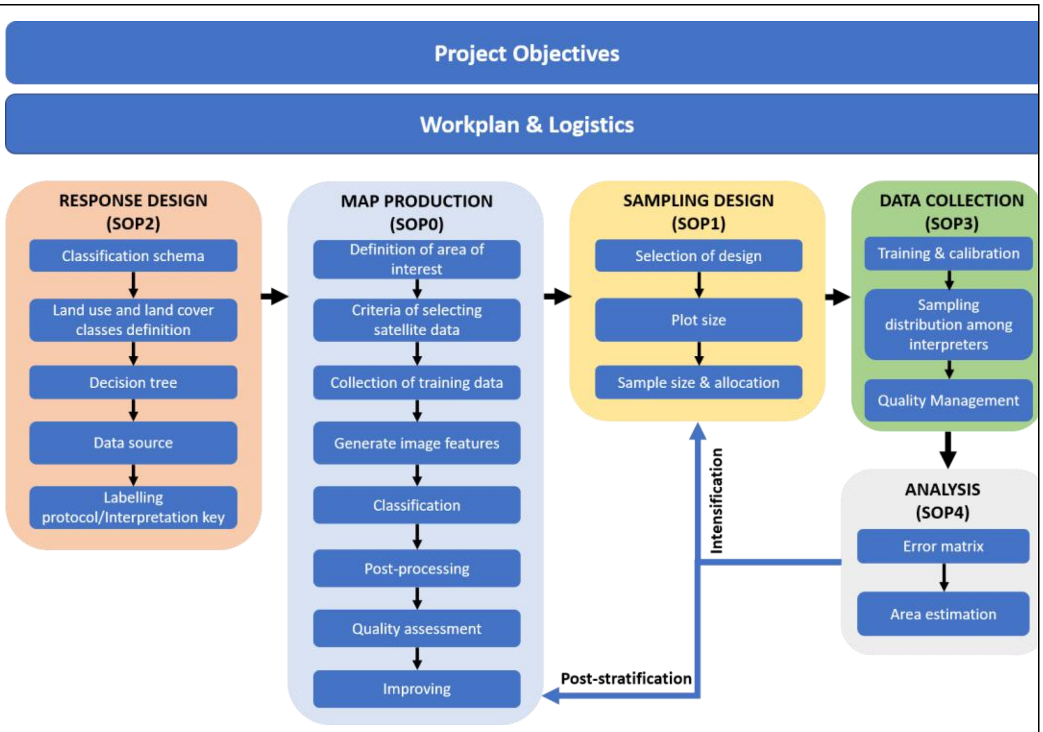
After all the samples have been assessed, the data collection manager compares the duplicated samples, whether they are self-checked or cross-checked samples, to the primary dataset. A summary of intra- and inter- interpreter agreement is reported. Uncertainty associated with variance in the interpreters can be estimated. The method agreement analysis applied is based on the Cohen's Kappa coefficient (Cohen, 1968) and their interpretation proposed by Landis and Koch (1977) (see Table 6). It can be conducted using Microsoft Office Excel and the procedures are described in the technical document “Passo a Passo para o Levantamento e Estimativa de Emissões do Sector de Uso da Terra, Mudanças do Uso da Terra e Florestas”.

Table 6 Interpretations of Cohen's Kappa values (Landis and Koch, 1977).

Cohen's Kappa values	Interpretation
<0.00	poor agreement
0 – 0.20	slight agreement
0.21 – 0.40	fair agreement
0.41 – 0.60	moderate agreement
0.61 – 0.80	substantial agreement
0.81 – 1.00	almost perfect agreement

All disagreements identified should be reviewed to identify and document the error source, and address a specific training for their respective interpreter to become more accurate and consistent.

Interpreters’ errors can be magnified in a stratified sample design (McRoberts *et al*, 2018). It is essential to reduce errors as much as possible, which can be done by following quality management practices.

<table>
<tr>
<td>Data Flow Diagram</td>
<td>


</td>
</tr>
<tr>
<td>References</td>
<td>
Cohen, J. 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin, 70, 213-220.

Landi, J.R & Koch, G.G. 1977. The measurement of observer agreement for categorical data. Biometrics, Vol. 33, No. 1. 159 – 174.

McRoberts, R.E., Stehman, S.V., Liknes, G.C., Næsset, E., Sannier, C., Walters, B.F., 2018. The effects of imperfect reference data on remote sensing-assisted estimators of land cover class proportions. ISPRS J. Photogramm. Remote Sens. 142, 292–300.

Patterson, M.S. and Mccallum, K. 2020. Project Development Manual For Collect Earth Online. U.S. Forest Service (USFS) Geospatial Technology and Applications Center (GTAC).

Schepaschenko, D.; See, L.; Lesiv, M.; Bastin, J.-F.; Mollicone, D.; Tsendbazar, N.-E.; Bastin, L.; McCallum, I.; Laso Bayas, J.C.; Baklanov, A.; Perger, C.; Dürauer, M. and Fritz, S. 2019. Recent advances in forest observation with visual interpretation of very high-resolution imagery. Surv. Geophys. Vol. 40, No 4, 839–862. https://link.springer.com/article/10.1007/s10712-019-09533-z
</td>
</tr>
</table>

SOP 4: ANALYSIS

STANDARD OPERATING PROCEDURE 4 (SOP4): SAMPLE-BASED AREA ESTIMATION ANALYSIS

Section 1 Overview	
Purpose	Derive area estimates and their uncertainties through the combined use of reference data and maps (i.e., sample-based area estimation). Specifically, create an error matrix and construct estimators of area with confidence intervals; and calculate accuracy measures
Scope	Wall-to-wall, geospatial data supported by a reference sample
Responsibilities	Technical decision and policy makers in REDD+ countries and international agencies, multilateral and bilateral programmes, in particular for the estimation of activity data (i.e., land use/land cover change) for emissions/removals in REDD+ activities.
Prerequisites	• Map production (SOP0) • Sampling design (SOP1) • Response design (SOP2) • Data collection (SOP3)
Requirements	• Interpreted reference data for all samples available • Map data for all samples available • Reference and map class labels are the same • Area of all strata in map. • Confidence level for the estimation of uncertainties **Important Note:**
Section 2 Procedure	
The error matrix	The current SOP procedure focuses on the basis that a stratified random sampling has been performed. Yet, it is a general case that reveals the same estimators as those used when one has performed either systematic or simple random sampling (see end of subsection *Estimating areas and their uncertainty*). The aim of the stratified area estimation is to eliminate bias as far as can be judged and reduce uncertainties as far as practicable. To attain both goals, reference data must be compared with map information. The reference data (from a probability sampling design) adjust for estimated systematic classification errors and estimating uncertainty, making use of the previously seen (see SOP1) stratified estimators. The error matrix contrasts reference and map data and constitutes the main tool for the area analysis. First, one can construct a matrix where reference data (sample points class

labels taken as ground truth) are represented in columns and map data (sample point class labels according to the available map) in rows.

Let us assume three change classes in our map and reference data. Although the objective is generalizable to any area estimation, let us also assume that we aim to detect area change. Our $H=3$ classes might be, in this case, *Forest* (stable), (Forest) *Loss* and *Non-Forest* (stable).

Two basic steps in building error matrices are considered:

1. First, an error matrix of sample counts of reference and map labels is built (Table 7). Along the diagonal, elements indicate matching labels (the map correctly describes the ground truth or reference), while off-diagonal elements in the tables below represent map omissions (top down direction) and commissions (left-right direction).
2. An extra column will calculate strata weights dividing the area of each class or stratum by the total area in the reporting land (i.e., country, province,..)

Table 7 *Error matrix of sample counts.*

Map data	Reference data			Area map	Strata weight (w_h)
	Deforestation	Stable forest	Stable non-forest		
High probability of deforestation	n_{11}	n_{12}	n_{13}	$n_{1.}$	$a_{1.}$
40 m Buffer	n_{21}	n_{22}	n_{23}	$n_{2.}$	$a_{2.}$
Low probability of deforestation	n_{31}	n_{32}	n_{33}	$n_{3.}$	$a_{3.}$
Stable forest	n_{41}	n_{42}	n_{43}	$n_{4.}$	$a_{4.}$
Stable non-forest	n_{51}	n_{52}	n_{53}	$n_{5.}$	$a_{5.}$
New deforestation	n_{61}	n_{62}	n_{63}	$n_{6.}$	$a_{6.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	n	a

3. To translate sample counts into actual estimated area proportions per class, it is necessary to operate on the colored cells. For each cell area proportion is defined as

$$\hat{p}_{hj} = w_h \cdot \frac{n_{hj}}{n_{h\cdot}}$$

where h and j stand for row and column, respectively. Accuracy measurements, though not relevant for the current SOP, can also be calculated from dividing individual cell area proportions by the respective map (user's accuracy) or reference (producer's accuracy) class area proportions.

Table 8 *Error matrix of area proportions.*

	Reference data			Total	User's accuracy ($\hat{U}_i$)
Map data	Deforestation	Stable forest	Stable non-forest		
High probability of deforestation	$\hat{p}_{11}$	$\hat{p}_{12}$	$\hat{p}_{13}$	$\hat{p}_{1\cdot}$	$\hat{p}_{11}/\hat{p}_{1\cdot}$
40 m Buffer	$\hat{p}_{21}$	$\hat{p}_{22}$	$\hat{p}_{23}$	$\hat{p}_{2\cdot}$	$\hat{p}_{22}/\hat{p}_{2\cdot}$
Low probability of deforestation	$\hat{p}_{31}$	$\hat{p}_{32}$	$\hat{p}_{33}$	$\hat{p}_{3\cdot}$	$\hat{p}_{33}/\hat{p}_{3\cdot}$
Stable forest	$\hat{p}_{41}$	$\hat{p}_{42}$	$\hat{p}_{43}$	$\hat{p}_{4\cdot}$	$\hat{p}_{44}/\hat{p}_{4\cdot}$
Stable non-forest	$\hat{p}_{51}$	$\hat{p}_{52}$	$\hat{p}_{53}$	$\hat{p}_{5\cdot}$	$\hat{p}_{55}/\hat{p}_{5\cdot}$
New deforestation	$\hat{p}_{61}$	$\hat{p}_{62}$	$\hat{p}_{63}$	$\hat{p}_{6\cdot}$	$\hat{p}_{66}/\hat{p}_{6\cdot}$
Total	$\boldsymbol{\hat{p}_{\cdot 1}}$	$\boldsymbol{\hat{p}_{\cdot 2}}$	$\boldsymbol{\hat{p}_{\cdot 3}}$	1	
Producer's accuracy (P_i)	$\hat{p}_{11}/\hat{p}_{\cdot 1}$	$\hat{p}_{22}/\hat{p}_{\cdot 2}$	$\hat{p}_{33}/\hat{p}_{\cdot 3}$		Overall accuracy ($\hat{O}$) $= \hat{p}_{11} + \hat{p}_{22} + \hat{p}_{33}$

Estimating areas and their uncertainty

The mean estimator for the area of each class can be directly obtained from Table 8. While one would be tempted to use map class area proportions ($\hat{p}_{h\cdot}$), unbiased stratified estimators are provided using reference class area proportions ($\hat{p}_{\cdot j}$). As implied in Table 8, these follow:

$$\hat{p}_{\cdot j} = \sum_{h=1}^{H} w_h \cdot \frac{n_{hj}}{n_{h\cdot}} = \sum_{h=1}^{H} \hat{p}_{hj}$$

Once the estimated reference class area proportions are obtained, the mean total area per class is calculated by multiplying them with the total reporting area a:

$$\hat{A}_j = \hat{p}_{\cdot j} \cdot a$$

Uncertainties in activity data were derived using non-parametric bootstrapping, where reference data points were re-sampled with replacement 100,000 times. For each permutation of reference data points, the bias-corrected area estimates were produced following the methods described in Olofsson *et al*. (2014). Uncertainty was estimated from the resulting distribution of area estimates.

Although more complex to implement, bootstrapping has the advantages of not requiring any assumption about the shape of the probability distribution function of each land cover transition class, and avoids the generation of negative areas in rare classes where a probability distribution function crosses zero. The method was implemented in R, and the scripts used are available in the "*Mozambique ERPA 2018*" shared folder.

The impact of using non-parametric bootstrapping to estimate uncertainties vs other methods was tested with a comparison of deforested areas derived from bootstrapping against sampling from a normal distribution with standard error calculated with the methods described in Olofsson *et al*. (2014). For the latter case two uncertainties were derived: one retaining any negative area estimates for rare transition classes, and another setting these to zero. The result (Figure SOP4**.22**) indicates that there is very little difference between any of the methods in either reference or monitoring periods, with the result that any chosen approach would produce equivalent emissions estimates.

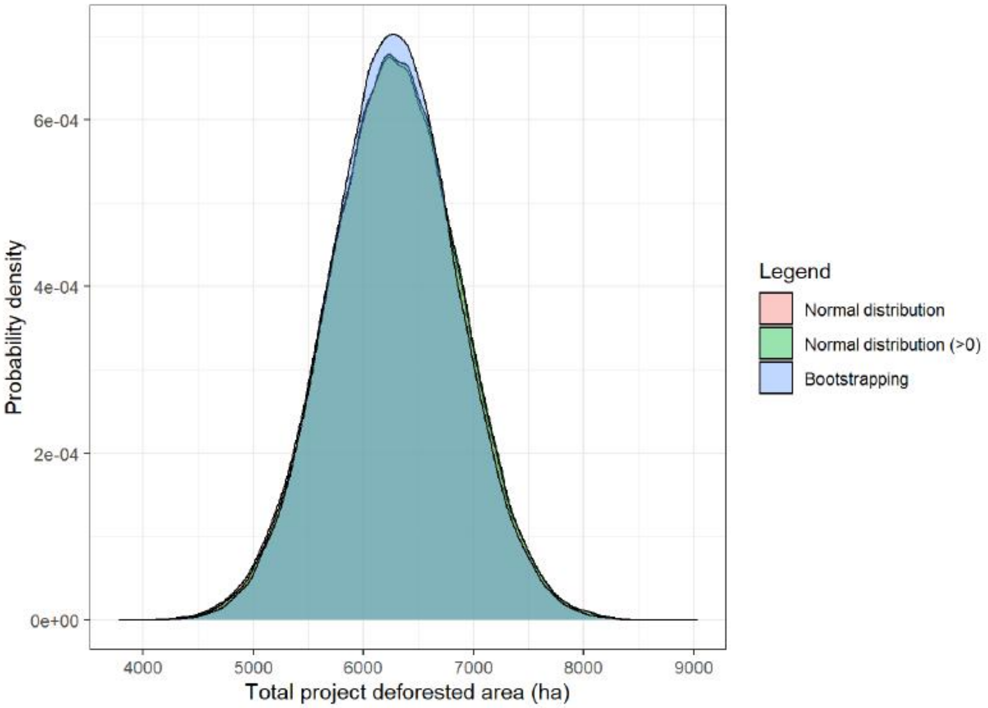
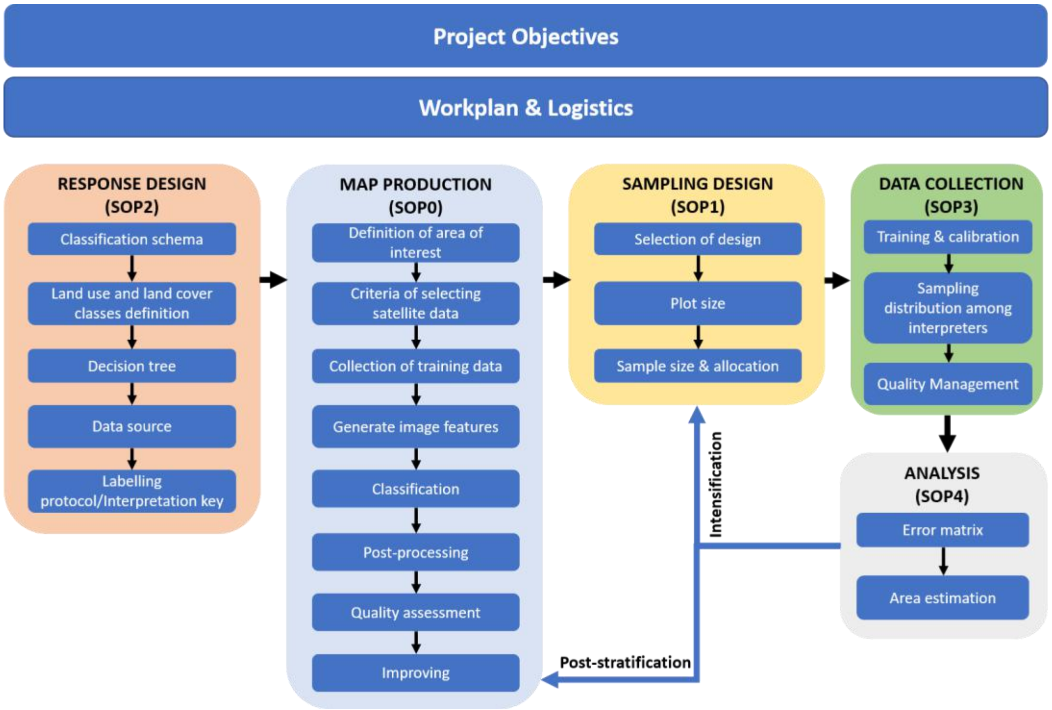


Figure SOP4.22: Total activity data area estimates for monitoring period using normal distributions for each transition class (red), normal distributions with a minimum area of 0 ha (green), and non-parametric bootstrapping (blue). All three methods result in equivalent uncertainty estimates.

Section 3 Quality management

Error matrix

- The error matrix, particularly that for sample counts, is usually an effective tool to quickly detect the overall quality of the data collection process. High numbers in off-diagonal elements (commission, or mostly, omission errors), when compared to those in the diagonal, may reflect a poor number of samples overall, imply the

<table>
<tr><td></td><td>need for a revision in data interpretation procedures or response design, or even reflect a poor map quality that may involve a new processing of the map. Often the people responsible may need to go back to review procedures. Although this may theoretically produce biases, it is often a necessary consequence derived by lack of expertise or knowledge in the specific area assessed.
• Report the error matrix in terms of estimated area proportions.</td></tr>
<tr><td>Area and uncertainty estimation</td><td>Set a maximum permissible uncertainty.</td></tr>
<tr><td>Data Flow Diagram</td><td>


</td></tr>
<tr><td>References</td><td>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 Sensing of Environment, 148:42–57.</td></tr>
</table>

For more information, please visit:

- Mozambique MRV website: https://www.fnds.gov.mz/mrv/

Please cite as: Fundo Nacional do Desenvolvimento Sustentável (FNDS). Unidade de MRV. 2020. Standard Operating Procedure for Sample-Based Area Estimation. 61. Maputo. Mozambique.