EU Support to the National Land Audit in Zambia

Contract No. EDF/ZM/2023/448-937

Output 1.2: Producing current Land Cover / Land Use maps at national scale

Activity 1.2.1: Land cover mapping

National Land Cover Mapping in Zambia – 2024

- Guidelines for image processing, analysis, and interpretation/validation
- Protocol for geometric and thematic accuracy assessment of the interpretation results
- Main results achieved and products available

Version 5: 14/02/2025

1 EXECUTIVE SUMMARY 5

2 INTRODUCTION 6

3 OBJECTIVES AND EXPECTED OUTPUTS 7

3.1 Main objective 7

3.2 Specific activities 7

3.3 Expected outputs 7

4 METHODOLOGY 8

4.1 Land cover legend 8

4.1.1 Definition of the classes 8

4.1.2 Challenges on the legend 9

4.2 Data preparation and preprocessing 9

4.2.1 Selection of the most suitable time frame 10

4.2.2 Exploration of available focus satellite data (Copernicus Sentinel-2) and auxiliary data for land cover mapping 11

4.2.3 Collection of input data for training and testing/validation of the results: 12

4.2.4 Collection of input data and validation for year 2017 15

4.2.5 S2 data pre-processing 15

4.2.6 Feature engineering – creation of input bands 15

4.2.7 Selection of the best-performing and state-of-the-art algorithms for classification and data post-processing 16

4.2.8 Post-classification filter 16

4.2.9 Enhancements in training and validation data after several rounds of statistical and visual inspections. 17

4.2.10 Enhancement with a multi-temporal approach 17

4.2.11 Enhancement with a one-leave-out permutation feature importance analysis 17

4.3 Geometric and thematic accuracy assessment 18

4.4 Export of the final map of land cover 20

4.5 Selected Methodology for Land Cover Change 2017-2023 21

5 TIMELINE AND MILESTONES 23

6 MAIN RESULTS ACHIEVED 24

6.1 Guidelines for processing and classification 24

6.1.1 The trained Random Forest model 24

6.1.2 Land cover legend 25
6.2 Validation report 25
6.3 National land cover map for 2023 and 2017 25
6.4 Land cover change map 28
6.5 Land cover change vs. documented deforestation 29
6.5.1 Overview 29
6.5.2 Forest Land Cover Mapping Efforts 29
6.5.3 Documented deforestation 29
6.5.4 Classification challenges 33
6.6 Capacity building in satellite data utilization using cloud-based platforms and in land cover mapping 34
7 NEXT STEPS 35
8 CONCLUSION 36
9 REFERENCES 37

LIST OF TABLES

Table 1: Definition of Land Cover classes 8
Table 2: Validation report for the National Land Cover of Zambia for 2023 19
Table 3: Validation report for the National Land Cover of Zambia for 2017 19
Table 4. Land cover composition of Zambia in 2023 based on classification results 27

LIST OF FIGURES

Figure 1: LC legend and colours 9
Figure 2: Mean precipitation and mean temperature in 2023 (above) and 2017 (below) in Zambia based on ERA 5-Land data with the identified rainy/wet and dry periods 10
Figure 3: The Normalised Difference Vegetation Index (NDVI) over Zambia in the wet period (left) and dry period (right) in 2017. Note: Some areas, specifically in the eastern and central part, are masked out, due to unavailability of cloud-free measurements during the selected wet season. 11
Figure 4: The NDVI over Zambia in the wet period (left) and dry period (right) in 2023 11
Figure 5: Available S2 image tiles over Zambia (116 in total) 12
Figure 6: Examples of the selected 8 land cover types (1) 14
Figure 8: Spatial distribution of training and testing/validation datasets in Zambia for 2023 15

Figure 9: All input features (left) and selected features after the permutation feature importance analysis (right). 18

Figure 10: Timeline of the National Land Cover mapping in Zambia 23

Figure 11 : General overview of the methodology 24

Figure 12. The final National Land Cover map for 2017 (zoomed in to an area around Kabwe city) 25

Figure 13. The final National Land Cover map for 2023 (zoomed in to an area around Kabwe city) 26

Figure 13. and Fig. 14. Land cover composition in Zambia in 2023; Land cover composition in Zambia in 2017. 27

Figure 15. An example of land cover change detection results (central coordinates of -14.647, 29.346). 28

1 EXECUTIVE SUMMARY

The output 1.2.1 of the EU support to the National Land Audit project aimed to produce up-to-date national land cover maps of Zambia for 2023 and 2017, as well as a land cover change map between these two years, using modern geospatial methodologies, specifically focusing on Copernicus Sentinel-2 satellite data. Year 2017 was selected as the first year that fully covered the entire country, while 2023 was selected as the last year that had all-year data. The selected land cover legend with 8 classes (forest, shrubland, grassland, cropland, built-up, bare land, water, and wetlands) is in line with other ongoing projects (e.g., EU funded" Sustainable Landscape through Integrated Management" (SLIM) project) and meets internationally accepted standards, allowing for cross-country comparisons. Various datasets were used as input data alongside the Sentinel-2 satellite data, such as Copernicus Digital Elevation Model, vegetation height information, built-up surface data from JRC, while training and validation data were collected using high-resolution imagery. Advanced machine learning algorithms were tested for land cover classification using multi-temporal approaches (considering the high variability in vegetation response due to seasonal changes) to enhance the classification results.

The produced national land cover maps for 2023 and 2017 achieved high overall accuracies (around 85%). The 2023 map showed that forest areas dominate the landscape (38.6%), followed by shrubland (34.8%), grassland (11.8%), and cropland (9.4%).

A three-day intensive training on cloud-computing, big data analysis and land cover mapping using Google Earth Engine was provided for local stakeholders and described in a parallel report. Training materials were prepared as well, covering essential workflows for data access, pre-processing, classification, time series analysis etc., laying the groundwork for capacity building. In the first quarter of 2025 (Q1) is expected to organize a second round of training activities, preferably online to allow attend a higher number of participants outside of Lusaka.

The main limitations of the land cover maps include the challenges in accurately distinguishing certain land cover types and the need for further validation by local stakeholders. Future work should focus on addressing these issues, potentially through validation of current results and input data, identification of omission and commission errors, as well as through the collection of additional ground-truth data in close cooperation with local stakeholders. These steps are necessary for improvement of the current results. The validation of results and input by stakeholders is expected until the end of 2024, while the subsequent improvement is expected in the Q1 of 2025.

2 INTRODUCTION

The EU support to the National Land Audit project in Zambia (NLA) aims at enhancing land administration and management across the country. This project addresses several fundamental issues related to land use and land cover, which are essential for sustainable development, environmental protection, and effective resource management. The primary focus of the project includes four key areas: i) improving the National Spatial Data Infrastructure (NSDI) by updating it with publicly available and accessible information, ii)producing comprehensive national-level land cover and land use maps, iii) carrying out 20 district land audits and 3 pilot audits, making key results of the NLA accessible and testing and validating methodologies for the NLA; iv) and providing guidance for a full implementation of the NLA along with a post-audit action plan.

This assignment is a crucial contribution to produce a current national-scale land use/land cover map. It directly addresses the audit question regarding the area and location of various land uses across the country. These maps are also crucial for understanding and managing the country's land resources effectively. The project has adopted the FAO Land Cover Classification System (LCCS) and the uses Google Earth Engine (GEE), a cloud-based platform developed primarily for Remote Sensing data analysis, for satellite data classification and processing ensure that this assignment is on the right track. The scope of work involves development of algorithms and pipelines in GEE to produce all national-level maps. This approach mirrors the methodology used in the SLIM project, ensuring consistency and reliability.

3 OBJECTIVES AND EXPECTED OUTPUTS

3.1 Main objective

The main objective of this assignment is to produce an up-to-date national land cover map of Zambia for 2023, as well as a land cover map for 2017, along with a land cover change map, which meets international standards, using modern and state-of-the-art geospatial methodologies of Earth Observation (EO) data processing, specifically focusing on Sentinel-2 Copernicus data.

3.2 Specific activities

1. definition of the land cover legend,
2. identification of the most suitable time frame for image mosaic creation,
3. exploration of available focus satellite data (Copernicus Sentinel-2) and auxiliary data for land cover mapping
4. collection of input data for training and testing/validation of the results,
5. selection of the best-performing and state-of-the-art algorithms for data pre-processing, classification and data validation,
6. selection of the best input features and algorithm parameters for the classification,
7. testing of a mono-temporal and a multi-temporal approach for land cover classification,
8. developing dynamic algorithms coded in JavaScript language, adaptable to various datasets and input parameters,
9. utilising GEE to generate national-level land cover maps for 2017 and 2023,
10. creation of a land cover change dataset, both in vector and raster format,
11. enhancing results with additional field validation data,
12. performing accuracy assessments of the produced maps using available validation points,
13. preparing training materials for data pre-processing, classification and validation
14. providing training for stakeholders

3.3 Expected outputs

The expected results from this assignment include: 1), **guidelines for image processing, analysis, and interpretation/validation** in GEE, which will be essential for land cover and land use mapping; 2) a **protocol for assessing the accuracy of the land cover and land use maps**; 3) **national maps of land cover and land use** that will be produced in standard formats with accepted metadata; and 4) Finally, **capacity development for technical staff from partner organisations**, including training materials on satellite data pre-processing, classification and validation for land cover and land use maps.

4 METHODOLOGY

4.1 Land cover legend

4.1.1 Definition of the classes

The legend, consisting of 8 major land cover classes, was already defined in the SLIM project, presented and validated at a technical meeting between the MLNR, MGEE, NRSC, SLIM and FAO. The land cover legend classes with a detailed description of each class are defined in Table 1.

Table 1: Definition of Land Cover classes

Land cover Class	Description
Tree cover	This class encompasses regions primarily covered by trees tall at least 5 m, where the tree canopy covers at least 50% of the area. Beneath the canopy, other types of land cover such as underbrush, shrubs, herbs, urban areas and permanent bodies of water may exist, though at a density less than that of the trees. This classification also incorporates areas that have been reforested or are used for plantation agriculture, including but not limited to crops like oil palm and olive trees.
Shrubland	This class includes any geographic area dominated by natural shrubs having a cover of 50% or more. Shrubs are defined as woody perennial plants with persistent and woody stems and without any defined main stem being less than 5 m tall. Trees can be present in scattered form if their cover is less than 30%. Herbaceous plants can also be present at any density. The shrub foliage can be either evergreen or deciduous.
Grassland	This class encompasses geographic regions primarily covered by natural herbaceous vegetation (including grasslands, prairies, steppes, savannahs, and pastures) that occupy 50% or more of the area, regardless of human or animal activities like grazing or managed burning. Woody vegetation such as trees and/or shrubs may exist, provided their coverage does not exceed 30%. This class also includes areas of cropland that are not harvested or are left with bare soil during the reference year, assuming they remain uncultivated.
Cropland	This class pertains to land areas utilised for annual crops, which are sown or planted and then harvested at least once within the 12 months following their sowing or planting date. Such croplands yield an herbaceous covering and may occasionally include some tree or woody vegetation. It's important to note that perennial woody crops are categorised according to the relevant tree cover or shrub classifications.
Built-up	This class is defined by land areas predominantly occupied by man-made structures, including buildings, roads and railroads, among others. It encompasses both residential and industrial buildings. Notably, urban green spaces such as parks and sports facilities are excluded from this classification. Additionally, areas

	designated for waste dumps and extraction sites are treated as bare land within this context.
Bare Land	This class refers to areas characterised by exposed soil, sand or rocks, where vegetation never covers more than 10% of the land at any point throughout the year.
Water bodies	This class encompasses any geographic region predominantly covered by water bodies, such as lakes, reservoirs and rivers, for the majority of the year (over 9 months).
Wetlands	This class pertains to land areas predominantly covered by natural herbaceous vegetation (with a coverage of 50% or more) that is consistently or frequently inundated by water. It specifically excludes areas of unvegetated sediment and swamp forests, which are categorised under open and closed forest classifications.

Proposal: using the internationally accepted Land Cover Classification System (LCCS) developed by UN FAO for global land cover mapping. It provides a systematic framework for classifying land cover based on the integration of biophysical, socioeconomic, and management criteria. LCCS is internationally accepted and provides comparability across different regions and projects. This method employs a consistent and standardised methodology for land cover mapping, ensuring coherence and comparability across different regions and time periods.



Figure 1: LC legend and colours

4.1.2 Challenges on the legend

The main challenges are the exact definition of the percentual share of the land cover class, e.g., percentage of canopy cover by forests, percentual coverage by shrubs. In the current state, all these classes are set to a minimum 50% threshold.

4.2 Data preparation and preprocessing

Two base years were selected for national land cover mapping in Zambia: 2023 as the latest full year covered by Sentinel-2 images and 2017, as the very first year fully covered by Sentinel-2 images. For the land cover change analysis, the year 2017 was selected. The Sentinel-2 mission started its

acquisitions in mid-2016. However, due to the availability of only Level-1C (Top-of-atmosphere - TOA reflectance) data for both 2017 and 2023 land cover maps the Level-1C data was used in the analysis to make them comparable.

4.2.1 Selection of the most suitable time frame

Based on the fact that different climatic conditions throughout the year in Zambia affects the spectral response in vegetation. Based on recommendations on the best differentiation of vegetation for Zambia in a EU forestry report by Pohjonen (2004), the dry season was selected for testing the mono-temporal approach.
Based on the fact that some land cover classes have great variations over the year, which proposes to test a multi-temporal approach using band composites from different parts of the year, i.e. to create and use separate input bands for dry and rainy seasons.
For the multi-temporal approach, separate image composites were created for wet and dry seasons. The selection of the most suitable time frame was driven by several factors:

1. First, the definition of wet and dry seasons was conducted using time series analysis of precipitation data using ERA5-Land data for both 2016-2017 and 2022-2023 over the entire Zambia (Fig. 2).

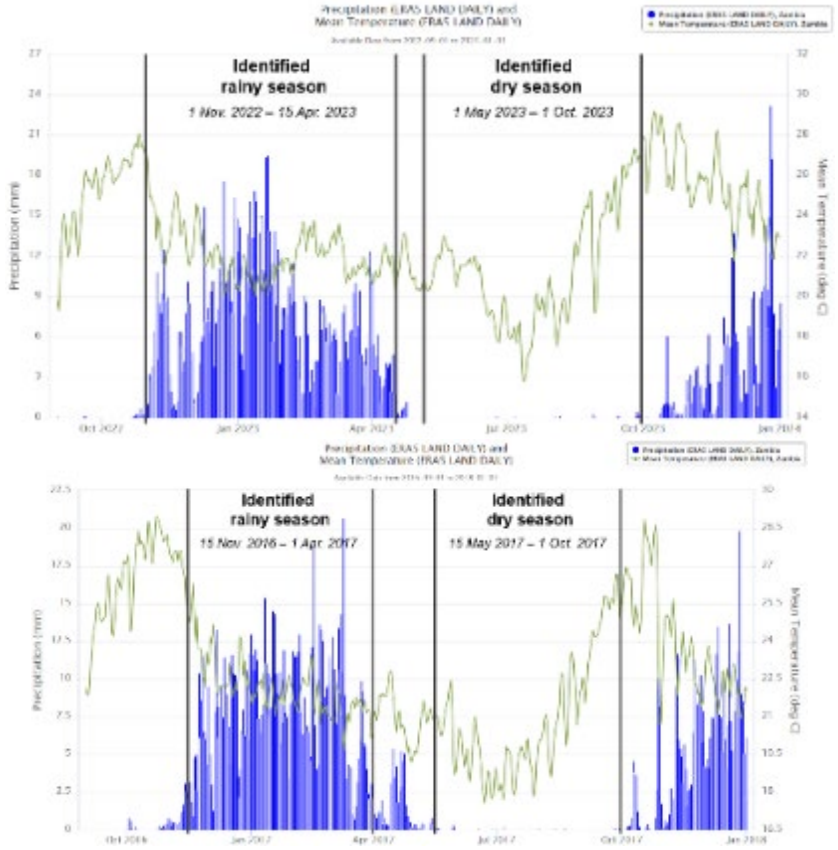


Figure 2: Mean precipitation and mean temperature in 2023 (above) and 2017 (below) in Zambia based on ERA 5-Land data with the identified rainy/wet and dry periods

2. The selected dates for rainy and dry seasons based on precipitation data needed further adjustments, taking into account the shifted vegetation response after the start/end of the rainy and dry seasons.
3. There was a need to select that time frame where the vegetation response still represents that period (wet or dry), while, at the same time, most of the country is covered by cloudless satellite imagery (Fig. 3 and Fig. 4 on the example of the NDVI).

Based on the considerations above, the following dates were selected for both 2017 and 2023:

- **Wet season:** January 1st - May 15th
- **Dry season:** June 15th - October 1st

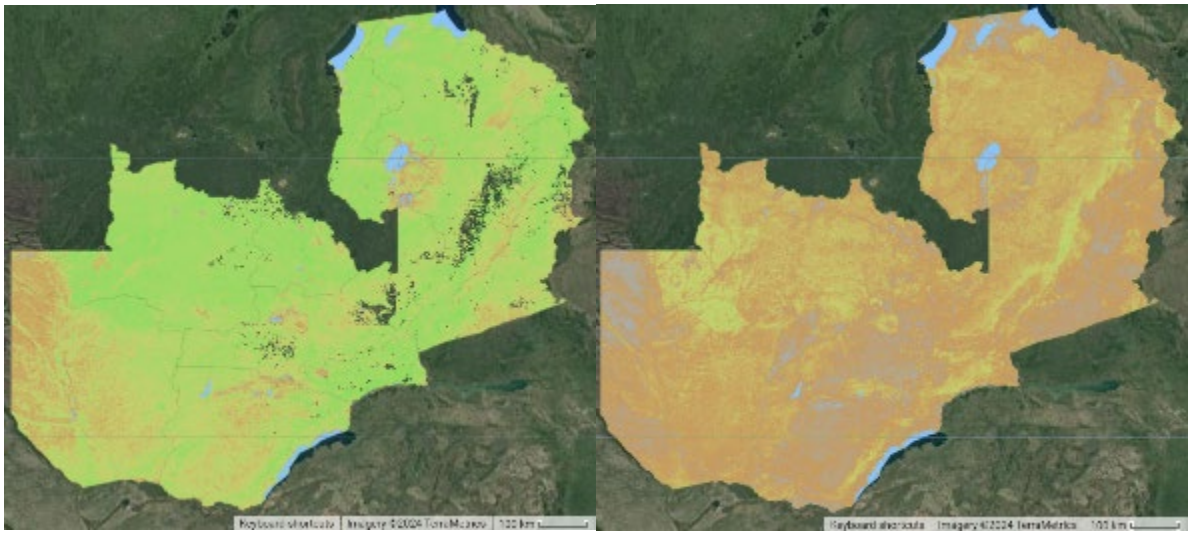


Figure 3: The Normalised Difference Vegetation Index (NDVI) over Zambia in the wet period (left) and dry period (right) in 2017. Note: Some areas, specifically in the eastern and central part, are masked out, due to unavailability of cloud-free measurements during the selected wet season.

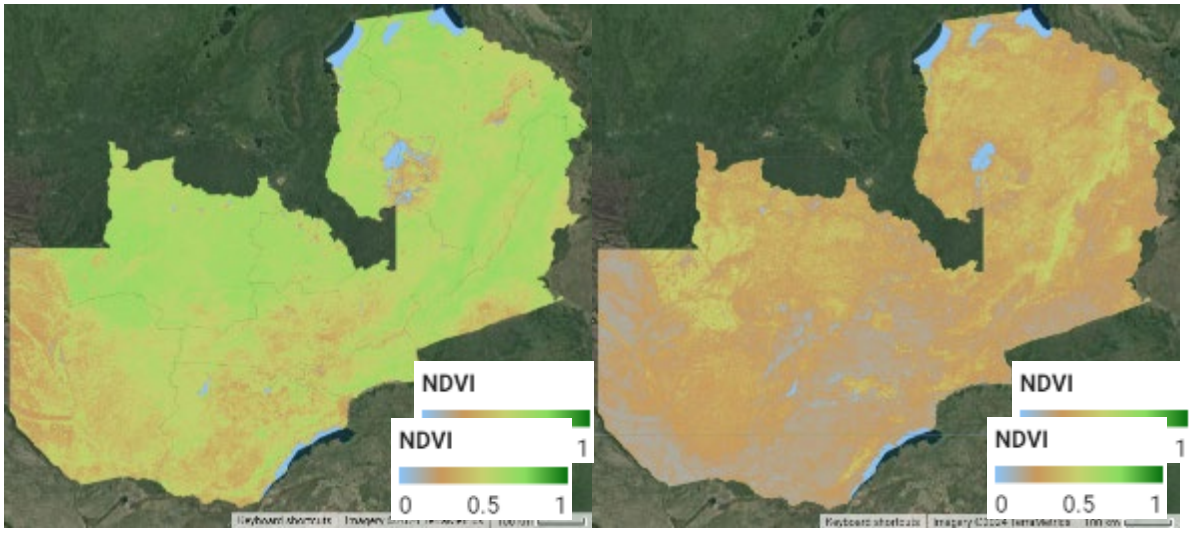


Figure 4: The NDVI over Zambia in the wet period (left) and dry period (right) in 2023

4.2.2 Exploration of available focus satellite data (Copernicus Sentinel-2) and auxiliary data for land cover mapping

The availability of Sentinel-2 Surface Reflectance data (S2) was explored in this step. For the latest year, for 2023, in total of 5235 Sentinel-2 images were found for the selected time range (May-

September 2023) for the entire Zambia covering more than 750 thousand of km^2. In total of 116 different S2 image tiles were detected over Zambia (Fig. 5).

Other auxiliary data were explored in the official GEE Data Catalog available at https://developers.google.com/earth-engine/datasets and the Awesome GEE Community Catalog available at https://gee-community-catalog.org/, including:

- Copernicus Digital Elevation Model (DEM) data, which was found to be the best performing DEM in current studies (Bielski et al. 2024, Gonzales-Moradas et al. 2023, Marešová et al. 2021, Li et al. 2022)
- Global Forest Canopy Height for year 2019 representing vegetation height information generated from a fusion of GEDI and Landsat data in a 30 m spatial resolution and height accuracy of mean absolute error of 6.36 m when compared to airborne lidar data. The tree height ranges from 1 to 30 metres, where the measurements are saturated (Potapov et al., 2021).
- JRC Global built-up surface dataset in 10 m spatial resolution for 2018. This raster dataset depicts the distribution of built-up surfaces, expressed in square metres per 10 m grid cell, for 2018 as observed from the S2 image data. Pesaresi et al. (2023).

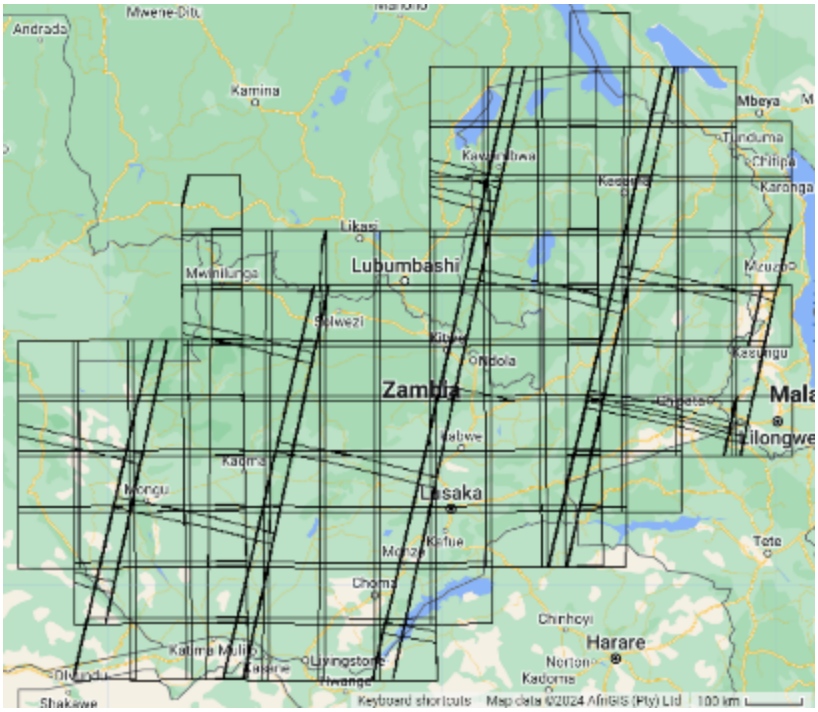


Figure 5: Available S2 image tiles over Zambia (116 in total)

4.2.3 Collection of input data for training and testing/validation of the results:

1. step - Points were generated using a stratified sampling method, based on the proportion of land cover types in ESA WorldCover as of 2021 for the entire Zambia

- ESA WorldCover 2021 was found to be one of the best performing land cover dataset for Sub-Saharan Africa (Kerner et al. 2024)

A threshold of 100 was set for areas that would not reach this number based on these proportions

The first step resulted in ~2200 points with a 20x20 m area

2. step - Assignment of land cover classes to generated areas was performed using high-resolution satellite imagery (Maxar and Airbus at 0.5 - 3 m spatial resolution) in Google Earth Pro, mostly using images from 2023 or 2023, while these images are available with a time series from ~2005 until 2024 (the availability of images varies by location)

Areas on the borders of land cover types and problematic ones e.g. with no coverage by recent images, were excluded.

The second step resulted in ~1400 validated input areas. Examples of land cover classes can be seen in Fig. 6 and Fig. 7.

3. step - Assessment of land cover types with low representation and their extension with samples from SLIM data: bare land, built-up and wetlands were used from the SLIM project.

Bare land - added 23 bare land areas from Lower Kafue (SLIM input data)

Built-up - added 53 built-up areas from Lower Kafue (SLIM input data)
Wetlands - added 100 wetlands from Lower Kafue (SLIM input data)

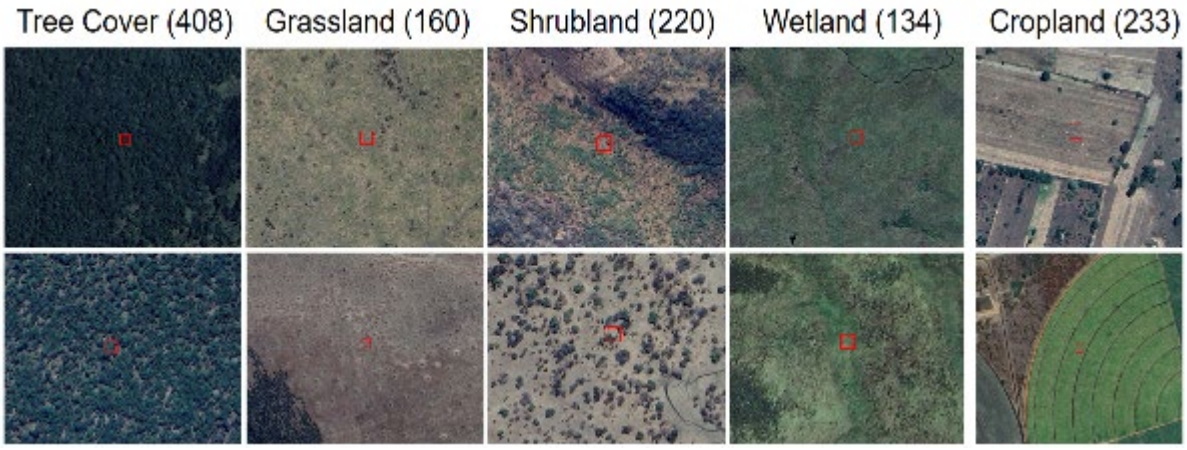


Figure 6: Examples of the selected 8 land cover types (1)

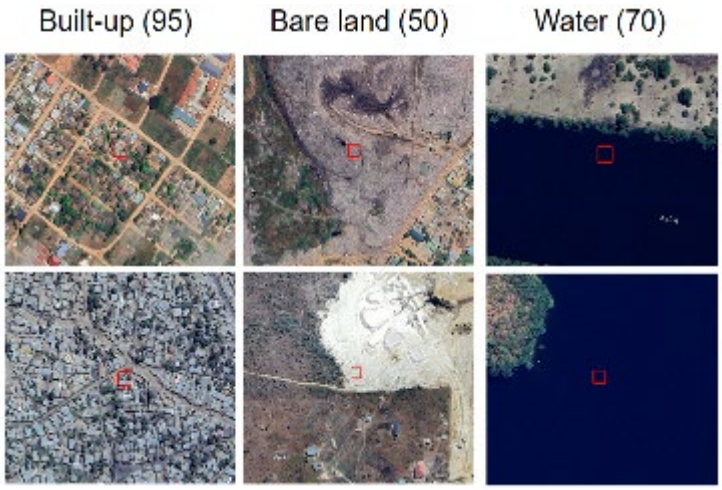


Figure 7: Examples of the selected 8 land cover types (2)

4. step - Input data exploration analysis - statistical analysis of input data in a Python Jupyter Notebook environment. An identification of outliers (using Tukey's fences) in input data based on vegetation/spectral indices (NDVI, NDMI, EVI, NDWI, NDBI) and vegetation height was performed. This step resulted in re-assignment of problematic outliers to other land cover classes or exclusion of outliers.
This step was iterated until all reasonable outliers were re-assigned or excluded from the input dataset.

5. step - Training and validation data creation: The validated data were divided into training (70% of each class) and into testing datasets (30% of each class). The spatial distribution of input training and validation dataset can be seen in Fig. 8.

Challenges in general:

- Tree covered areas with <5m height detected based on the Global canopy height dataset;
- Irrigated croplands have totally different spectral characteristics compared to regular croplands in Zambia. Their spectral characteristics are more similar to wetlands. Therefore, a separate class of irrigated croplands were created. These data were input to the classifiers.

- Overlay of spectral response of grasslands and shrublands can cause misclassifications.

Next steps:
Validation of the created input dataset by stakeholders and project partners is recommended, focusing on problematic areas and classes.

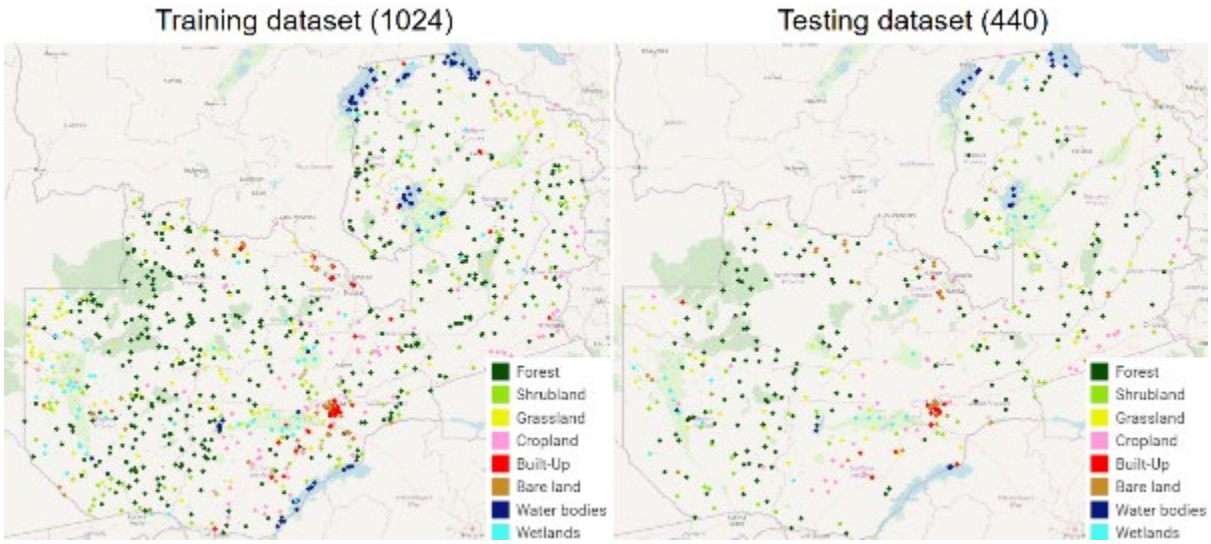


Figure 8: Spatial distribution of training and testing/validation datasets in Zambia for 2023

4.2.4 Collection of input data and validation for year 2017

For the 2017 Land Cover, the same training and testing input set was adjusted to match the state of land cover classes in that year. The validation was performed using high resolution satellite imagery in Google Earth Pro, followed by designed steps in chapter 3.2.3, similarly as for year 2023.

4.2.5 S2 data pre-processing

The original temporally and spatially filtered image collection of 5235 S2 images were further filtered by metadata. The cloud coverage metadata information was used to filter out S2 image tiles having higher than 30% cloud coverage. This filtered in 4811 S2 images for the May-September 2023 period over Zambia.

On top of metadata filtering, the state-of-the-art cloud masking approach, the CloudScore+ introduced at the end of 2023 (Pasquarella et al., 2023 and Pasquarella 2024), was applied to exclude cloud-affected areas on the remaining images.

4.2.6 Feature engineering – creation of input bands

A multi-modal (multi-source) approach was utilised in this project including not only S2 satellite data and its spectral indices, but also DEM data and other globally available auxiliary data, such as

vegetation height or built-up areas, extended by textural and multi-temporal time series features. The final selection of input features included:

- Topographic data (elevation and slope from the Copernicus DEM)
- Optical spectral/vegetation indices (NDVI, EVI, NDMI, NDWI, NDBI)
- Vegetation height information - Global Forest Canopy Height from GEDI & Landsat – Potapov et al. (2021).
- JRC Global built-up surface dataset in 10 m spatial resolution for 2018. This raster dataset depicts the distribution of built-up surfaces, expressed in square metres per 10 m grid cell, for 2018 as observed from the S2 image data. Pesaresi et al. (2023).
- multi-temporal time series features - calculated time series variance of NDVI, NDMI, NDWI, EVI and NDBI for each pixel
- GLCM textural features - sum average (SAVG) with a 5x5 kernel window size.

4.2.7 Selection of the best-performing and state-of-the-art algorithms for classification and data post-processing

Two machine learning algorithms, Random Forest and Gradient Boosting, were tested for classifications. Random Forest efficiently handles high dimensional data, requires few tuning parameters, discovers complex patterns, and yields highly accurate results (Shafizadeh-Moghadam et al. 2020). Random Forest was used also in the SLIM project and is currently one of the most used and best performing algorithms in GEE (Tamiminia et al. 2020, Waśniewski et al. 2020). On the other hand, Gradient Boosting is a newer algorithm with potential higher accuracy, better error handling, therefore it is considered to be one of the state-of-the-art algorithms for machine learning tasks (Grinsztajn, Oyallon, and Varoquaux 2022).

Various hyperparameter combinations were tested for the Random Forest, while for Gradient Boosting, the number of trees parameter was crucial in accuracy improvement. Both the accuracy of individual classifications (2017 and 2023) and the temporal consistency between the two classifications were important factors in selection of the best performing algorithm. At the end, the Random Forest algorithm with 150 trees was selected, as it yielded to high accuracies, and at the same time, comparable spatial results for 2017 and 2023 classifications, ensuring high temporal consistency compared to Gradient Boosting. Another advantage of the Random Forest algorithm is that in GEE it is possible to save the trained model. Although the Gradient Boosting algorithm achieved higher accuracies, the temporal consistency between the two time-steps was very low, mainly because it was not possible to transfer the trained model.

At the end, the trained Random Forest model on 2023 data was used to classify both the 2017 and 2023 land cover maps. The trained Random Forest model used in this work is available in GEE: https://code.earthengine.google.com/?asset=projects/danielp-cuni/assets/NLA-Zambia/RF_classifier

4.2.8 Post-classification filter

Application of a post-classification majority filter to exclude “salt-and-pepper” effect, specifically to exclude areas less than 1 ha (10x10 S2 pixel area).

4.2.9 Enhancements in training and validation data after several rounds of statistical and visual inspections.

The problematic classes, with low per-class accuracies were further inspected and re-validated using a time series of high-resolution satellite imagery in GEE. By an iterative process, the differentiation between tree cover, shrublands and grasslands was improved. Moreover, a new land cover subclass was created for irrigated croplands due to their differences from regular croplands. For this reason, 100 irrigated cropland areas were included in the input dataset. It helped with a clear differentiation between croplands, irrigated croplands and wetlands. At the end, regular croplands and irrigated croplands were merged into one class, into cropland land cover class.

4.2.10 Enhancement with a multi-temporal approach

To further improve the accuracy, a multi-temporal classification approach (described in section 2 of this chapter). The dry season was defined as described above – from 1st May to 30th September. There were two candidate periods for the wet season: the wet season starting at the end of 2022 and ending at the beginning of 2023 and the wet season starting at the end of 2023 and ending at the beginning of 2024. Two main considerations were taken into account: 1) the availability of high-resolution images (there are more available images for the state of the landscape for the earlier wet season) and 2) amount of rain (based on the rainfall analysis in Zambia, there was significantly higher mean rainfall over Zambia in the earlier wet season, see Fig. 1). For these two reasons, the wet season used in this study was temporally defined as the period between the 1st November 2022 and 31st April 2023.

In this step, spectral bands, vegetation/spectral indices were calculated for each season separately. On the other hand, the time series bands representing variance in spectral indices were calculated only once – for the entire selected time period.

4.2.11 Enhancement with a one-leave-out permutation feature importance analysis

To further improve the accuracy and use only the most important input bands (features), a Leave-One-Out Permutation Feature Importance Analysis was used. It measures the contribution of each feature (input band) to the overall performance of the model. This analysis aimed to explore how input features might affect the results. The Leave-One-Out method refers to leaving the worst performing feature out from the next step and assessing the accuracy again. This iterative workflow ran until the overall accuracy and f-scores of individual classes were increasing. On the left part of the Fig. 8, all the 38 input features can be seen, including features from wet and dry season, multi-temporal features and other ancillary features. On the right hand side of Fig. 8, the final selection of 23 input features can be seen after the permutation feature importance analysis. It should be noted that the 5x5 SAVG textural feature of the canopy height dataset is the most important driver of the classification, followed by the minimum composite of the NDWI, the Blue band (B2) from the wet season and elevation derived from Copernicus DEM. It should be noted that the JRC Global built-up surface dataset (*bulit_surface* in Fig. 9) kept in the feature space even though it expressed lower importance for overall results. However, this dataset was the most important driver to classify built-up areas correctly.

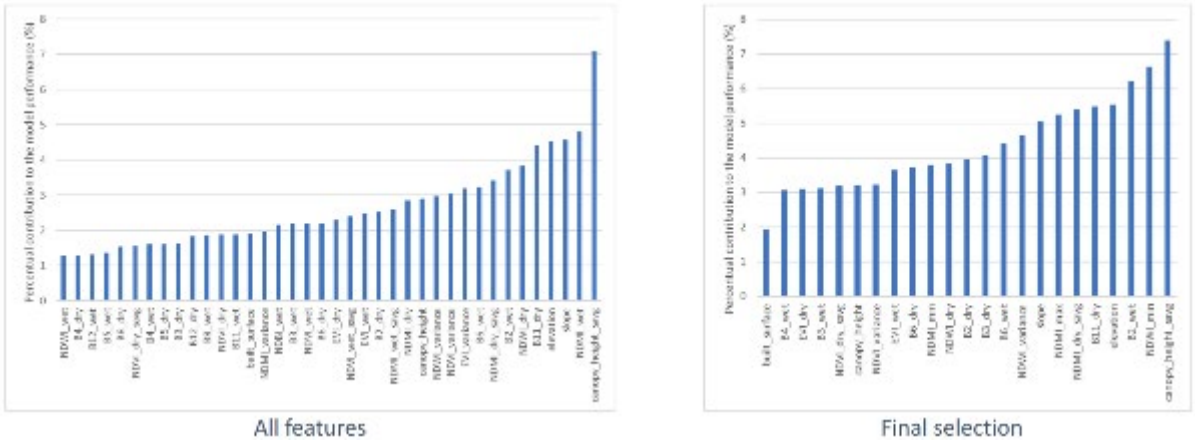


Figure 9: All input features (left) and selected features after the permutation feature importance analysis (right).

4.3 Geometric and thematic accuracy assessment

Statistical analysis – the overall accuracy, the macro F-score, user's and producer's accuracy and individual F-scores for each of the classes were assessed. The overall accuracy (range: 0-100%) represents the number of correctly classified areas over the number of validation points. The F-score (range: 0-1) of individual classes is a measure which combines precision and recall to evaluate the model's accuracy, balancing the rates of omission errors (missed labels) and commission errors (incorrectly predicted labels). The macro F-score calculates the F-score independently for each label and then averages these scores, treating all labels with equal importance regardless of their frequency. User's Accuracy (UA), also precision, indicates the probability that a pixel classified as a specific class actually belongs to that class. It measures the reliability of the classification from the user's perspective. Producer's Accuracy (PA), also recall, indicates the probability that a reference pixel of a certain class has been correctly classified.

The highest overall accuracy achieved was 85.06%, with the overall macro F-score of 0.852. The best performing classes were water, built-up areas, tree cover and cropland (1, 0.90, 0.89, and 0.88 of F-score respectively). The lowest accuracy in terms of F-score was achieved by shrublands and grasslands (0.70 and 0.71, respectively). All other classes also achieved F-scores higher than 0.81.
The F-scores in 2023 for each class were:

- Forest: 0.89
- Shrubland: 0.70
- Grassland: 0.71
- Cropland: 0.88
- Built-Up: 0.90
- Bare land: 0.86
- Water bodies: 1.00
- Wetlands: 0.85

For the 2017 land cover map, the highest overall accuracy achieved was 84.00%, with an overall macro F-score of 0.85. The best-performing classes were Water bodies, Wetlands, Built-Up, and Bare land

with F-scores of approximately 1.00, 0.95, 0.91, and 0.92, respectively. The lowest accuracy in terms of F-score was achieved by Shrubland and Grassland with F-scores of around 0.63 and 0.66. All other classes achieved F-scores higher than 0.76, indicating solid model performance across most classes but some room for improvement in distinguishing shrubland and grassland areas.

Table 2: Validation report for the National Land Cover of Zambia for 2023

Actual \ Predicted	Forest	Shrubland	Grassland	Cropland	Built-Up	Bare land	Water bodies	Wetlands	Row Total	PA (%)
Forest	120	12	0	0	0	0	0	0	132	90.91
Shrubland	13	46	3	0	0	0	0	0	62	74.19
Grassland	2	7	32	3	0	0	0	1	45	71.11
Cropland	0	5	4	57	0	0	0	2	68	83.82
Built-Up	0	0	1	1	28	0	0	0	30	93.33
Bare land	0	0	0	0	4	12	0	0	16	75.00
Water bodies	0	0	0	0	0	0	46	0	46	100.00
Wetlands	2	0	5	0	0	0	0	29	36	80.56
Column	137	70	45	61	32	12	46	32	OA: **85.06%**	
UA (%)	87.6	65.7	71.1	93.4	87.5	100.0	100.0	90.6		

Table 3: Validation report for the National Land Cover of Zambia for 2017

Actual \ Predicted	Forest	Shrubland	Grassland	Cropland	Built-Up	Bare land	Water bodies	Wetlands	Row Total	PA (%)
Forest	90	10	1	0	0	0	0	0	101	89.11
Shrubland	3	23	0	0	0	0	0	0	26	88.46
Grassland	3	8	25	7	0	1	0	1	45	55.56
Cropland	1	3	1	22	0	0	0	0	27	81.48
Built-Up	0	3	1	2	31	0	0	0	37	83.78
Bare land	0	0	0	0	0	6	0	0	6	100.00
Water bodies	0	0	0	0	0	0	20	0	20	100.00
Wetlands	0	0	3	0	0	0	0	35	38	92.11
Column	97	47	31	31	31	7	20	36	OA: **84.00%**	
UA (%)	92.78	48.94	80.65	70.97	100.0	85.71	100.0	97.22		

The F-scores in 2017 for each class were:

- Forest: 0.91
- Shrubland: 0.63
- Grassland: 0.66
- Cropland: 0.76
- Built-Up: 0.91
- Bare land: 0.92
- Water bodies: 1.00
- Wetlands: 0.95

Misclassifications between certain land cover types, such as forest and shrublands, can arise due to the similar spectral and structural characteristics these ecosystems may share, particularly in regions with transitional vegetation. Irrigated agriculture and wetlands also tend to be confused because the presence of water in both can lead to overlapping signatures, especially when using remote sensing data that captures moisture content. During dry periods, the distinction between wetlands and grasslands becomes even more challenging, as wetlands may lose visible water/moisture, making them resemble grasslands in spectral imagery.

Furthermore, the lack of adequate training and testing data for specific land cover types, such as grassland, bare land, and wetland, can contribute to these classification issues. Insufficient sample data limits the model's ability to learn the unique features of these classes, leading to reduced classification accuracy. This is particularly problematic in remote or underrepresented regions where collecting ground truth data is difficult, which increases the likelihood of misclassifications and weakens the overall robustness of land cover maps. The scarcity of reliable data makes it difficult to improve machine learning models, creating a feedback loop of lower precision for these critical land types.

These misclassifications cause lower accuracies for selected land cover classes. This problem can be overcome by further validation of input training and testing sets by local stakeholders, by providing additional reference areas from ground-truth measurements, high-resolution image interpretation or from their archives from earlier projects.

Next steps:

A thorough validation of the Land Cover Map for both 2017 and 2023 is needed by local stakeholders. Problematic areas, including commission and omission errors (areas which falsely detected and missing areas, respectively) should be identified for further improvement of the obtained results.

Crucial will be to further validate the input training and testing datasets by local stakeholders, by providing additional reference areas from ground-truth measurements, high-resolution image interpretation or from their archives from earlier projects.

4.4 Export of the final map of land cover

Challenges: The main challenge is the computational complexity of all the mosaicking, calculations, auxiliary data integration, data pre-processing, feature engineering, the main analysis, post-processing and accuracy assessment, even for the robust cloud-based platform GEE. The export of the

final, best-performing product is not possible as one image due to its large geographical coverage in a 10-meter pixel size.

The final product was therefore divided into districts, while the export of each district takes 1-4 hours, depending on the size of that district, making the export of one national product a 2-day long export process. Moreover, when we need to export data to our local computer, a two-step export is needed: the first export is an export to a GEE Asset and then a follow-up export of the GEE Asset to Google Drive is needed. From our Google Drive we can then download the resulting data in 72 GeoTiff parts (each part representing one district). These separate parts can then be merged in a desktop platform, e.g., QGIS.

4.5 Selected Methodology for Land Cover Change 2017-2023

For the land cover change map, two approaches were tested.

In the first approach, two land cover maps were created with the same methodology. Then, the land cover change map was created by subtracting these two land cover maps: 2017-2023. The resulting land cover map was able to show areas with no change and areas with various changes from one land cover type to another.

When generating land cover change maps, noisy results can still be problematic, even when an overall accuracy (OA) of over 82% is achieved in both (2017 and 2023) classifications. Despite high accuracy, errors present in both the initial and final land cover classifications tend to compound over time, leading to a multiplied effect in the resulting land cover change detection. This means that even small inaccuracies in each classification step can result in significant errors in the final change map.

Applying a 1-hectare post-filter can help to reduce some of this noise by smoothing out smaller, scattered misclassifications, but this technique is often insufficient to fully address the underlying errors. While the post-filter may reduce the appearance of small, isolated misclassifications, it does not eliminate the larger, more systematic errors that arise from compounded classification mistakes. As a result, noisy and inaccurate land cover change maps can still emerge, potentially undermining the reliability of the analysis.

For these reasons, to improve the accuracy of land cover change detection and mitigate the issues of noisy results and compounding errors, the following method was developed:

- **Generate two land cover maps**: The first step involves creating two separate land cover maps for 2017 and 2023. These maps represent the land cover classification for these two years and serve as the basis for analysing changes over time.
- **Create a land cover change map from the two LC maps** and apply a 1-hectare post-filter (PF): Using the two generated land cover maps, a land cover change map is created by comparing the classifications between 2017 and 2023. A 1-hectare post-filter is applied to reduce small, isolated misclassifications and smooth the change map, improving its visual clarity and accuracy by focusing on more significant changes rather than scattered noise.
- **Generate a change-no change mask based on multi-temporal vegetation index differences and thresholding** for both wet and dry seasons + apply a 1-ha PF: To further refine the change detection, a change-no change mask is generated using multi-temporal vegetation indices

(NDVI and NDMI). By calculating the difference between vegetation indices from both the wet and dry seasons and applying a threshold, areas of significant vegetation change are identified. This approach captures seasonal variations, providing a more comprehensive view of land cover dynamics. A 1-hectare post-filter is then applied to this mask to remove small patches of noise.

Apply the threshold mask to the result from step 2 + apply a 1-ha PF: The change-no change mask generated in step 3 is applied to the land cover change map created in step 2. This ensures that only the areas identified as undergoing real changes (based on vegetation index thresholds) are retained, filtering out areas where no significant change occurred. A 1-hectare post-filter is again applied to improve the map's clarity.

Update no-change areas from step 2 with the mask from step 3: The areas identified as experiencing no significant change in the change-no change mask (step 3) are used to update the corresponding areas in the land cover change map (step 2). This ensures that regions showing no real changes in both change masks are classified as no-change.

Generate final land cover change maps based on the updated land cover change mask: With the no-change areas updated and the noise filtered out, the final land cover change maps are produced. These maps reflect more accurate land cover transitions over time, with significantly reduced noise and errors, ensuring a more reliable representation of actual land cover dynamics.

This approach strategically combines multi-temporal vegetation data, thresholding, and post-filtering techniques to address the limitations of traditional simple bi-temporal change detection methods, enhancing the quality and accuracy of land cover change maps.

Next steps:

A thorough validation of the Land Cover Change Map 2017-2023 is needed by local stakeholders. Problematic areas, including commission and omission errors (areas which falsely detected and missing areas, respectively) should be identified for further improvement of the obtained results.

5 TIMELINE AND MILESTONES

The timeline for the NLA Land Cover generation (June–November 2024) outlines a structured progression of activities:

1. June 2024: The analysis starts, marked by the first on-site mission to produce the first draft of the land cover map.
2. September 2024: The results are refined in September, with the focus on the generation of a land cover change map between 2017 and 2023. A technical workshop on land use and land cover was held and an intensive 3 full-day training on land cover (LC) mapping was conducted. Final results of the 2023 Land cover and a Land Cover Change map was generated for further validation by local stakeholders.
3. November-December 2024: The validation phase on 2023 National Land Cover Map and for the National Land Cover Change Map 2017-2023.
4. January-March 2025: Improvement of the results based on obtained feedback
5. January-March 2025: Online training activities on satellite data processing and land cover monitoring using satellite data and cloud-based technologies.

Fig 10. shows milestones, visibility activities, and data processing to ensure the smooth progression of tasks.

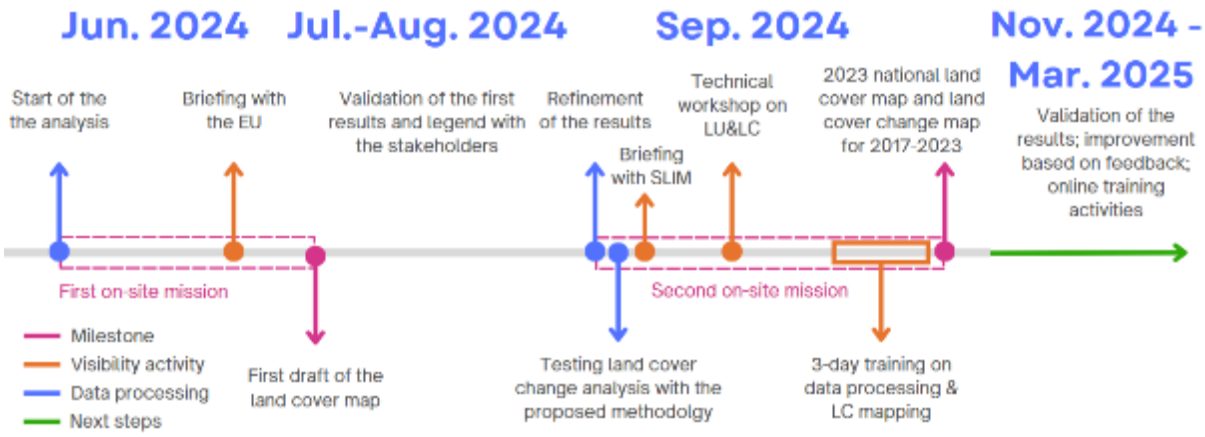


Figure 10: Timeline of the National Land Cover mapping in Zambia

6 MAIN RESULTS ACHIEVED

6.1 Guidelines for processing and classification

Detailed description for image pre-processing, feature engineering, classification, post-processing and accuracy assessment in Google Earth Engine to be used for land cover mapping is available in this report. The finalised methodology of land cover classification can be seen in the attached methodology overview in Fig. 11.

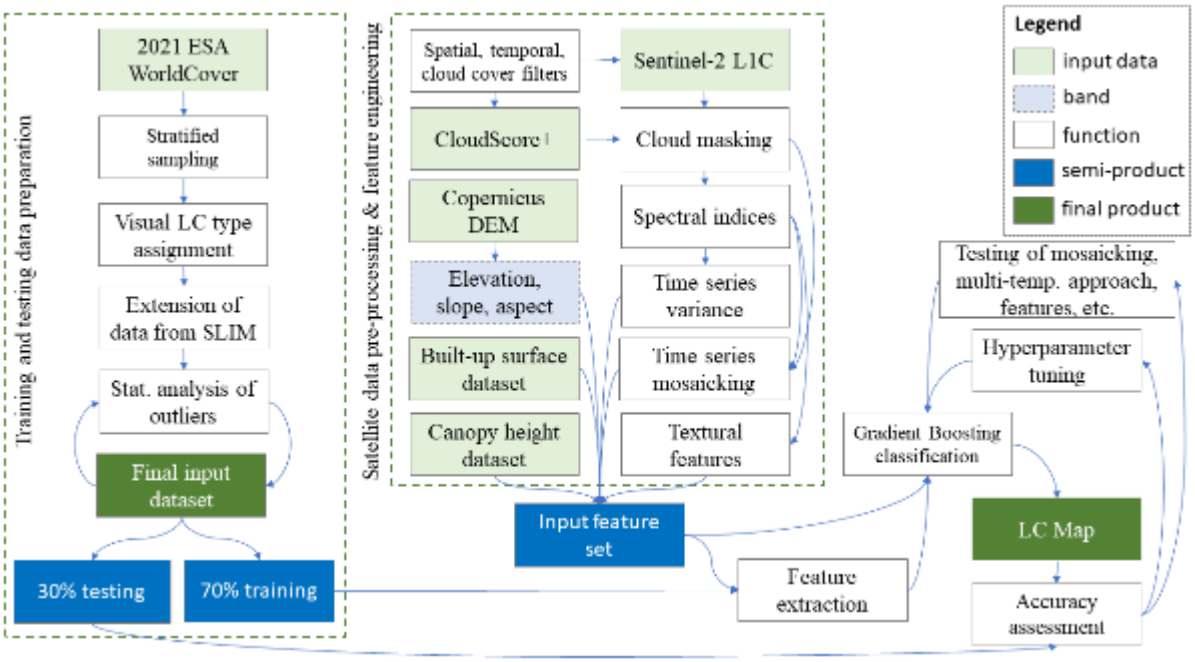


Figure 11 : General overview of the methodology

The following sub-products are available as well:

- Validated training dataset for 2023 (used in the model training): https://code.earthengine.google.com/?asset=users/danielp/Zambia-NLA/training_2023
- Validated testing dataset for 2023 (used in the validation of the 2023 LC product): https://code.earthengine.google.com/?asset=users/danielp/Zambia-NLA/testing_2023
- Validated testing dataset for 2017 (used in the validation of the 2017 LC product): https://code.earthengine.google.com/?asset=projects/danielp-cuni/assets/testing_2017

Prepared, but not used in the final product generation:

- Validated training dataset for 2017: https://code.earthengine.google.com/?asset=projects/danielp-cuni/assets/training_2017

6.1.1 The trained Random Forest model

The trained Random Forest model, described in chapter 4.2.7, is available in GEE: https://code.earthengine.google.com/?asset=projects/danielp-cuni/assets/NLA-Zambia/RF_classifier.

6.1.2 Land cover legend

The proposed land cover legend with 8 classes (forest, shrubland, grassland, cropland, built-up, bare land, water, wetlands) defined in Chapter 4.1.1. was used. Please refer to this Chapter for detailed information.

6.2 Validation report

The first draft of the National Land Cover map for 2023 using the accuracy metrics defined in chapter 4.3 of this report, available in Table 2.

6.3 National land cover map for 2023 and 2017

First draft of the national land cover map of Zambia for 2023 with an 85% of overall accuracy

- Available as GeoTiff for local use
  i. For 2017
  ii. For 2023

- Available as GEE Asset for use in GEE, available here:
  i. For 2017
  ii. For 2023

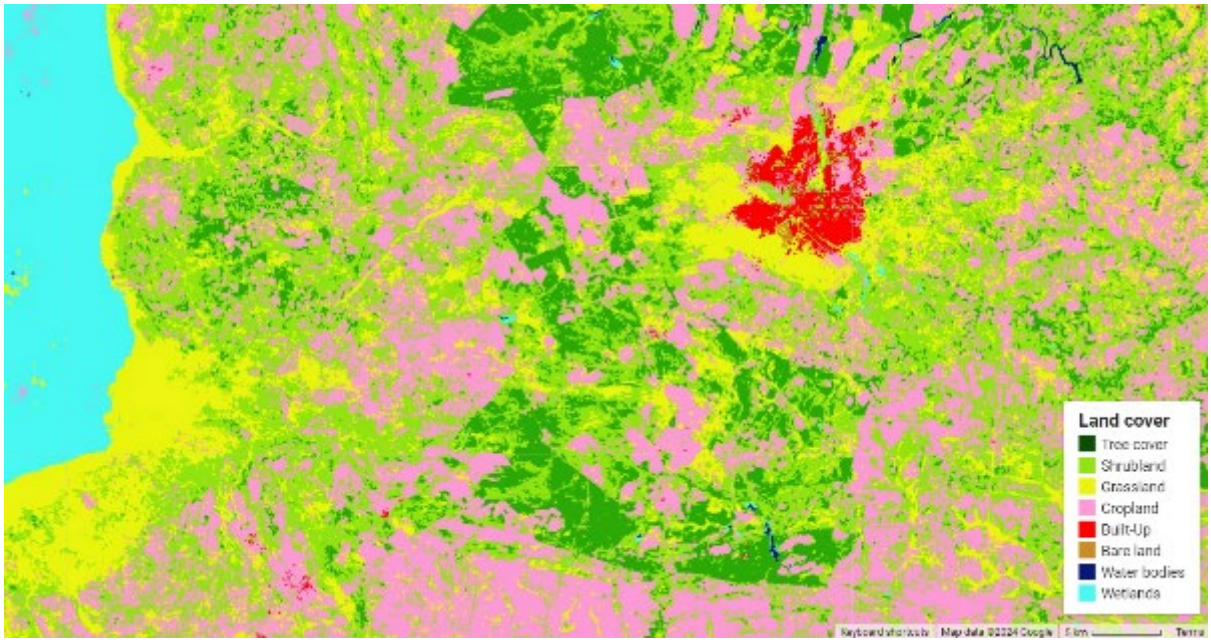


Figure 12. The final National Land Cover map for 2017 (zoomed in to an area around Kabwe city)

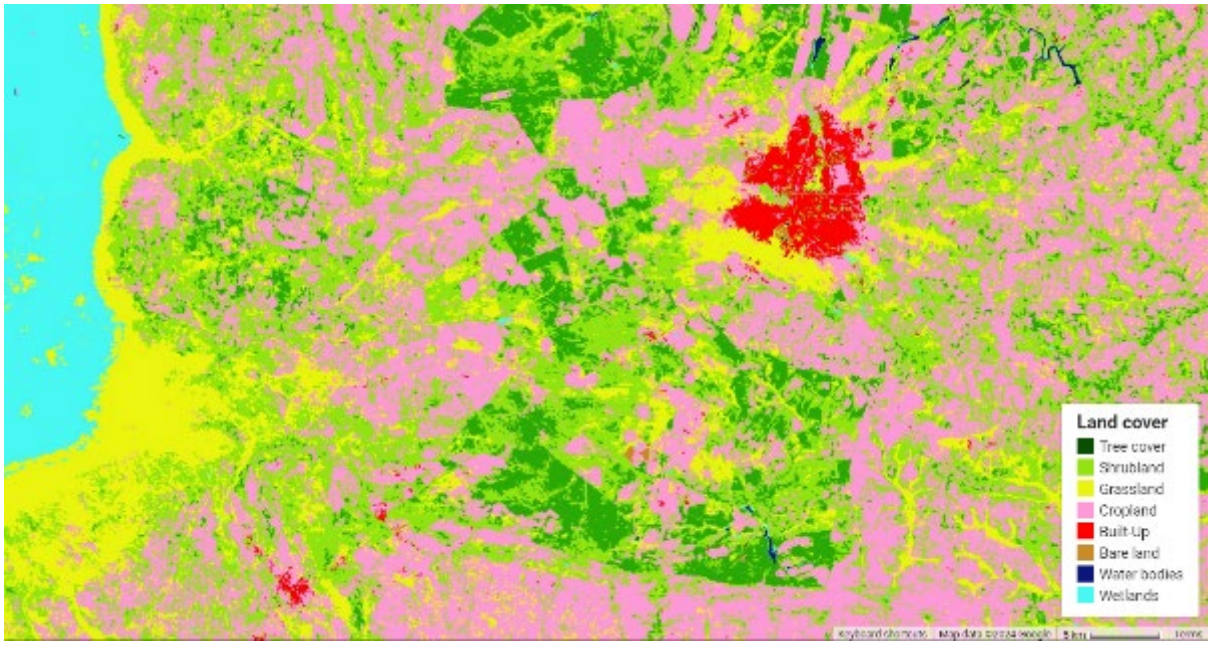


Figure 13. The final National Land Cover map for 2023 (zoomed in to an area around Kabwe city)

The final land cover maps of 2017 and 2023 have been uploaded on the NSDI.

Based on classification results, in 2017, forest areas dominate the landscape, covering 293,529 km² (39.1% of the total land area), followed by shrubland at 252,323 km² (33.6%). Grasslands and croplands are also significant, accounting for 103,948 km² (13.9%) and 64,919 km² (8.7%), respectively. Wetlands and water bodies occupy smaller portions, with 20,086 km² (2.7%) and 12,369 km² (1.6%), while built-up areas cover 2,147 km² (0.3%), and bare land represents the smallest share at 872 km² (0.1%).

Based on classification results, in 2023, forest areas dominate the landscape, covering 290,207 km² (38.6% of the total land area), followed by shrubland at 261,667 km² (34.8%). Grasslands and croplands are also significant, accounting for 88,407 km² (11.8%) and 70,775 km² (9.4%), respectively. Wetlands and water bodies occupy smaller portions, with 24,524 km² (3.3%) and 12,994 km² (1.7%), while built-up areas cover 1,802 km² (0.2%), and bare land represents the smallest share at 688 km² (0.1%). This composition reflects the diversity of Zambia's ecosystems, highlighting substantial forest and shrubland areas alongside other key land cover types.

The decrease in built-up areas and bare lands in 2023 compared to 2017 can be attributed to misclassifications between these two classes in 2023, as reported in Table 3.

Table 4. Land cover composition of Zambia in 2023 based on classification results

	2023		2017	
Land cover class	**Area in km²**	**Percentage share**	**Area in km²**	**Percentage share**
Forest	290,207	38.6	293,529	39.1
Shrubland	261,667	34.8	252,323	33.6
Grassland	88,407	11.8	103,948	13.9
Cropland	70,775	9.4	64,919	8.7
Wetland	24,524	3.3	20,086	2.7
Water	12,994	1.7	12,369	1.6
Built-up	1,802	0.2	2,147	0.3
Bare land	688	0.1	872	0.1

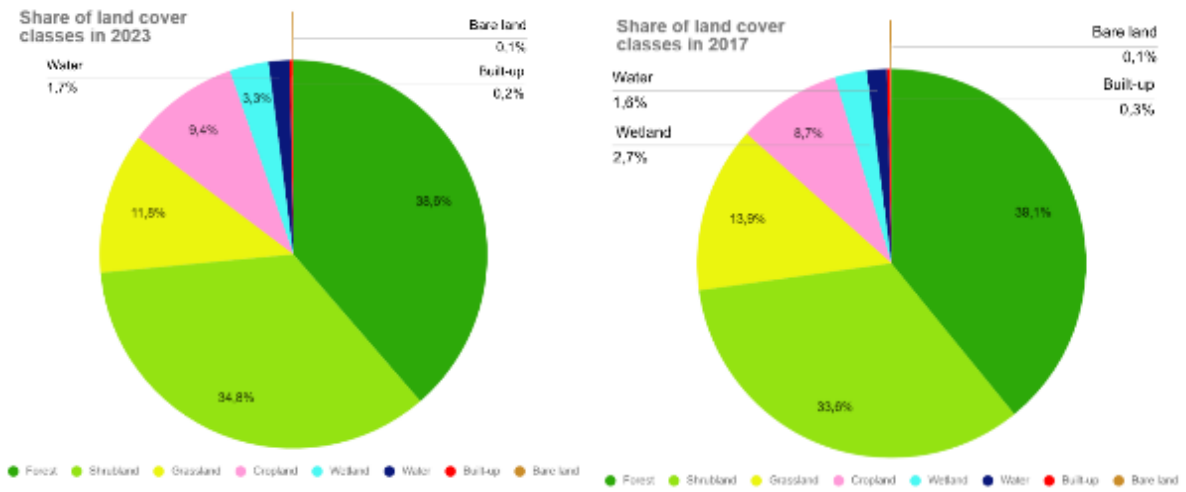


Figure 13. and Fig. 14. Land cover composition in Zambia in 2023; Land cover composition in Zambia in 2017.

6.4 Land cover change map

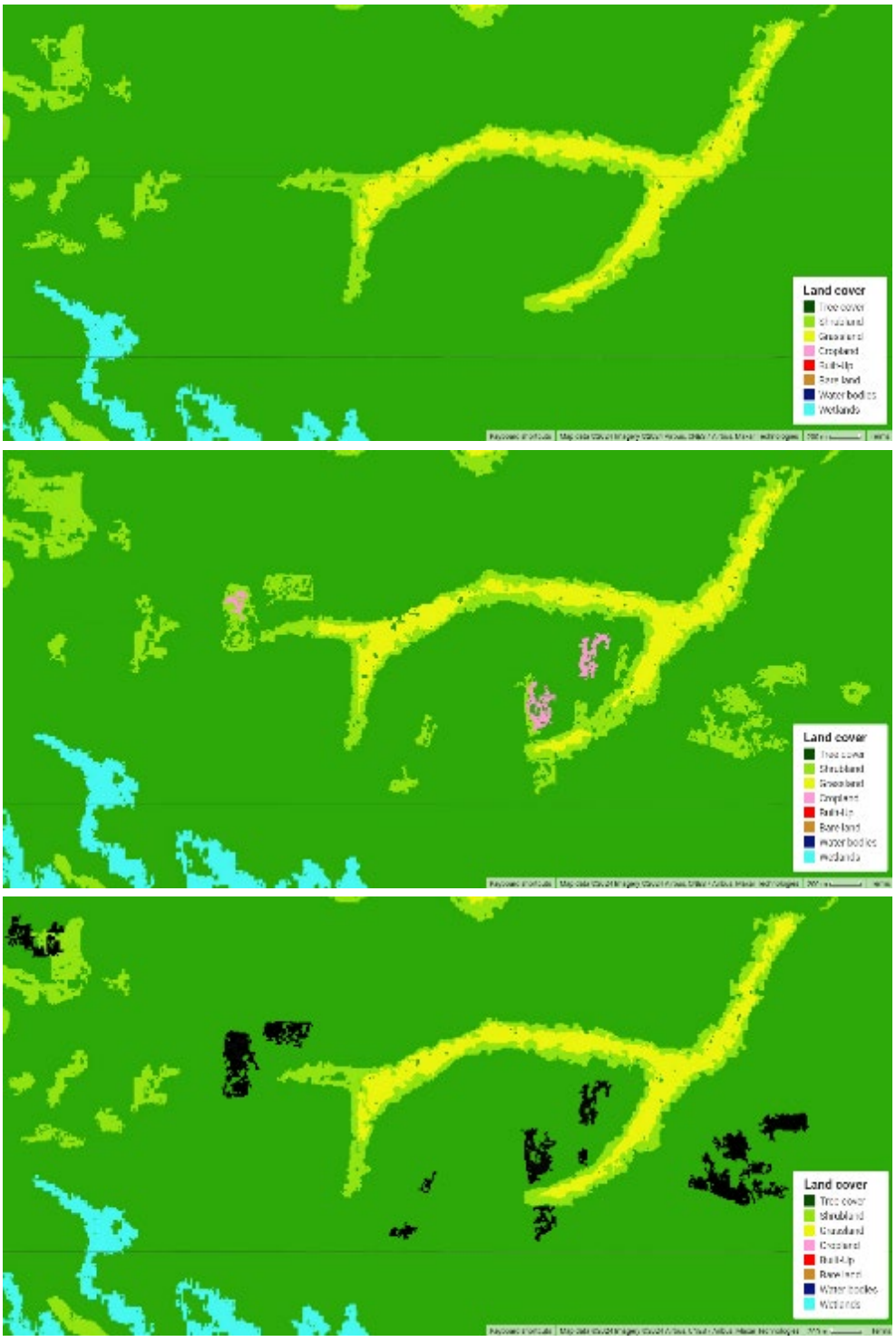


Figure 15. An example of land cover change detection results (central coordinates of -14.647, 29.346).

The land cover composition in 2017 is depicted on the top image, middle image is the land cover in 2023 and the bottom image shows the 2017 image overlaid with the land cover change layer (black areas represent land cover change).

6.5 Land cover change vs. documented deforestation

6.5.1 Overview

Selective logging is a common practice in Zambia's forests, where trees are carefully removed based on criteria like height, diameter, or species. The goal is to harvest timber while preserving the forest ecosystem. However, the effectiveness of selective logging in Zambia is challenged by several factors, including, i**ndigenous timber exports.** Only 75% of harvested indigenous timber can be exported, while the remaining 25% must be used locally. Most exotic timber harvested comes from plantations and is often used for construction.

Zambia has experienced high rates of deforestation, primarily due to agriculture (both subsistence and commercial), logging, infrastructure development, and mining. The country loses an estimated 250,000 to 300,000 hectares of forest annually.

Several key factors are driving deforestation. One major cause is the expansion of agriculture, with large areas of forest being cleared to grow crops, especially maize and soybeans. Charcoal production is another significant contributor, as many households rely on charcoal as their main source of energy, leading to continuous forest cutting. Logging, both legal and illegal, also plays a major role in the decline of forest resources. In addition, infrastructure development—such as road construction and urban growth—results in habitat loss and the fragmentation of forested areas.

6.5.2 Forest Land Cover Mapping Efforts

Organizations like the Zambia National Remote Sensing Centre (ZRS) use satellite imagery to monitor land use and land cover changes. Landsat and Sentinel satellite imagery are commonly utilized for mapping forest extent and change over time. The Zambian government has conducted national inventories to assess forest resources, which include mapping efforts to understand the distribution and health of various forest types.

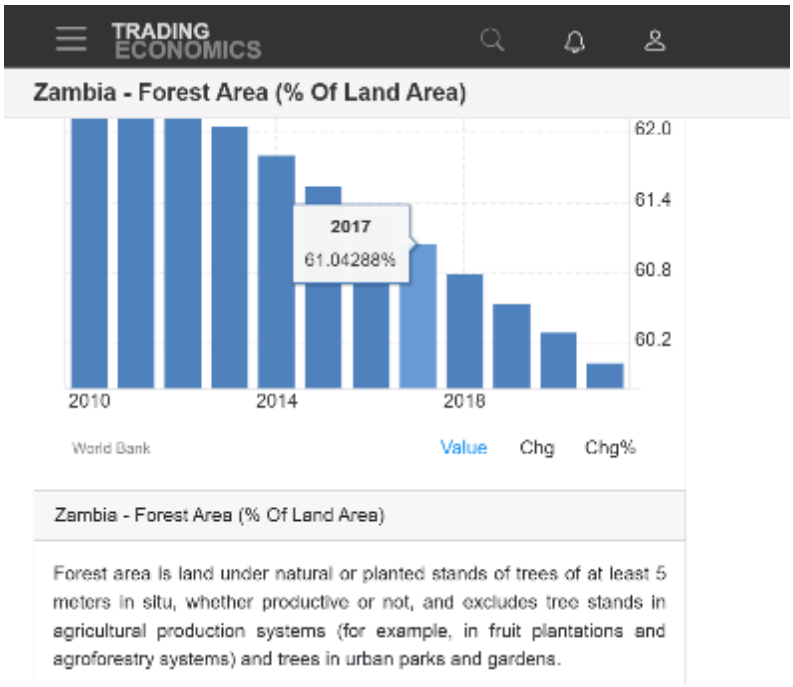
Various NGOs and international agencies, such as the Food and Agriculture Organization (FAO) and World Wildlife Fund (WWF), collaborate with the Zambian government on projects aimed at sustainable land use, conservation, and restoration of deforested areas.

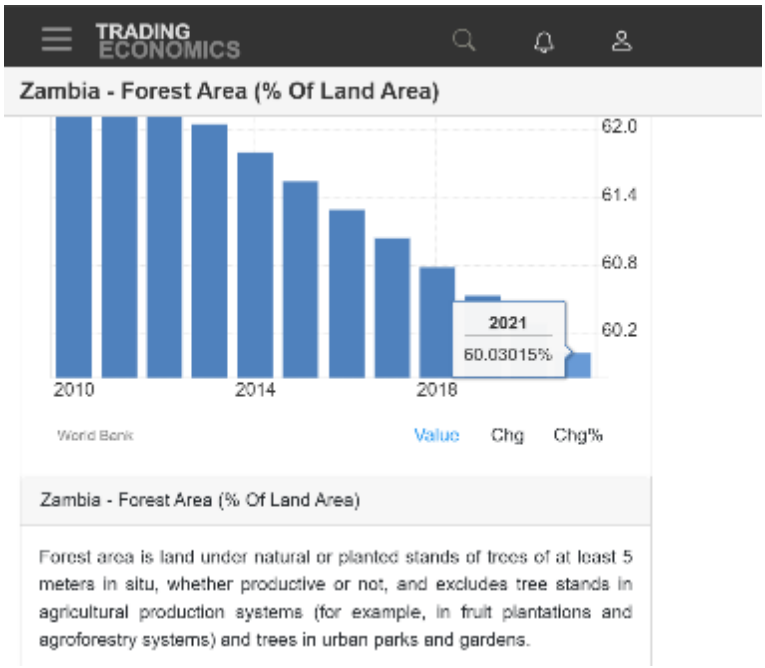
Zambia is also involved in REDD+ (Reducing Emissions from Deforestation and Forest Degradation), which aims to create financial incentives for developing countries to reduce forest loss and invest in sustainable land management.

6.5.3 Documented deforestation

The statistics shown below shows very different results presented by the different information sources. In addition, most documented deforestation statistics use a different forest definition, - e.g., - the Trading Economics source uses 5m height for forest cover (5-15m falls into the shrub category in the NLA land cover result). Other sources (Global Forest Watch) do not give detailed definition of the

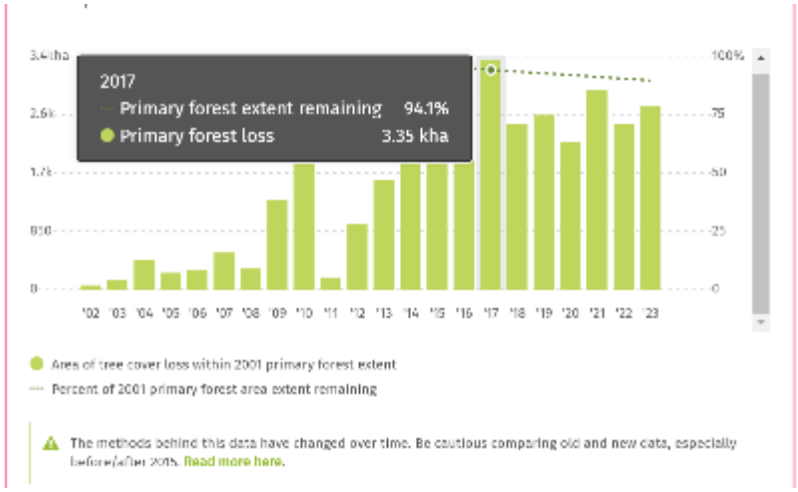
Forest land cover class. This makes a comparison between NLA forest class change and documented deforestation in a comparable time period difficult, if not impossible.

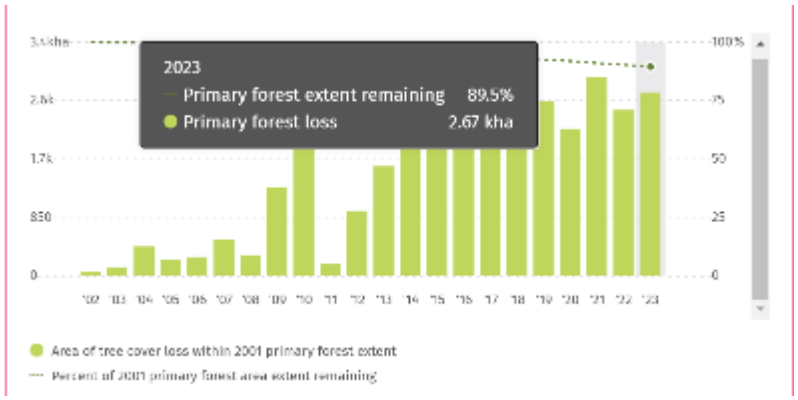




Forest loss using 5m height 2017-2021: **1.01273%**

Comparable forest loss according to globalforestwatch:





Difference between 2017 and 2023: 4.6%, or 2,275,528 ha.

Land cover	Y2018 (ha)	Y2019 (ha)	Y2020 (ha)
Cropland	2,571,702.10	3,152,100.41	2,971,463.34
Forest	41,002,568.01	39,017,664.48	38,899,290.28
Grassland	27,887,840.48	29,660,235.80	29,691,066.61
Other land	1,118,396.05	1,140,446.59	1,134,001.14
Settlement	332,193.95	361,325.67	378,909.78
Wetland	2,348,699.41	1,929,627.05	2,186,668.85
TOTAL	75,261,400.00	75,261,400.00	75,261,400.00

Changes to forest cover were observed in protected forests. Forest area reduced from 5.5 million to 5.3 million hectares 27, a reduction of 2.6 percent (Table 2).

Forest reduction 2018-2020, according to World Bank (above): 2 103 278 ha.

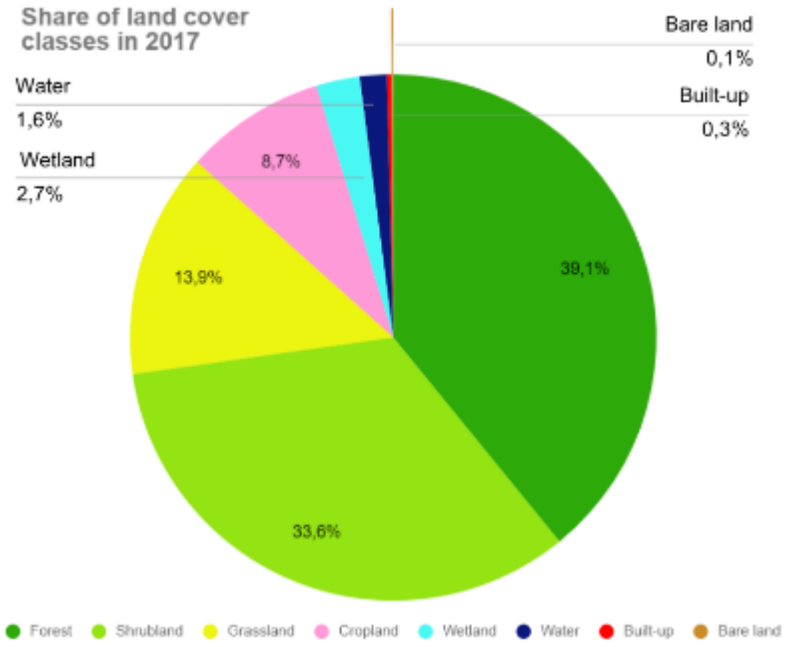
	2023		2017	
Land cover class	Area in km²	Percentage share	Area in km²	Percentage share
Forest	290,207	38.6	293,529	39.1
Shrubland	261,667	34.8	252,323	33.6
Grassland	88,407	11.8	103,948	13.9
Cropland	70,775	9.4	64,919	8.7
Wetland	24,524	3.3	20,086	2.7
Water	12,994	1.7	12,369	1.6
Built-up	1,802	0.2	2,147	0.3
Bare land	688	0.1	872	0.1

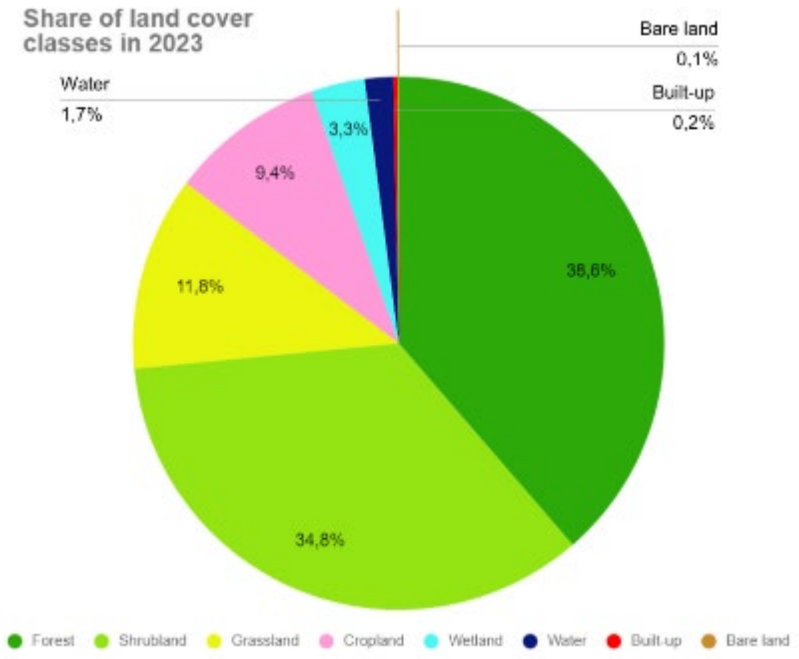
Difference:

Forest: **- 3322 sq km, or 332,200 ha**

Shrub: **+9344 sq km, or 934,400 ha**

Results show an increase of shrub-covered areas and a decrease of forested areas.





Land cover composition of Zambia in 2017 and 2023 based on classification results. The shown deforestation is constrained to clearcuts of primary dense forests (mature trees higher than 15m and 80% canopy/crown closure). However, the dominant de-forestation in Zambia is caused by selective logging and burning. Selective land cover change validation confirms this.

Primary dense forest clearcuts between 2017-2023: **332,200 ha**
Selective logging and burns between 2017-2023: **1,943,328 ha**

6.5.4 Classification challenges

Classifying land cover becomes a challenging exercise when trying to tell the difference between similar types like grasslands, shrublands, and forests. These challenges show that it's important to adjust international classification systems, like the one from the FAO, to fit the specific purpose of a project. The FAO system is meant to be general and used around the world, but it doesn't always work well for detailed local studies or specific goals.

One clear example of this issue is the difference in forest loss numbers between the National Land Assessment (NLA) and other global or local datasets. These differences often happen because each source may use different definitions, categories, or methods. This shows how important it is to make sure everyone uses the same standards and approaches if we want to compare or understand land cover data correctly.

In some cases, putting similar land cover types into one group can give a clearer and more useful picture. For example, combining forests and shrublands might help show deforestation more accurately, especially in places where it's hard to separate the two. The same idea applies to croplands

and grasslands, which often change back and forth in farming areas. Grouping them could give a better view of land used for agriculture.

Being able to adjust or combine land cover classes depending on what the data is for is a simple but powerful tool. Whether we are tracking forest loss, planning land use, or studying ecosystems, using the right classification for the job—along with clear definitions and methods—helps make the results more accurate and useful.

6.6 Capacity building in satellite data utilization using cloud-based platforms and in land cover mapping

The training on using cloud-based platforms for satellite data processing and on land cover mapping was held in Lusaka, in the DIA Belgium Global SRL office on September 23-25 2024. The trainer of this training was Daniel Paluba, DAI Senior Non-Key Export for Land Cover and Land Use mapping, affiliated also with Charles University in Prague, Czechia.

The main objective of the training was to equip participants with the necessary skills and knowledge to effectively use the GEE platform for cloud-based remote sensing data processing and analysis, particularly for land cover mapping. The training was conducted over three days and consisted of a mix of theoretical and hands-on practical sessions. The topics covered included an introduction to GEE, basic image processing, advanced data exploration, time series analysis, and land cover classification using machine learning algorithms. The training provided the participants with a strong foundation in using GEE for remote sensing applications, particularly for environmental monitoring and land cover analysis. The hands-on exercises enabled the participants to gain practical experience in applying these tools to real-world scenarios.

The training was attended by 14 participants from various government agencies and universities in Zambia, mainly from Lusaka. The participants learned how to navigate the GEE interface, access and manipulate remote sensing datasets, perform cloud masking, calculate vegetation indices, and conduct land cover classification. The feedback from the participants was overwhelmingly positive, with many reporting increased confidence in using GEE for their remote sensing tasks.

The successful completion of the training program and the positive feedback from the participants suggest that the training achieved its objectives of equipping the participants with the necessary skills and knowledge to use GEE for land cover mapping and analysis. Moreover, based on the feedback, the participants are interested in and willing to attend similar training activities.

7 NEXT STEPS

The next steps are to validate the created National Land Cover maps for 2023 and 2017 and the Land Cover Change Map 2017-2023. Field measurements and validation of the land cover map will also be crucial. Even with the relatively high statistical results achieved, it is recommended to follow-up with visual inspection of the results by local stakeholders, inspection of the training and testing data used for the classification and accuracy assessment, respectively. Field measurements to explore the problematic areas classified incorrectly are recommended as well. Only based on the feedback from local stakeholders, we can further improve the thematic accuracy of the land cover results.

Based on the obtained feedback and validation from local stakeholders, we will be able to improve the results. In case of spatial discrepancies of the generated results, the most important data to validate by local experts are the training and testing datasets for both 2023 and 2017 provided in Chapter 6.1.

Further improvement of the results based on obtained feedback is expected in the first quarter of 2025.

Further capacity building activities in land cover mapping are recommended for local stakeholders with GIS and Remote Sensing expertise, located outside of Lusaka, is recommended through online training in the first quarter of 2025.

8 CONCLUSION

In conclusion, the objectives set for land cover mapping in this project have been successfully achieved. Up-to-date National Land Cover Maps for 2023 and 2017, alongside with a land cover change map between 2017 and 2023, have been produced, meeting international standards and achieving high overall accuracies.

The land cover legend was clearly defined, establishing a consistent classification system essential for mapping accuracy and clarity. Through extensive analysis, the most suitable time frame for image mosaic creation was identified, ensuring optimal data quality and seasonal consistency. Copernicus Sentinel-2 data, along with relevant auxiliary datasets, were explored and incorporated, supporting the robustness of the classification process.

A rigorous data collection process provided high-quality input for training and validating the machine learning models. Advanced algorithms were evaluated and selected for pre-processing, classification and validation, ensuring that state-of-the-art methods were applied at every stage. Additionally, optimal input features and algorithm parameters were chosen, enhancing the precision of both mono-temporal and multi-temporal classification approaches.

Dynamic algorithms were developed in GEE JavaScript API, providing flexibility for adapting to various datasets and parameters. These algorithms enabled efficient processing of large-scale data, culminating in the generation of national-level land cover maps for both 2017 and 2023.

A three-day intensive training on cloud-computing, big data analysis and land cover mapping using GEE was provided for local stakeholders described in a different. Training materials were prepared as well, covering essential workflows for data access, pre-processing, classification, time series analysis etc., laying the groundwork for capacity building.

This project has successfully met its goals, setting basis for future land cover monitoring initiatives in Zambia.

9 REFERENCES

Bielski C. et al., 2024, "Novel Approach for Ranking DEMs: Copernicus DEM Improves One Arc Second Open Global Topography," in IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1-22, 2024, Art no. 4503922, doi: 10.1109/TGRS.2024.3368015

Grinsztajn, Leo, Edouard Oyallon, and Gael Varoquaux. 2022. “Why Do Tree-Based Models Still Outperform Deep Learning on Typical Tabular Data?” Advances in Neural Information Processing Systems 35 (December):507–20.

Kerner H, Nakalembe C, Yang A, Zvonkov I, McWeeny R, Tseng G, Becker-Reshef I. How accurate are existing land cover maps for agriculture in Sub-Saharan Africa? Sci Data. 2024 May 10;11(1):486. doi: 10.1038/s41597-024-03306-z

Li, H., Zhao, J., Yan, B., Yue, L., & Wang, L. (2022). Global DEMs vary from one to another: an evaluation of newly released Copernicus, NASA and AW3D30 DEM on selected terrains of China using ICESat-2 altimetry data. International Journal of Digital Earth, 15(1), 1149–1168. https://doi.org/10.1080/17538947.2022.2094002

M. Amani et al., "Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 5326-5350, 2020, doi: 10.1109/JSTARS.2020.3021052.

Marešová, J.; Gdulová, K.; Pracná, P.; Moravec, D.; Gábor, L.; Prošek, J.; Barták, V.; Moudrý, V. Applicability of Data Acquisition Characteristics to the Identification of Local Artefacts in Global Digital Elevation Models: Comparison of the Copernicus and TanDEM-X DEMs. Remote Sens. 2021, 13, 3931. https://doi.org/10.3390/rs13193931

Maria del Rosario González-Moradas, Willem Viveen, Raúl Andrés Vidal-Villalobos, Juan Carlos Villegas-Lanza: A performance comparison of SRTM v. 3.0, AW3D30, ASTER GDEM3, Copernicus and TanDEM-X for tectonogeomorphic analysis in the South American Andes, CATENA, 228, 2023, https://doi.org/10.1016/j.catena.2023.107160

PASQUARELLA, V. (2024): All Clear with Cloud Score+, Google Earth and Earth Engine, Medium article Online: https://medium.com/google-earth/all-clear-with-cloud-score-bd6ee2e2235e (Last accessed 17. 6. 2024).

PASQUARELLA, V. J., BROWN, C. F., CZERWINSKI, W., RUCKLIDGE, W. J. (2023): Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2125–2135. https://doi.org/10.1109/CVPRW59228.2023.00206

Pesaresi, Martino; Politis, Panagiotis (2023): GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) doi: 10.2905/9F06F36F-4B11-47EC-ABB0-4F8B7B1D72EA

Pohjonen, V., 2004: Zambia Forest Resource Assessment 2004, EU – Forestry Support Programme in Zambia – 8 ACP/ 051.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., & Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*, 112165. https://doi.org/10.1016/j.rse.2020.112165

Samapriya Roy, Muhammed Abdelaal, Ujaval Gandhi, & Swetnam, T. (2024). samapriya/awesome-gee-community-datasets: Community Catalog (2.7.0). Zenodo. https://doi.org/10.5281/zenodo.12057445

Santoro, M.; Cartus, O. (2023): ESA Biomass Climate Change Initiative (Biomass_cci): Global datasets of forest above-ground biomass for the years 2010, 2017, 2018, 2019 and 2020, v4. NERC EDS Centre for Environmental Data Analysis, 21 April 2023. doi:10.5285/af60720c1e404a9e9d2c145d2b2ead4e.

Shafizadeh-Moghadam, H., Khazaei, M., Alavipanah, S. K., & Weng, Q. (2021). Google Earth Engine for large-scale land use and land cover mapping: an object-based classification approach using spectral, textural and topographical factors. GIScience & Remote Sensing, 58(6), 914–928. https://doi.org/10.1080/15481603.2021.1947623

Svoboda, J.; Štych, P.; Laštovička, J.; Paluba, D.; Kobliuk, N. Random Forest Classification of Land Use, Land-Use Change and Forestry (LULUCF) Using Sentinel-2 Data—A Case Study of Czechia. Remote Sens. 2022, 14, 1189. https://doi.org/10.3390/rs14051189.

Tamiminia, H., B. Salehi, M. Mahdianpari, L. Quackenbush, S. Adeli, and B. Brisco. 2020. “Google Earth Engine for Geo-big Data Applications: A Meta-analysis and Systematic Review.” ISPRS Journal of Photogrammetry and Remote Sensing 164: 152–170. doi:https://doi.org/10.1016/j.isprsjprs.2020.04.001.

Waśniewski, A.; Hościło, A.; Zagajewski, B.; Moukétou-Tarazewicz, D. Assessment of Sentinel-2 Satellite Images and Random Forest Classifier for Rainforest Mapping in Gabon. Forests 2020, 11, 941. https://doi.org/10.3390/f11090941



This report was produced with the financial support of the European Union.
Its contents are the sole responsibility of
DAI Global Belgium SRL and do not necessarily
reflect the views of the European Union