OBJECT-BASED MAPPING AND CLASSIFICATION FEATURES FOR TROPICAL HIGHLANDS USING SENTINEL-1, SENTINEL-2, AND GEDI CANOPY HEIGHT DATA - A CASE STUDY OF THE MURINGATO CATCHMENT, KENYA

Marcus Goebel, Kuria Thiong'o and Andreas Rienow

With 11 figures and 4 tables
Received 20 July 2022 · Accepted 14 March 2023

Summary: Tropical highlands remain a challenging target for remote sensing due to their high heterogeneity of the landscape and frequent cloud cover, causing a shortage of high-quality and reliable comprehensive data on land use and land cover on a local or regional scale. These, however, are urgently needed by local stakeholders and decisionmakers. This applies for example to the Muringato sub-catchment in Nyeri County, Kenya, where acute water problems have been identified to be usually directly related to specific land use and land cover. This article contributes to the understanding of tropical highlands from a remote sensing perspective by examining Sentinel-1, Sentinel-2 and Global Forest Canopy Height Model data from the Global Ecosystem Dynamics Investigation, all provided by the Google Earth Engine. To do so, we assess classifiers derived from these datasets for different land cover types, analyzing the performance of promising candidates identified in the literature, using 2,800 samples extracted from high-resolution image data across Nyeri County. We also propose an object-based classification strategy based on sequential masking. This strategy is adapted to very heterogeneous landscapes by refining image objects after re-evaluating their homogeneity. Small buildings, which constitute a significant part of the settlement structure in the area, are particularly difficult to detect. To improve the recognition of these objects we additionally consider the local contrast of the relevant classifier to identify potential candidates. Evaluating our sample data, we found that especially optical indices like the Sentinel Water Index, the Enhanced Normalized Difference Impervious Surfaces Index or specific Sentinel-2 bands combined with canopy height data are promising for water, built-up or tree cover detection. With these findings, our proposed object-based classification approach is applied to the Muringato sub-catchment as a representative example of the Kenyan tropical highland region. We achieve a classification accuracy of approximately 88% in the Muringato sub-catchment, outperforming existing products available for the study area. The knowledge gained in the study will also be used for future remote sensing-based monitoring of the region.

Zusammenfassung: Das tropische Hochland ist aufgrund großer Heterogenität der Landschaft und der häufigen Bewölkung nach wie vor ein schwieriges Ziel für die Fernerkundung, woraus ein Mangel an hochwertigen, zuverlässigen und umfassenden Daten über Landnutzung und Bodenbedeckung auf lokaler oder regionaler Ebene in diesen Regionen resultiert. Dies gilt beispielsweise auch für das Teileinzugsgebiet Muringato in Nyeri County, Kenia, wo akute Wasserprobleme meist direkt auf die Landnutzung und Bodenbedeckung zurückzuführen sind. Dieser Artikel trägt zum Verständnis des tropischen Hochlands aus der Perspektive der Fernerkundung bei, indem er von der Google Earth Engine bereitgestellte Daten von Sentinel-1, Sentinel-2 und des Global Forest Canopy Height Model aus der Global Ecosystem Dynamics Investigation untersucht. Zu diesem Zweck bewerten wir aus diesen Datensätzen abgeleitete Klassifikatoren für verschiedene Landbedeckungsarten und analysieren die Leistung vielversprechender, in der Literatur identifizierter Kandidaten anhand von 2800 Stichproben, die aus hochauflösenden Bilddaten von Nyeri County extrahiert wurden. Außerdem schlagen wir eine objektbasierte Klassifizierungsstrategie vor, die auf sequenzieller Maskierung basiert. Diese Strategie ist an sehr heterogene Landschaften angepasst, indem Bildobjekte nach der Generierung hinsichtlich ihrer Homogenität bewertet gegebenenfalls neu auf einer feineren Ebene segmentiert werden. Ergänzend werden kleine Gebäude, welche einen bedeutenden Teil der Siedlungsstruktur in diesem Gebiet ausmachen und besonders anspruchsvoll zu erkennen sind, zusätzlich mit mithilfe des lokalen Kontrastes des entsprechenden Klassifikators identifiziert. Bei der Auswertung unserer Stichproben haben wir festgestellt, dass insbesondere optische Indizes wie der Sentinel Water Index, der Enhanced Normalized Difference Impervious Surfaces Index oder bestimmte Sentinel-2-Bänder in Kombination mit Daten zur Vegetationshöhe vielversprechend für die Erkennung von Wasser, Bebauung oder Baumbedeckung sind. Unter Berücksichtigung dieser Erkenntnisse wenden wir den von uns vorgeschlagenen Klassifizierungsansatz auf das Teileinzugsgebiet Muringato an, welcher als repräsentatives Beispiel für die tropische Hochlandregion Kenias dient. Dabei erzielen wir eine Klassifizierungsgenauigkeit von ca. 88 % und übertreffen damit die für das Untersuchungsgebiet bislang verfügbaren Produkte. Die in der Studie gewonnenen Erkenntnisse werden für die künftige fernerkundungsbasierte Überwachung der Region genutzt.

Keywords: Remote sensing, Sentinel-1, Sentinel-2, GEDI Global Forest Canopy Height Model, object-based classification, tropical highlands

https://doi.org/10.3112/erdkunde.2023.01.03 ISSN 0014-0015 (Print) · ISSN 2702-5985 (Online)

1 Introduction

The Muringato sub-catchment in Nyeri County, Kenya, is a tropical highland region where accurate land surface data are urgently needed. The local Water Resources User Association (WRUA) released a Sub-Catchment Management Plan (SCMP) in 2014, highlighting the region's most pressing water resource challenges. Most of these problems are directly related to poor land use practices or unfavourable changes in land cover such as deforestation or unsuitable farming methods (Muringato Water Resource User Association 2014), which have accelerated in sub-Saharan Africa over the last two decades (Thonfeld et al. 2020). In response, a baseline data survey was undertaken in February 2018 (IGGReS 2018). This inventory was a first step in characterising the sub-catchment to support decision making regarding the specific water-related problems in Muringato. However, given that the situation is known to be particularly dynamic and regular in-situ data acquisition is not practical or economical, automated and repeatable methods are needed to support the work of the local WRUA and to efficiently monitor the constant changes across this large region of approximately 225 km² in size.

Remote sensing can be used to complement existing sources for data acquisition and may be the only practicable option when dense in-situ measurement networks are not available (Sheffield et al. 2018). Several land use and land cover (LULC) products of continental or global coverage already exist encompassing a range of different resolutions. This includes the ESA WorldCover (10 m), ESA CCI Landcover (20 m), Copernicus Global Land Cover (100 m), and the ESA GlobCover (300 m). However, these kinds of products tend not to accurately depict the regional characteristics of heterogeneous and dynamic regions such as tropical highlands (Schulz et al. 2021, Nabil et al. 2020). Studies on LULC used to rely predominantly on Landsat data, given that it is the only earth observation system encompassing the last 45 years (Thonfeld et al. 2020). This is advantageous for creating historical time series datasets but not may be optimal for tracking small targets due to the 30 m spatial resolution.

Higher spatial resolution can be achieved using data from the European Space Agency's (ESA) Copernicus Sentinel-2 mission (Drusch et al. 2012), which also offers superior spectral resolution from the near infrared spectrum regions with its Multispectral Instrument (MSI) (Kaplan & Avdan 2018). Sentinel-2 is a two-Satellite constellation (Sentinel-2A and Sentinel 2B) which systematically acquires observations from land and coastal areas from 56° south to 84° north with a frequent revisit time at the equator of five days (Drusch et al. 2012). Despite this improvement in resolution, the accuracy of mapping products for Africa often remains unsatisfactory due to highly challenging conditions. High levels of landscape heterogeneity, frequent cloud cover, and small sized farmland, which is common in African agriculture, are the main causes of uncertainties in land use and land cover (LULC) products for many African countries (Nabil et al. 2020, Kumar & Reshmidevi 2013, Bégué et al. 2020).

Therefore, it may be advisable to increase the data density and add additional levels of information by using additional sources and complementary sensor types. Due to the high level of cloud occurrence in tropical highlands, weather-independent microwave sensors such as ESA's Sentinel-1 C-band synthetic aperture radar (SAR) seem a logical choice (Kaplan & Avdan 2018, Nabil et al. 2020, Torres et al. 2012). Sentinel-1 is also a two-Satellite constellation (Sentinel-1A and Sentinel 1B) following a preprogrammed conflict-free scenario mapping global landmasses with a six day repeat cycle at the equator (Torres et al. 2012). Several studies have highlighted potential benefits of synergistic Sentinel-1-SAR and Sentinel-2-MSI for use in other cases and landscapes including mapping impervious surfaces in Pakistan using a random forest machine learning classifier (Shrestha et al. 2021), wetlands in Turkey using a supervised object-based classification approach (Kaplan & Avdan 2018), and creating segmentation maps for water, bare soil and rice plants in Spain using a deep leaning approach (Gargiulo et al. 2020)). The main advantages indicated are an improvement in classification accuracy and the possibility to fill data gaps caused by cloudiness.

Another instrument that has recently received a lot of attention in remote sensing-based environmental studies is the National Aeronautics and Space Administration (NASA) Global Ecosystem Dynamics Investigation (GEDI). This is a full-waveform Light Detection and Ranging (LiDAR) sensor attached to the International Space Station (ISS). GEDI has been operating since April 2019 and samples the characteristics of vegetation structure almost globally from approximately 51.6° S to approximately 51.6°N. GEDI is designed to sample information on topography, object height, and various relative height metrics (RH) (Kacic et al. 2021, Pereira-Pires et al. 2021, Dubayah et al. 2020). Therefore, it offers a chance to include additional

information in the classification process on the vertical structure of the vegetation which can be challenging to access for the tropics. The most practical and reproducible way to include GEDI data is to use analysis-ready follow-on products like the Global Forest Canopy Height Product 2019 (GFCH) (Potapov et al. 2021). The dataset combines Landsat and GEDI data to obtain elevation values for areas covered by vegetation, thereby providing a potential substitute to costly and often unavailable digital surface models in developing and emerging countries.

Combining various dataset types and increasing spatial resolution also leads an increase in complexity for the analysis. Therefore, strategies for solving issues related to mixed signal pixels are needed (Nabil et al. 2020). Object-based image analysis methods (OBIA) group neighbouring single measurement pixels into image objects that comprise real-word objects, thus taking contextual features into account (Blaschke 2010, Kucharczyk et al. 2020). Comparative studies indicate that this also holds great potential for improving accuracies in Sentinel-2-based classification attempts (Phiri et al. 2020). However, many studies on object-based classification are specifically adapted to their respective study areas and derive inconsistent results for other regions (Ma et al. 2017). An important consideration in this context is that object-based approaches typically rely on image segmentation as a crucial prerequisite, which strongly depends on the scale of the investigation (Im et al. 2014, Hao et al. 2021). While some studies focus on the determination of an optimal scale for a specific target or dataset (Qiu et al. 2016, Witharana & Civco 2014, Yang et al. 2019), heterogeneous tropical highlands, however, are often composed of areas that actually need to be addressed at different scales. For example, contiguous surface types in Nyeri's rural areas tend to be considerably smaller than those in the national parks. Consequently, not only the optimal features for characterising the area are relevant, but also an adapted segmentation strategy to meet the heterogeneous environment in Muringato and similar regions.

In this paper, we assess the potential of Sentinel-2-MSI alongside the near-globally available Sentinel-1-SAR and GFCH-datasets to be used in a predominantly automated and object-based mapping strategy for heterogeneous tropical highland landscapes. To achieve this, we (i) examine possible classification features that can be derived from these datasets and (ii) develop a classification strategy that aims at Muringato's characteristics as a heterogeneous tropical highland and could potentially be applied to similar regions.

To generate suitable data for the study area (Section 2) several pre-processing steps were undertaken on a range of datasets using the Google Earth Engine (GEE) cloud platform. This included advanced cloud masking and mosaicking for generating comprehensive optical imagery as well as advanced border noise reduction, radiometric terrain normalization, and speckle filtering for the SAR data (Section 3.1). We present the class nomenclature for our monitoring concept, generate matching sample image objects using the pre-processed data, and present validation data for our classification strategy (Section 3.2). We elaborate on possible classification features for use from the datasets available (Section 3.3) and analyse our samples from across the Nyeri County from the dry seasons in 2020 and 2021 regarding these features (Section 3.4). The findings were then incorporated in the customized object-based classification process and were applied to the Muringato sub-catchment (Section 3.5). The results of the study are presented, examined regarding their accuracy using our validation data, and briefly compared to a novel global product based on Sentinel (Section 4). We will discuss the approach presented (Section 5), which will form the basis for permanent monitoring in this region.

2 Study area

Nyeri County is approximately 80 km north of Nairobi in the highlands of central Kenya between Aberdare Range in the west and Mount Kenya in the east. The Muringato sub-catchment lies in the mid-western part of the county and forms part of the Tana Basin, which is one of Kenya's five main basins. The area encompasses parts of the Aberdare Range National Park and the Kimathi, Muhoya, Kamakwa, Kiganjo, Mukaro, and Mweiga areas comprising a total area of 225 km². The city of Nyeri borders the area to the southeast. The Locations of Nyeri County and the Muringato sub-catchment is displayed in Fig. 1.

The region is characterized by a tropical climate and two rainy seasons with long rains from March to May and short rains from October to December, and is therefore affected by frequent cloud cover. The present LULC types cover large scale farming of mostly tea and coffee and small-scale farming mainly encompassing maize and beans (Muringato Water Resource User Association 2014). There are also urban and rural built-up structures, bare land,

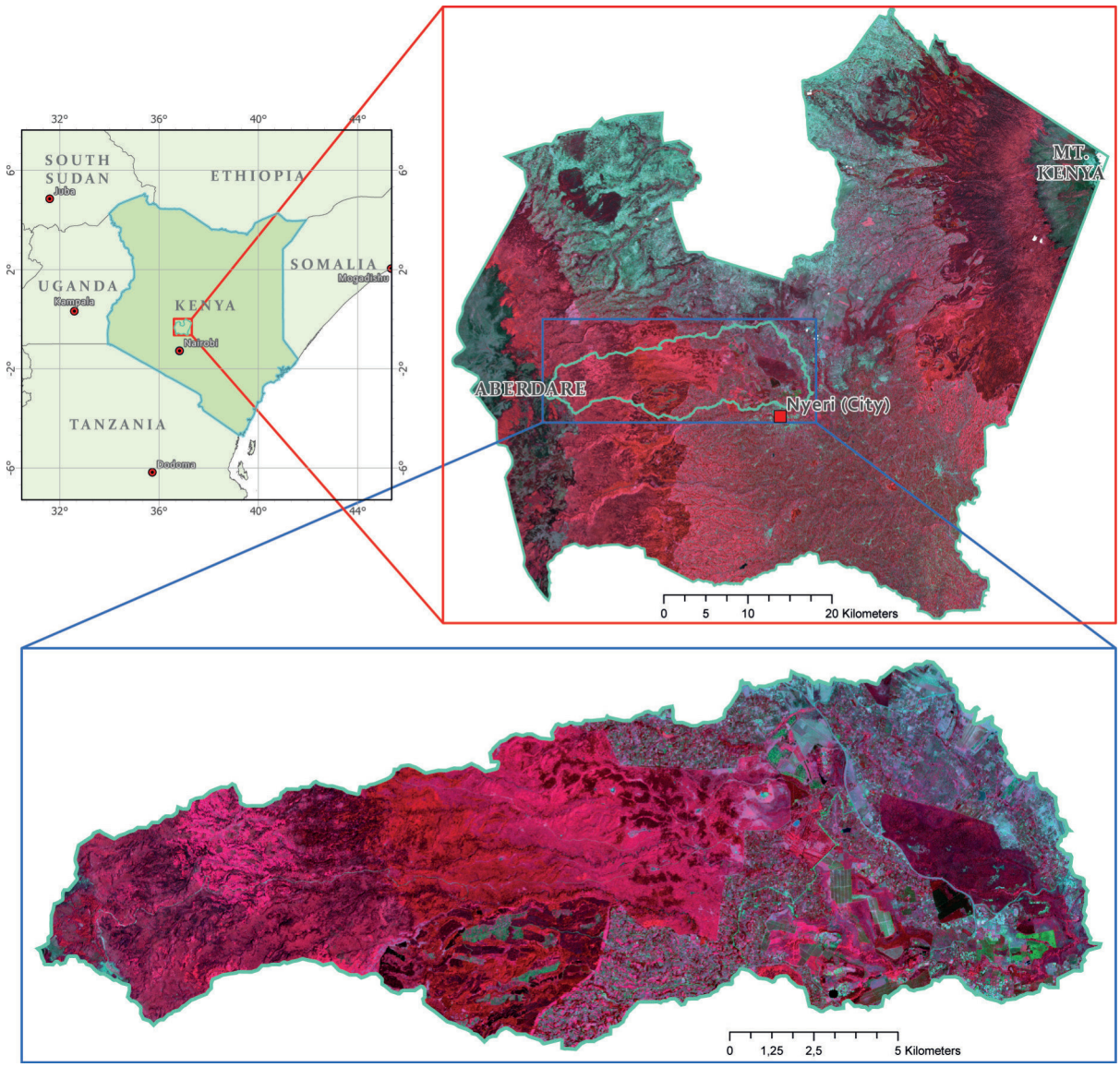

Fig. 1: Sentinel-2 image composite outlining Nyeri County and the Muringato sub-catchment in 2021 (RGB: B8, B4, B3). Administrative boundaries: Food and Agriculture Organization of the United Nations.

and open water, as well as different forest types, and shrublands and grasslands of varying sizes. This includes larger vegetation formations in the national park and smaller formations in the inhabited rural areas, which are generally structured much more heterogeneously. This composition of land cover types and the general structure of a relatively homogeneous national park or forest reserve component, either Aberdare or Mt. Kenya, and the lower lying relatively heterogeneous rural structures can also be observed in all sub-catchments of Nyeri. Therefore, we consider Muringato as a representative example of the highland region.

3 Data and methods

3.1 Data acquisition and pre-processing

Data selection and pre-processing was performed using the Google Earth Engine cloud computing platform (GEE) (Gorelick et al. 2017) which has been used successfully for environmental studies such as mapping wetlands (Hardy et al. 2020), floods (DeVries et al. 2020), and vegetation cover (Xie et al. 2019). The process is summarized in Fig. 2.

Sentinel-2-MSI scenes are available as Level-2A Bottom of Atmosphere (BOA) products in GEE, which

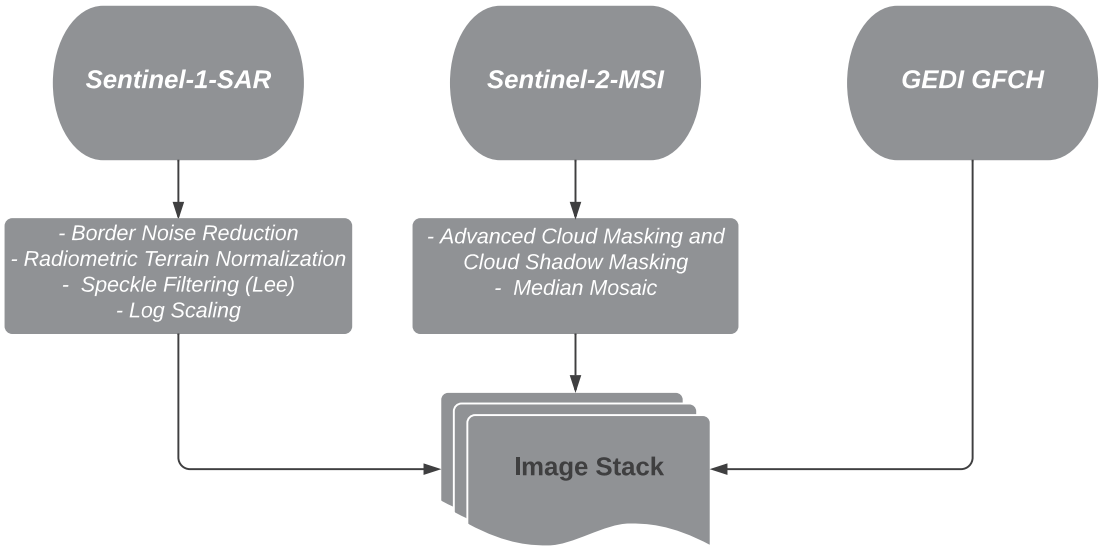

Fig. 2: Data preprocessing

are analysis-ready orthoimages corrected for atmospheric effects using the Sentinel-2-specific Sen2Cor-algorithm (Main-Knorn et al. 2017). However, frequent cloud cover still significantly affects data quality and availability in the mountainous study area even during dry seasons. To obtain comprehensive images, we fused data available data for 2020 and 2021 on a biannual basis, resulting in a total of four per-pixel-median images for the analysis of optical classification features. When possible, we used only dry seasons images with less than 20% total cloud coverage to prevent artifacts from highly cloudy, foggy, or hazy scenes. Due to exceptionally high cloud cover during the second dry season in 2021, the threshold was increased to 25% and additional rainy season images were considered to compile a spatially comprehensive dataset. All eligible scenes have been individually processed using advanced cloud and cloud shadow masking techniques as outlined in (Braaten 2022) with priority given to the removal of clouds and shadows (pixels with a cloud or shadow probability over 5%) over conserving ambiguous pixels.

Sentinel-1-SAR data in GEE are orthorectified Ground Range Detected (GRD) scenes with thermal noise removal, radiometric calibration, and terrain correction already having been applied. The values (x) are converted to decibels (dB) through log scaling (Eq. (1)):

$$dB = 10 * log_{10}(x) \quad (1)$$

As recommended for land monitoring by the ESA S-1 observation strategy, the GRD-Data used in this study was obtained in interferometric wide (IW) swath mode with 240 km swath, which supports dual-polarization with VV and VH bands (Shrestha et al. 2021). To improve the data quality and to anticipate terrain-related difficulties, an additional pre-processing framework was used as described by (Mullissa et al. 2021). Besides radiometric terrain normalization, the framework includes additional border noise reduction as well as multiple options for speckle filtering and can be directly applied in GEE. Following the suggestions of (Verhelst et al. 2021), an improved Lee Filter (Lee et al. 2009) was used in this study for speckle filtering. SAR data was selected to be in the middle of the period covered by each optical dataset, as they are intended for joint use instead of filling gaps in the rainy season.

The GFCH data is ready to use and does not require further pre-processing. Combining the datasets results in a stack of orthorectified measurements from the various sensors in gridded format. An overview of the data with their respective gridded spatial resolution is given in Table 1.

3.2 Class design, sample and validation data

Based on the available datasets, the main land cover types should be identified, namely water, tree cover, shrubs, lower vegetation (dense and sparse), built-up areas, and bare land (Tab. 2). These classes are assumed to be sufficient to form a basis for the description of changing environmental resources including deforestation, human land-use intensification, loss of waterbodies, and agricultural use once the concept is transferred into monitoring and regu-

Tab. 1: Pre-processed datasets used in this study

Spatial Resolution [m]	Band	Sensor
10	VV, VH, (VV+VH)/2	S1-SAR
	blue (B2), red (B3), red (B4), near infrared (B8)	S2-MSI
20	red edge (B5, B6, B7, B8a) short wave infrared (B11, B12)	S2-MSI
30	Global Forest Canopy Height	LiDAR (GEDI) + optical (Landsat)

lar maps are produced on a semi-annual basis. This class nomenclature is relatively general and excludes, for example, specific tree species to avoid unnecessary sources of error and ensure a level of transferability for the concept.

For our study, we created two complementary datasets aligned with this class design: a sample dataset was utilized to assess the effectiveness of potential classification features, while a validation dataset of ground truthing points was used to verify the accuracy of our classification approach in our primary study area: the Muringato sub-catchment.

The Sample dataset comprises 700 sample data objects across Nyeri County for each semi-annual image stack to cover different periods, allowing for a meaningful number of samples while keeping spatial bias low, so that the findings can, in principle, also be applied to other catchments. These objects were obtained by manually delimiting and visually interpreting the optical data, along with the aid of higher resolution imagery from Google Earth. Consequently, a total of 2,800 samples were obtained for feature evaluation. Since this work was carried remotely, only objects that could be clearly identified and delimited in the image material were considered as the risk of incorrect assignment of classes due to subjective errors was considered too great. Therefore, we also did not collect samples for transitional classes 'Sparse vegetation' and 'Shrubs', as these are not clearly identifiable in the images. Instead, we assume that they fall between the 'Tree cover' and 'Low vegetation' classes or the vegetation classes and non-vegetation, respectively.

Tab. 2: Target class descriptions.

Landcover Type	Description
Tree cover	forest, dense tree crowns
Shrubs	larger shrubs and bushes, young trees, shrub-like crops (e.g., coffee)
Low vegetation	dense grass, vital low grown crops (e.g., tea), forbs, herbs, very small shrubs, young secondary vegetation
Sparse vegetation	low vegetation cover, crops with lesser photosynthetic activity (e.g., young maize, potatoes, beans)
Bare land	open soil, bare or freshly cultivated cropland, rocks, quarries, minor sealed surfaces, burnt areas
Built-up	settlements, buildings, larger sealed surfaces
Water	water holes, water reservoirs, sewage treatment plants

For validation purposes, an additional total of 556 independent and randomly distributed ground truthing data points using the complete class nomenclature were collected across the Muringato sub-catchment area during a fieldwork campaign in early December 2021. Ground Truthing points in restricted areas that were not accessible during this work were interpreted using a Maxar orthoimage with a resolution of 50 cm as a substitute. This affects largely the western part of the area that is part of the Aberdare National Park.

3.3 Possible classification feature overview

Given that the aggregation of optical data from several scenes can also be expected to introduce some smoothing, it is particularly important to derive robust features that work effectively with the median mosaics. Since there are no specific investigations on this matter, different indices and specific band combinations given in the literature were considered.

Numerous water bodies in Nyeri County are of small size and highly dynamic including water holes with fluctuating levels composed of only a few pixels. This requires a highly robust approach for successful capture. Methods for water detection predominantly require the use of specific indices such as the Normalized Difference Water Index (NDWI) and the Modified Normalized Difference Water Index (MNDWI) (Xu 2006, Du et al. 2016, Sekertekin et al. 2018)) or sensor-specific variants including the Sentinel-2 Water Index (SWI) (Jiang et al. 2020, Jiang et al. 2021). These indices highlight water by relating two spectral bands and are calculated as shown in Eq. (2), (3) and (4):

$$NDWI = B_3 - B_8 \,/\, B_3 + B_8 \tag{2}$$

$$MNDWI = B_3 - B_{11} \,/\, B_3 + B_{11} \tag{3}$$

$$SWI = B_5 - B_8 \,/\, B_5 + B_8 \tag{4}$$

Sentinel-1-SAR can also be used for rapid mapping under cloudy conditions because of the relatively low backscatter from the waterbodies (Nasirzadehdizaji et al. 2019, Zheng & Shao 2018)).

The presence of vital vegetation is most widely tracked using the Normalized Difference Vegetation Index (NDVI) in remote sensing. Compared with other indices, it is sensitive to soil and shadow influences and has a sensitive response even for low vegetation cover (Xue & Su 2017). Therefore, it can also be used to capture sparse vegetation cover. The formula is shown in Eq. (5).

$$NDVI = B_8 - B_4 \,/\, B_8 + B_4 \tag{5}$$

To separate the tree cover from the low vegetation, the Sentinel-2-MSI bands B2, B3, B6 and B12 are expected to hold the most valuable spectral information (Ottosen et al. 2020).

Distinguishing open ground from built-up areas is one of the most challenging tasks in remote sensing due to their high spectral similarity. This is especially true for the rural areas in Nyeri, where built-up structures are frequently small and scattered, making them highly inconspicuous and difficult to detect. In addition to the SAR and plain MSI-data, several optical indices for mapping impervious surfaces that are discussed in comparative studies (Chen et al. 2020, Ettehadi Osgouei et al. 2019, Li et al. 2021)) were considered. The performance of the respective indices can vary considerably depending on the region and the types of soil present (Chen et al. 2020), which can also be affected by seasonal variation (Li et al. 2021). Therefore, in addition to the widely used Normalized Difference Built-Up Index (NDBI), the most promising novel indices have been selected regarding their performance in the respective study and calculated for the research area. These are the Normalized Difference Tillage Index (NDTI) (Ettehadi Osgouei et al. 2019) and the Enhanced Normalized Difference Impervious Surfaces Index (ENDISI) (Chen et al. 2019). The formulas for the NDBI and NDTI are given in Eq. (6) and (7).

$$NDBI = B_{11} - B_8 \,/\, B_{11} + B_8 \tag{6}$$

$$NDTI = B_{11} - B_{12} \,/\, B_{11} + B_{12} \tag{7}$$

The ENDISI uses a more complex formula given in Eq. (8)

$$ENDISI = \frac{B_2 - \alpha * \left[\frac{B_{11}}{B_{12}} + (MNDWI)^2\right]}{B_2 + \alpha * \left[\frac{B_{11}}{B_{12}} + (MNDWI)^2\right]} \tag{8}$$

with a correction factor α using the scene mean values (Eq. (9))

$$\alpha = \frac{2 * B_{2\,mean}}{\left(\frac{B_{11}}{B_{12}}\right)_{mean} + \left[(MNDWI)^2\right]_{mean}} \tag{9}$$

and the previously described MNDWI (Eq. (3)).

3.4 Evaluation of classification feature candidates

To investigate the performance of the potential classification features, we evaluated our reference data from four datasets comprising a total of 2,800 samples regarding the spread of the mean image object values on a z-score standardized value range. Particular attention is paid to the different options for detecting water, the separability of the tree cover from the low vegetation and of built-up areas to bare land because these surface types are particularly relevant or expected to be particularly difficult to separate.

When comparing these options, we made the following observations in our reference data:

For water masking (also see Fig. 3), the SWI-Index outperforms other optical indices because water has the smallest overlaps, especially with the built-up class when the outliers are taken into account. The SAR backscatter was also found to be relatively low. However, this is also true for other

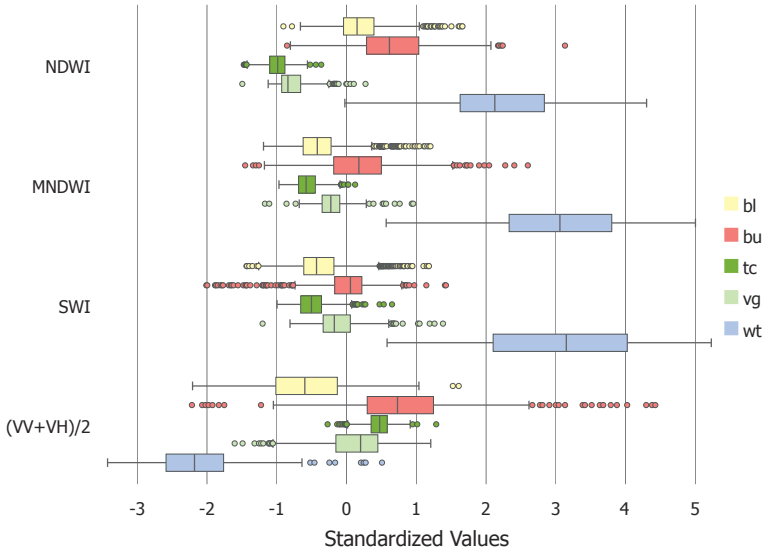

Fig. 3: Boxplots for water indices and SAR-backscatter according to class (bl: bare land; bu: built-up; tc: tree cover; vg: low vegetation; wt: water)

surfaces such as bare land, which might also represent dry or low level water holes. However, since extremely low backscatter values certainly indicate water, this can be used as a supplementary criterion. According to our observations, this is occasionally helpful in shaded areas which in a few cases are falsely identified as water. A combination of SWI and SAR-backscatter therefore represents a promising solution with only a few overlapping objects remaining in the feature space.

For tree cover detection (see also Fig. 4), the sample data show that the tree covered areas tend to be less responsive in the relevant spectral ranges compared with lower vegetation. If the spectral values are compared with the canopy height values in the samples from the vegetation and tree cover classes, the presence of a relatively strong correlation (R2 = 0.72) can also be confirmed. Therefore, GFCH data is likely to be suitable as a complementary or additional data source when the colours falsely indicate the presence of tree cover. This is conceivable, for example, in the case of certain agricultural plants or terrain-related shading effects. SAR-backscatter patterns are comparatively ambiguous in these cases. However, dB values of less than approximately - 13 dB only occur with low vegetation and therefore represent a potential exclusion criterion for non-tree-covered areas.

In terms of detection of built-up areas (also see Fig. 5), the NDBI shows no significant separability,

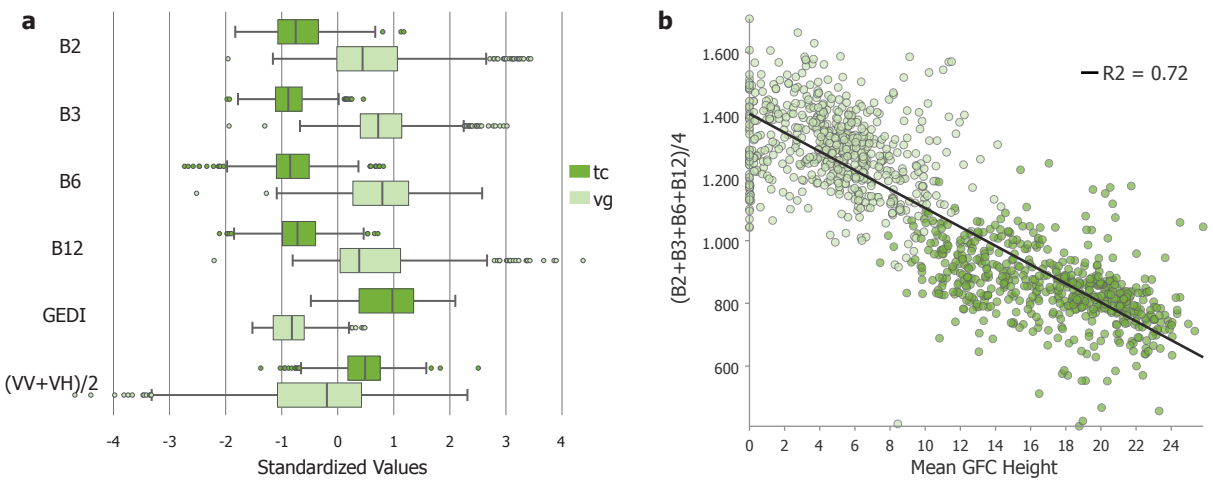

Fig. 4: (a) Boxplots for tree cover and low vegetation for relevant features and (b) scatterplot showing the correlation between the optical and GFCH data (tc: tree cover; vg: low vegetation)

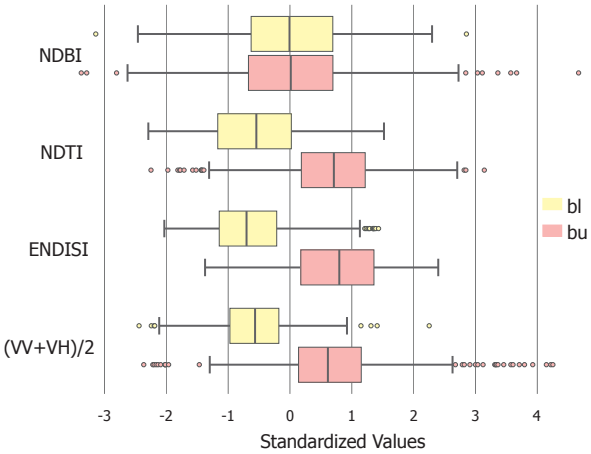

Fig. 5: Boxplots for built-up-indices and SAR-backscatter according to class (bl: bare land; bu: built-up)

while the SAR backscatter tends to be higher on built-up structures. However, it is also subject to strong noise and outliers. High extrema indicate the presence of buildings. A visual interpretation of the data shows that many of the smaller buildings cannot be seen in SAR. The ENDISI and NDTI indices show potential for extraction given that the interquartile ranges for the built-up area and bare soil classes do not overlap. According to our sample data, the ENDISI is preferable due to fewer outliers and a larger distance for the class medians and the interquartile ranges on the standardised scale.

3.5 Classification strategy

To meet the requirements of the local conditions in Nyeri and Muringato, we developed a sequential classification scheme using Trimble eCognition software. In this study, primary image segmentation was performed using the popular multi-resolution segmentation, which merges single pixels into objects based on a local homogeneity criteria (Baatz & Schäpe 2000). These contain a parent scale parameter and a weight of shape against colour as well as a weight of compactness against smoothness. Authors suggest different settings for Sentinel-2-data based on the individual task including for mapping wetlands (Kaplan & Avdan 2018) or for mapping large-scale cropland (Belgiu & Csillik 2018). We found that none of the setups in the literature met the requirements of our study area because the more general target classes can occur at multiple scales including small- and large-scale farming, rural single housing, and larger settlements or small trees in stands in rural areas and larger forests in national parks and forest reserves.

Therefore, we propose a sequential application at multiple scales while checking whether an object is homogenous regarding important classification features or needs further refinement. To do so, we chose to evaluate the interquartile ranges of the relevant features along with the difference in quartiles for the respective mean values. With this method, segments that contain multiple classes can be identified while excluding extrema that are deliberately suppressed, such as shadows in forests, by using an object-based approach. Another aim of this multi-scale approach is to evaluate large areas such as large scale agriculture and forests stands which have higher measured values differentiated from small, but also relevant objects such as detached single buildings and small-scale agriculture, whose demarcation is limited by the geometric resolution of the data. Fig. 6 shows an example of this re-evaluation process using an object homogeneity criterion. Our segmentation uses all Sentinel-2 bands with a resolution of 10 meters and starts with a scale of 40, which we consider to be suitable for the larger homogeneous objects. Inhomogeneous objects are then further segmented with a scale of 10.

In this case, the rural areas with subsistence farming and smaller dwellings are grouped into objects, and are under-segmented, while the larger vegetation structures and agricultural areas were correctly segmented. By checking the homogeneity using the NDVI and ENDISI, it can be determined if there are likely to be different target classes like buildings, vegetation, or bare land composed within a single segment. These inhomogeneous image objects can then be flagged accordingly and sub-segmented at a finer scale.

At the finest scale we suggest an additional local contrast analysis because it is expected that even relatively low values can indicate an object if it shows a certain contrast and gradient with its surroundings. In Muringato, this concerns the detection of small individual buildings whose size is often on the borderline for geometric resolution. Although the ENDISI values for image objects are high in relation to their size, the absolute values are frequently still lower than, for example, larger dry arable areas but still predominantly represent a local extreme. Therefore, the local contrast from ENDISI is also used as a classification feature with an example being shown in Fig. 7.

The image stack is then classified using a sequence of masks with the classification features shown in Fig. 8. The classification nomenclature allows for an intermediate classification into the

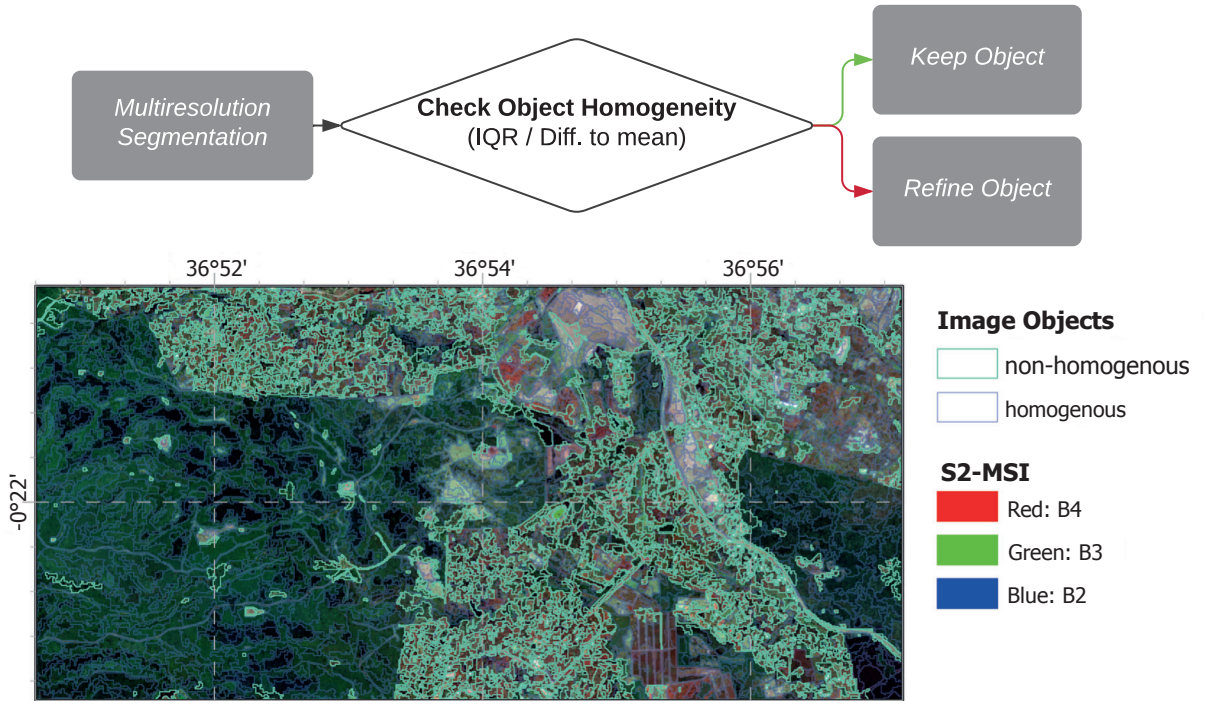

Fig. 6: Image objects flagged as inhomogeneous at a scale factor of 40

generalized parent classes. By doing so, we expect to minimize omission and commission errors between subclasses that compose certain parent classes like vegetation and non-vegetation.

These masks can either be applied directly using the image objects mean values for homogeneous image objects or in conjunction with a multi-threshold-segmentation, which further subsegments objects identified as heterogeneous at the value edges.

Due to the smoothing and inconsistencies that arise when creating per-pixel-median images, we perform the masking in a supervised process. The thresholds in this study are chosen exploratory. This is because they are expected to differ depending on the season, the time range and the number of images processed into the underlying image mosaic.

For all the classification steps, there are fuzzy transition areas in which objects cannot be clearly assigned. For vegetation, we use 'sparse vegetation' and 'shrubs' as intermediate classes. The former marks the transition between vegetation and non-vegetation, i.e., areas that have too high a NDVI to

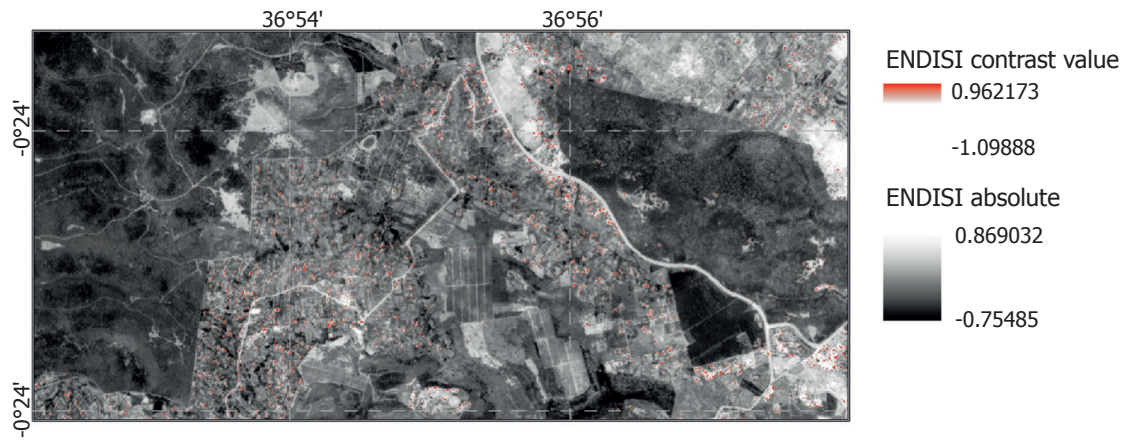

Fig. 7: ENDISI-based building detection using absolute and contrast values

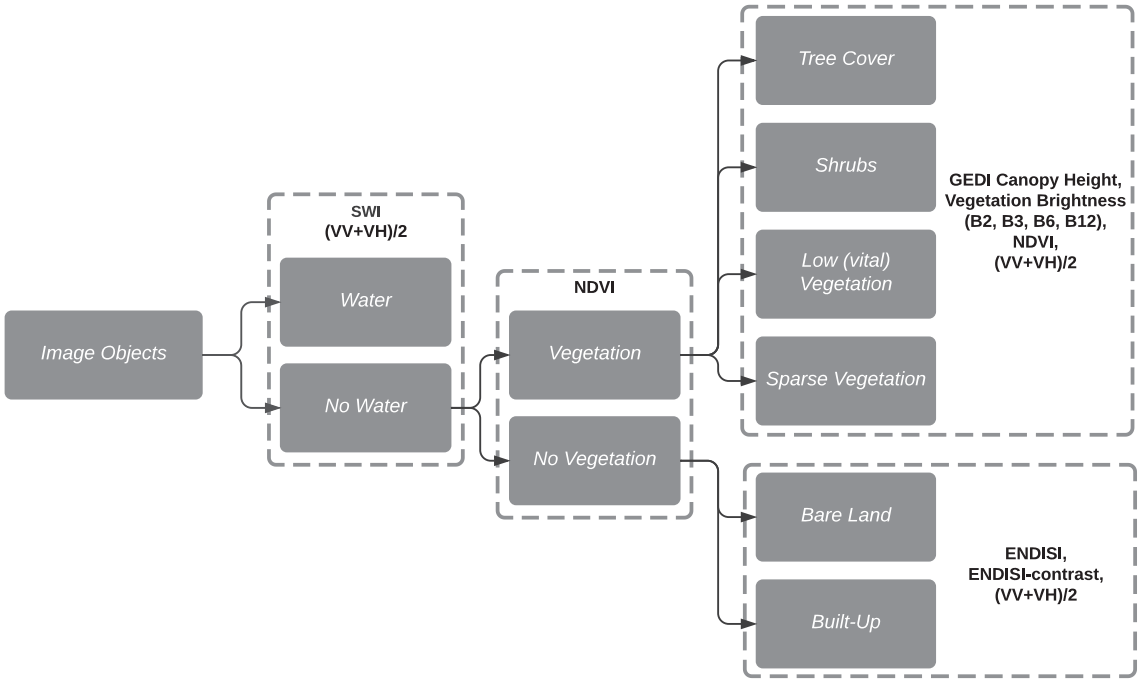

Fig. 8: Masks applied by thresholding the given features

be clearly classified as non-vegetation and too low a NDVI to be clearly classified as vegetation. The latter marks the transition between tree cover and low vegetation. Shrubs are assumed to lie between trees and low vegetation in terms of object height and colour values. To classify the remaining ambiguous image objects in all other steps where there is no meaningful semantic intermediate class, we chose to use a k-nearest Neighbours (kNN) machine learning classifier using the unambiguous objects as reference. kNN compares the unknown object against training data and assigns them to the class of the nearest trainings samples within the feature space (Maxwell et al. 2018). In this study, kNN is performed using Sentinel-2 B2, B3 and B4 bands in addition to the respective class properties.

4 Results and accuracy assessment

Based on the literature review and our findings from the classification feature evaluation, we performed the LULC classification for the Muringato sub-catchment using the feature combination for the class masks as summarised in Tab. 3. Due to the findings from our reference data, raw SAR-backscatter is used as an additional exclusion criterion for the tree cover, water, and built-up masks to identify false positives, while a combination of NDVI, ENDISI, and SWI indices or a vegetation-specific band combination (B2, B3, B6, B12) and GFCH data provide the most promising classification basis. Fig. 9 shows the classification results for the Muringato sub-catchment median composite of the second half of 2021.

Tab. 3: Suitable classification features based on the literature review and analysis of 2,800 samples

Mask	Features
Water / Non-water	SWI, (VV+VH)/2
Tree cover / Shrub / Low vegetation	GFCH, Vegetation Brightness (B2, B3, B6, B12), (VV+VH)/2
(vital) Vegetation / Sparse vegetation / Non vegetation	NDVI
Built-up	ENDISI, ENDISI-contrast, (VV+VH)/2

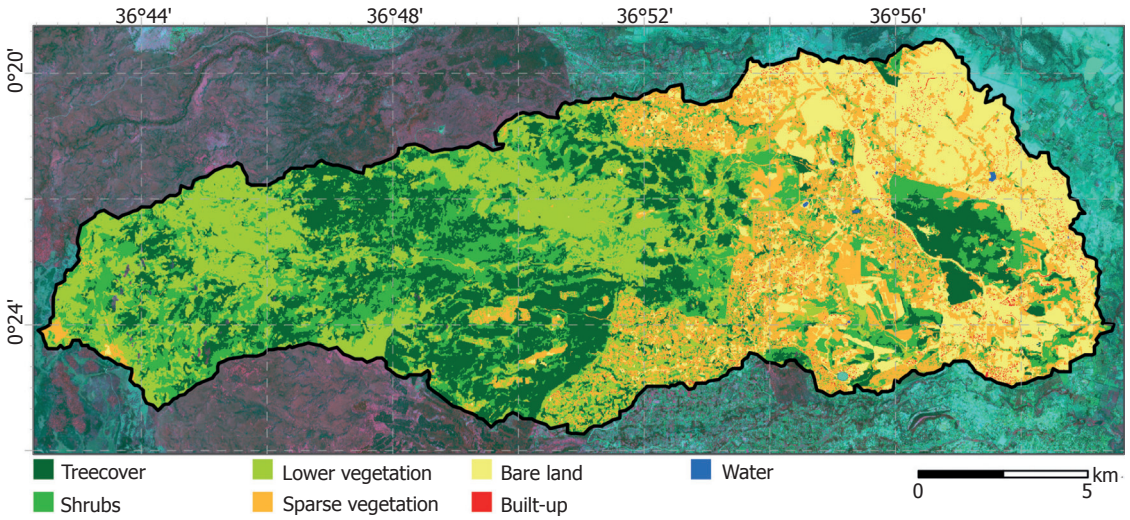

Fig. 9: Classification results for the Muringato sub-catchment 2021 (second half of the year)

Accuracy Assessment was performed using a total of 556 independent validation points. A statistical comparison between the predicted classes and the reference classes is given in Tab. 4.

The overall accuracy is estimated to be approximately 88% with a kappa coefficient of 0.85. Confusion generally occurs primarily between the different vegetation classes and the spectrally similar non- or sparsely vegetated classes, which dominated the eastern part of Muringato sub-catchment under the particularly dry conditions during the study period. This is also visible in the corresponding optical imagery.

Given that the aim of this study was also to generate alternatives to global products for the study area, it is also useful to make a brief comparison. A visual comparison with the relatively novel ESA WorldCover 10 m product for 2020 (Zanaga et al. 2021) also shows a considerably higher level of detail recognition compared to the global products available (see Fig. 10).

Tab. 4: Confusion matrix

	Tree cover	Shrubland	Low vegetation	Sparse vegetation	Bare land	Built -up	Water	Total	Users	Kappa
Tree cover	90	13	0	0	0	0	0	103	0.874	
Shrubland	1	115	3	0	0	0	0	119	0.966	
Low vegetation	0	14	95	2	0	0	0	111	0.856	
Sparse vegetation	2	10	1	66	3	5	0	87	0.759	
Bare land	0	0	0	7	67	5	0	79	0.848	
Built-up	0	0	0	0	1	39	0	40	0.975	
Water	0	0	0	0	0	0	17	17	1	
Total	93	152	99	75	71	49	17	556	0	
Producers	0.968	0.756	0.9595	0.88	0.944	0.796	1	0	**0.879**	
Kappa										**0.855**

Even though a complete statistical assessment is of limited significance due to the different class nomenclature, only 16 out of 45 buildings in our validation data are present in the ESA WorldCover dataset. In addition, 468 validation points are classified as being tree covered, shrubland, or grassland in the global dataset while only 341 are classified as tree cover, shrubland or lower vegetation in our classification. This indicates a considerable overestimation of vegetation cover compared with our LULC. 51 of our validation points are located on cropland in the global dataset. These could represent both vegetation and non-vegetation depending on the state of growth and have been excluded for the purposes of this comparison.

5 Discussion and conclusion

Tropical highlands remain a challenging target for LULC due to their high heterogeneity of the landscape and frequent cloud cover, causing a shortage of high quality and reliable comprehensive data. With this paper, we investigated the usability of GEE-processed Sentinel-1-SAR, Sentinel-2-MSI and GFCH data by analysing the classification features derived from these datasets and performed an adapted multi-scale classification scheme for complex tropical highland regions.

Analysing 2,800 sample data objects from four semi-annual datasets after undertaking the literature review, we found the SWI- and ENDISI-indices to be highly promising features for water and building extraction. Meanwhile, a relatively differentiated determination of vegetation types can be achieved by combining the optical data selected (B2, B3, B6 and B12 bands of Sentinel-2-MSI) with the CFCH information. The robustness of the optical features is important given the high number of images usually required to compose a cloud free mosaic in tropical highlands and to introduce a certain degree of smoothing due to temporal variation. The fact that some of these feature combinations cover wavelength ranges that other comparable freely available satellites lack already indicates a particular utility for Sentinel-2 in the region. Moreover, its relatively high resolution allows enables meaningful analysis of the diverse rural area in the highlands, as demonstrated through the use of local contrasts in classifying built-up structures.

The SAR data was only used in a complementary role in this study because we could not see a clear demarcation of our target classes in the reference data with the values strongly scattering. Many objects of interest are not visible or are barely visually recognisable in the data and could not be constantly recorded. This is especially true for smaller objects including stand-alone buildings but was also the case for many vegetation areas which can be clearly recognised in the corresponding optical features but not in the SAR data (see Fig. 11).

Other authors have achieved high levels of accuracy for certain tasks using SAR in other regions. However, a priori information is frequently used for masking irrelevant areas beforehand (Dostálová et al. 2016, Khabbazan et al. 2019, Ygorra et al. 2021). These datasets are not always available, especially in developing countries. In this study, Sentinel-1-SAR-implementation was performed using GRD data from GEE with the aim of deriving direct classification features to be used within our object-based classification approach. Other authors

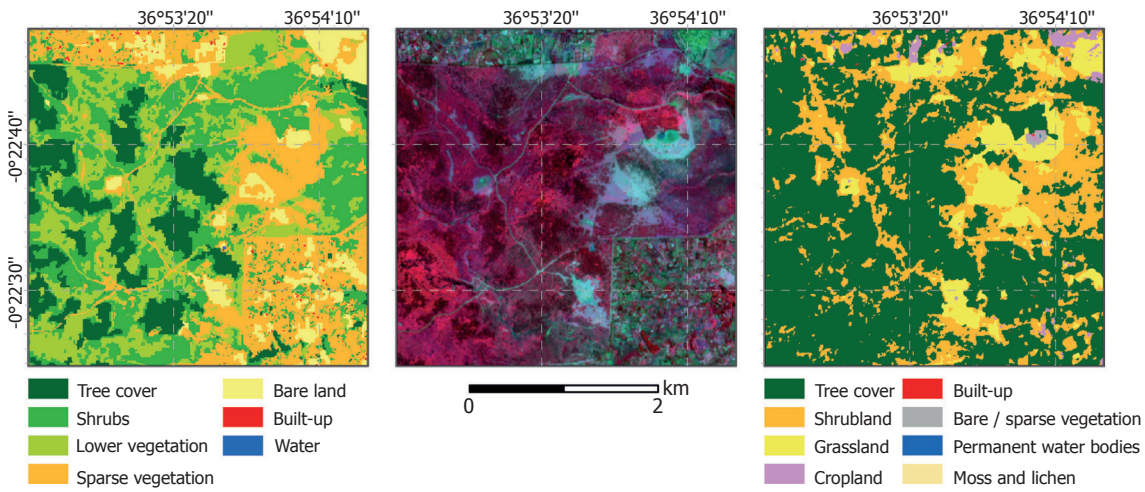

Fig. 10: Comparison of the classification results (left), optical data (middle) and the ESA WorldCover product (right)

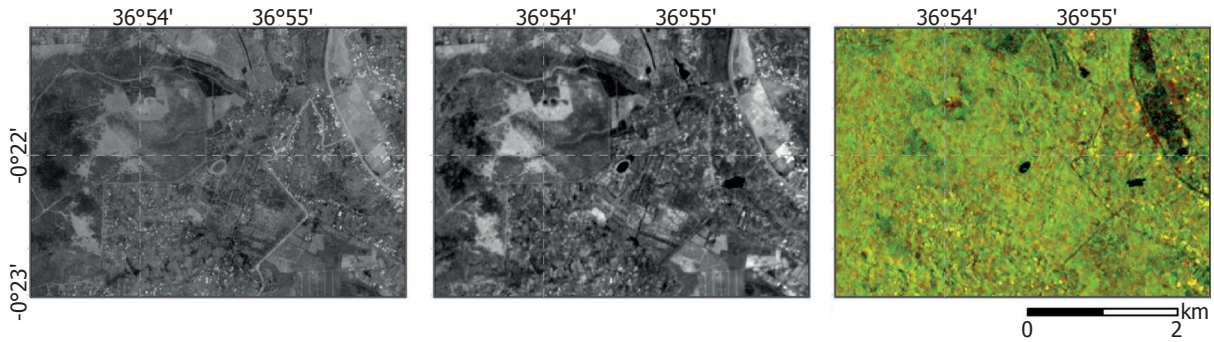

Fig. 11: ENDISI (left), B2-B3-B6-B12-band average (middle) and Sentinel-1-GRD (right; RGB: VV, VH, ratio)

suggest using the coherence of multiple observations from Sentinel-1-SLC (Single Look Complex) in a time series (Borlaf-Mena et al. 2021, Jacob et al. 2020, Nikaein et al. 2021) but SLC-data is not available in GEE. In this case, its cloud computing functions would have to be dispensed with. Sentinel-1-SAR-backscatter could also be used for sequential change detection (Canty et al. 2020) instead of the classification features. This could be used to visualise changes independent of the explicit land use type, which may be helpful given that some classes such as buildings can be more persistent than others.

Because it provides the highest resolution among the freely available datasets, Copernicus data was deliberately used for this work. Other SAR data using bands at different wavelengths, such as TerraSAR-X using X-band, or polarizations sources may be considered. The use of quad-polarized sensors would allow for the calculation of a full quad-polarized Radar Vegetation Index (RVI) (Dinesh Kumar et al. 2013, Kim et al. 2014) in comparison with Sentinel-1 dual-polarization. Kim et al. (2012) found L-band derived RVI to be well correlated with several optical vegetation indices using a ground-based radar system. An L-band based SAR satellite mission 'Tandem L' with high temporal and spatial resolution is planned by the German Aerospace Centre (Krieger et al. 2009).

The GEDI data used in this study are solely from the Global Canopy Height Model 2019 (Potapov et al. 2021), and also subject to all the inaccuracies related to its genesis, such as artefacts and cloud gaps caused by the optical component (Verhelst et al. 2021) and will also involve some temporal deviations. The spatial resolution is not entirely optimal, given that they are less accurate than the Sentinel-2 data that is used to build image objects in this study. This gap in resolution could be closed with future products with a higher resolution from fusion with other sensors (Qi et al. 2019, Kacic et al. 2021). A future direct integration of Sentinel 2 data is also planned for the Global Canopy Height product, which is also expected to improve thematic accuracy (GLAD 2022). However, being the only ready-to-use product at the time of this study, the GFCH product was considered the most practical and reproducible option for method development until other full or improved products are released.

In our sample data, the canopy height values show a relatively high correlation to the colour values that also separate the tree cover and low vegetation. We noticed a helpful effect when certain agricultural crops such as coffee, which appear darker than other vegetation areas and would otherwise be classified as 'tree-covered', are correctly classified by the GCHM integration. However, given that we did not subdivide specific forms of lower vegetation besides sparse vegetation via NDVI, we cannot quantify this observation.

Some aspects could also be altered in terms of the methods used. Smoothing data from calculating a median composite is a potential source of uncertainties. Another option that could be considered is to classify the images individually and then merge the results. We also did not include a change detection or time series across multiple semi-annual products. This could contribute substantially to the accuracy given that the scene we evaluated as an example is in a period with low levels of vegetation, which is particularly susceptible to confusion between sparse vegetation and non-vegetation. By adding a temporal component by comparing individual classification results, additional classification depth including agriculture could also be incorporated and might also improve the accuracies as indicated by (Schulz et al. 2021). This could not be included in this study due to the lack of multi-year validation data and is planned as a further step in the operational implementation of the concept.

The customized multi-scale classification strategy we propose is designed to meet the special properties of heterogeneous tropical highland landscapes and achieved an 87.9% accuracy in the Muringato sub-catchment. This is comparable to other studies in East Africa, which use different data, such as Landsat time series and a range of ancillary datasets (87.7%) (Mariye et al. 2022), or combined Landsat and RapidEye data (four maps ranging from 85.7% to 93.2%) (Kindu et al. 2013). However, some of these data are lower resolution or less accessible. Schulz et al. (2021) used comparable seasonal Sentinel-1 and 2 data stacks in a heterogeneous environment in Niger for machine learning approaches, achieving up to 73.3%. This allows for a higher degree of automation but suggests that this comes at the expense of accuracy. Also, although this is not statistical evidence, a visual comparison and cross-checking with our validation data suggests a noticeable improvement over current available global products. This manifests itself primarily in the large-scale and heterogeneous areas that are not represented with sufficient differentiation in global products for catchment-level tasks. Here, an object-based approach in combination with a homogeneity assessment of the areas under consideration and a local contrast analysis for particularly critical classes is likely to be particularly helpful. Therefore, we conclude that multi sensor earth observation data is suitable for improving LULC in an object-based approach despite the difficult conditions in these regions. Due to the transferable class design, we think that this classification strategy could be applicable with minor adjustments for similar regions. However, this remains to be evaluated given that resources for the validation data in this study were limited to Muringato.

Findings from this paper will be incorporated into a regional environmental monitoring strategy, including a Water and Energy Budget-based Distributed Hydrological Model (WEB-DHM) (Wang et al. 2009a, Wang et al. 2009b) that uses LULC information as an input.

Acknowledgements

We would like to thank the German Federal Ministry of Education and Research (01DG20022) for funding this research within the project 'CITGI4Muringato'. The corresponding author would also like thank Johanna A. Wanjala, Simon Muthee, Wisdom Kipkemboi and Anna Bartels for their support in capturing and processing validation data.

References

Baatz M, Schäpe A (2000) Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. Strobl J (ed) *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg 2000 XII:* 12–23. Salzburg.

Bégué A, Leroux L, Soumaré M, Faure J-F, Diouf AA, Augusseau X, Touré L, Tonneau J-P (2020) Remote sensing products and services in support of agricultural public policies in Africa: Overview and challenges. *Frontiers in Sustainable Food Systems* 4: 58. https://doi.org/10.3389/fsufs.2020.00058

Belgiu M, Csillik O (2018) Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment* 204: 509–523. https://doi.org/10.1016/j.rse.2017.10.005

Blaschke T (2010) Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Borlaf-Mena I, Badea O, Tanase MA (2021) Assessing the Utility of Sentinel-1 coherence time series for temperate and tropical forest mapping. *Remote Sensing* 13: 4814. https://doi.org/10.3390/rs13234814

Braaten J (2022) Sentinel-2 cloud masking with s2cloudless. https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless (date: 24.02.2022).

Canty MJ, Nielsen AA, Conradsen K, Skriver H (2020) Statistical analysis of changes in Sentinel-1 time series on the Google Earth Engine. *Remote Sensing* 12: 46. https://doi.org/10.3390/rs12010046

Chen J, Chen S, Yang C, He L, Hou M, Shi T (2020) A comparative study of impervious surface extraction using Sentinel-2 imagery. *European Journal of Remote Sensing* 53: 274–292. https://doi.org/10.1080/22797254.2020.1820383

Chen J, Yang K, Chen S, Yang C, Zhang S, He L (2019) Enhanced normalized difference index for impervious surface area estimation at the plateau basin scale. *Journal of Applied Remote Sensing* 13: 1. https://doi.org/10.1117/1.JRS.13.016502

DeVries B, Huang C, Armston J, Huang W, Jones JW, Lang MW (2020) Rapid and robust monitoring of flood events using Sentinel-1 and Landsat data on the Google Earth Engine. *Remote Sensing of Environment* 240: 111664. https://doi.org/10.1016/j.rse.2020.111664

Dinesh Kumar S, Srinivasa Rao S, Sharma JR (2013) Radar vegetation index as an alternative to NDVI for monitoring of soyabean and cotton. *Indian Cartographer* 33: 91–96

Dostálová A, Hollaus M, Milenković M, Wagner W (2016) Forest area derivation from Sentinel-1 data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* III-7: 227–233. https://doi.org/10.5194/isprs-annals-III-7-227-2016

Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, Hoersch B, Isola C, Laberinti P, Martimort P, Meygret A, Spoto F, Sy O, Marchese F, Bargellini P (2012) Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment* 120: 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Du Y, Zhang Y, Ling F, Wang Q, Li W, Li X (2016) Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by Sharpening the SWIR band. *Remote Sensing* 8: 354. https://doi.org/10.3390/rs8040354

Dubayah R, Blair JB, Goetz S, Fatoyinbo L, Hansen M, Healey S, Hofton M, Hurtt G, Kellner J, Luthcke S, Armston J, Tang H, Duncanson L, Hancock S, Jantz P, Marselis S, Patterson PL, Qi W, Silva C (2020) The global ecosystem dynamics investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing* 1: 100002. https://doi.org/10.1016/j.srs.2020.100002

Ettehadi Osgouei P, Kaya S, Sertel E, Alganci U (2019) Separating built-up areas from bare land in Mediterranean cities using Sentinel-2A imagery. *Remote Sensing* 11: 345. https://doi.org/10.3390/rs11030345

Gargiulo M, Dell'Aglio DAG, Iodice A, Riccio D, Ruello G (2020) Integration of Sentinel-1 and Sentinel-2 data for land cover mapping using W-Net. *Sensors* 20: 2969. https://doi.org/10.3390/s20102969

GLAD (Global Land Analysis and Discovery) (2022) Global forest canopy height, 2019. https://glad.umd.edu/dataset/gedi (date: 03.07.2022)

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* 202: 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hao S, Cui Y, Wang J (2021) Segmentation scale effect analysis in the object-oriented method of high-spatial-resolution image classification. *Sensors* 21: 7935. https://doi.org/10.3390/s21237935

Hardy A, Oakes G, Ettritch G (2020) Tropical wetland (TropWet) mapping tool: The automatic detection of open and vegetated waterbodies in Google Earth Engine for tropical wetlands. *Remote Sensing* 12: 1182. https://doi.org/10.3390/rs12071182

IGGReS (Dedan Kimathi University of Technology Institute of Geomatics, GIS and Remote Sensing) (2018) Report on baseline mapping and documentation of Muringato WRUA sub-catchment characteristics. Basic Inventory. Nyeri.

Im J, Quackenbush LJ, Li M, Fang F (2014) Optimum scale in object-based image analysis. Weng Q (ed) Scale issues in remote sensing 197–214. Hoboken, New Jersey. https://doi.org/10.1002/9781118801628.ch10

Jacob AW, Vicente-Guijalba F, Lopez-Martinez C, Lopez-Sanchez JM, Litzinger M, Kristen H, Mestre-Quereda A, Ziolkowski D, Lavalle M, Notarnicola C, Suresh G, Antropov O, Ge S, Praks J, Ban Y, Pottier E, Mallorqui Franquet JJ, Duro J, Engdahl ME (2020) Sentinel-1 InSAR coherence for land cover mapping: A comparison of multiple feature-based classifiers. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13: 535–552. https://doi.org/10.1109/JSTARS.2019.2958847

Jiang W, Ni Y, Pang Z, He G, Fu J, Lu J, Yang K, Long T, Lei T (2020) A new index for identifiying water body from Sentinel-2 satellite remote sensing imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-3-2020: 33–38. https://doi.org/10.5194/isprs-annals-V-3-2020-33-2020

Jiang W, Ni Y, Pang Z, Li X, Ju H, He G, Lv J, Yang K, Fu J, Qin X (2021) An effective water body extraction method with new water index for Sentinel-2 imagery. *Water* 13: 1647. https://doi.org/10.3390/w13121647

Kacic P, Hirner A, Da Ponte E (2021) Fusing Sentinel-1 and -2 to model GEDI-derived vegetation structure characteristics in GEE for the Paraguayan Chaco. *Remote Sensing* 13: 5105. https://doi.org/10.3390/rs13245105

Kaplan G, Avdan U (2018) Sentinel-1 and Sentinel-2 data fusion for wetlands mapping: Balikdami, Turkey. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-3: 729–734. https://doi.org/10.5194/isprs-archives-XLII-3-729-2018

Khabbazan S, Vermunt P, Steele-Dunne S, Ratering Arntz L, Marinetti C, van der Valk D, Iannini L, Molijn R, Westerdijk K, van der Sande C (2019) Crop monitoring using Sentinel-1 data: A case study from the Netherlands. *Remote Sensing* 11: 1887. https://doi.org/10.3390/rs11161887

Kim Y, Jackson T, Bindlish R, Lee H, Hong S (2012) Radar vegetation index for estimating the vegetation water content of rice and soybean. *IEEE Geoscience and Remote Sensing Letters* 9: 564–568. https://doi.org/10.1109/LGRS.2011.2174772

Kim Y-H, Oh J-H, Kim Y-I (2014) Comparative analysis of the multispectral vegetation indices and the radar vegetation index. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* 32: 607–615. https://doi.org/10.7848/ksgpc.2014.32.6.607

Kindu M, Schneider T, Teketay D, Knoke T (2013) land use/land cover change analysis using object-based classification approach in Munessa-Shashemene landscape of the Ethiopian highlands. *Remote Sensing* 5: 2411–2435. https://doi.org/10.3390/rs5052411

Krieger G, Hajnsek I, Papathanassiou K, Eineder M, Younis M, Zan F de, Prats P, Huber S, Werner M, Fiedler H, Freeman A, Rosen P, Hensley S, John-

son W, Veilleux L, Grafmueller B, Werninghaus R, Bamler R, Moreira A (2009) The Tandem-L mission proposal: Monitoring earth's dynamics with high resolution SAR interferometry. *2009 IEEE Radar Conference, Pasadena, CA, USA*: 1–6. https://doi.org/10.1109/RADAR.2009.4977077

Kucharczyk M, Hay GJ, Ghaffarian S, Hugenholtz CH (2020) geographic object-based image analysis: A primer and future directions. *Remote Sensing* 12: 2012. https://doi.org/10.3390/rs12122012

Kumar DN, Reshmidevi TV (2013) Remote sensing applications in water resources. *Journal of the Indian Institute of Science* 93: 163–188. http://civil.iisc.ernet.in/~nagesh/pubs/65_JIISc_RS_review_Jun13.pdf (date: 26.04.2021).

Lee J-S, Wen J-H, Ainsworth TL, Chen K-S, Chen AJ (2009) Improved sigma filter for speckle filtering of SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing* 47: 202–213. https://doi.org/10.1109/TGRS.2008.2002881

Li C, Shao Z, Zhang L, Huang X, Zhang M (2021) A comparative analysis of index-based methods for impervious surface mapping using multiseasonal Sentinel-2 satellite data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 3682–3694. https://doi.org/10.1109/JSTARS.2021

Ma L, Li M, Ma X, Cheng L, Du P, Liu Y (2017) A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 130: 277–293. https://doi.org/10.1016/j.isprsjprs.2017.06.001

Main-Knorn M, Pflug B, Louis J, Debaecker V, Müller-Wilm U, Gascon F (2017) Sen2Cor for Sentinel-2. *Proceedings of SPIE volume* 10427: 1042704. https://doi.org/10.1117/12.2278218

Mariye M, Jianhua L, Maryo M (2022) Land use and land cover change, and analysis of its drivers in Ojoje watershed, Southern Ethiopia. *Heliyon* 8: e09267. https://doi.org/10.1016/j.heliyon.2022.e09267

Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing 39:2784-281. https://doi.org/10.1080/01431161.2018.1433343

Mullissa A, Vollrath A, Odongo-Braun C, Slagter B, Balling J, Gou Y, Gorelick N, Reiche J (2021) Sentinel-1 SAR backscatter analysis ready data preparation in Google Earth Engine. *Remote Sensing* 13: 1954. https://doi.org/10.3390/rs13101954

Muringato Water Resource User Association (2014) Muringato WRUA sub-catchment management plan (SCMP). Nyeri.

Nabil M, Zhang M, Bofana J, Wu B, Stein A, Dong T, Zeng H, Shang J (2020) Assessing factors impacting the spatial discrepancy of remote sensing based cropland products: A case study in Africa. *International Journal of Applied Earth Observation and Geoinformation* 85: 102010. https://doi.org/10.1016/j.jag.2019.102010

Nasirzadehdizaji R, Akyuz DE, Cakir Z (2019) Flood mapping and permanent water bodies change detection using Sentinel SAR data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-4-W18: 797–801. https://doi.org/10.5194/isprs-archives-XLII-4-W18-797-2019

Nikaein T, Iannini L, Molijn RA, Lopez-Dekker P (2021) On the value of Sentinel-1 InSAR coherence time-series for vegetation classification. *Remote Sensing* 13: 3300. https://doi.org/10.3390/rs13163300

Ottosen T-B, Petch G, Hanson M, Skjøth CA (2020) Tree cover mapping based on Sentinel-2 images demonstrate high thematic accuracy in Europe. *International Journal of Applied Earth Observation and Geoinformation* 84: 101947. https://doi.org/10.1016/j.jag.2019.101947

Pereira-Pires JE, Mora A, Aubard V, Silva JMN, Fonseca JM (2021) Assessment of Sentinel-2 spectral features to estimate forest height with the new GEDI data. Camarinha-Matos LM, Ferreira P, Brito G (eds) *Technological innovation for applied AI systems:* 123–131. Cham. https://doi.org/10.1007/978-3-030-78288-7_12

Phiri D, Simwanda M, Salekin S, Nyirenda V, Murayama Y, Ranagalage M (2020) Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing* 12: 2291. https://doi.org/10.3390/rs12142291

Potapov P, Li X, Hernandez-Serna A, Tyukavina A, Hansen MC, Kommareddy A, Pickens A, Turubanova S, Tang H, Silva CE, Armston J, Dubayah R, Blair JB, 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

Qi W, Lee S-K, Hancock S, Luthcke S, Tang H, Armston J, Dubayah R (2019) Improved forest height estimation by fusion of simulated GEDI Lidar data and TanDEM-X InSAR data. *Remote Sensing of Environment* 221: 621–634. https://doi.org/10.1016/j.rse.2018.11.035

Qiu Y, Ming D, Zhang X (2016) Object oriented land cover classification combining scale parameter preestimation and mean-shift segmentation. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS):* 6332–6335. https://doi.org/10.1109/IGARSS.2016.7730655

Schulz D, Yin H, Tischbein B, Verleysdonk S, Adamou R, Kumar N (2021) Land use mapping using Sentinel-1 and Sentinel-2 time series in a heterogeneous landscape in Niger, Sahel. *ISPRS Journal of Photogrammetry and Remote Sensing* 178: 97–111. https://doi.org/10.1016/j.isprsjprs.2021.06.005

Sekertekin A, Cicekli SY, Arslan N (2018) Index-based identification of surface water resources using Sentinel-2

satellite imagery. *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*: 1–5. https://doi.org/10.1109/ISMSIT.2018.8567062

Sheffield J, Wood EF, Pan M, Beck H, Coccia G, Serrat-Capdevila A, Verbist K (2018) Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions. *Water Resources Research* 54: 9724–9758. https://doi.org/10.1029/2017WR022437

Shrestha B, Stephen H, Ahmad S (2021) Impervious surfaces mapping at city scale by fusion of radar and optical data through a random forest classifier. *Remote Sensing* 13: 3040. https://doi.org/10.3390/rs13153040

Thonfeld F, Steinbach S, Muro J, Kirimi F (2020) Long-term land use/land cover change assessment of the Kilombero catchment in Tanzania using random forest classification and robust change vector analysis. *Remote Sensing* 12: 1057. https://doi.org/10.3390/rs12071057

Torres R, Snoeij P, Geudtner D, Bibby D, Davidson M, Attema E, Potin P, Rommen B, Floury N, Brown M, Traver IN, Deghaye P, Duesmann B, Rosich B, Miranda N, Bruno C, L'Abbate M, Croci R, Pietropaolo A, Huchler M, Rostan F (2012) GMES Sentinel-1 mission. *Remote Sensing of Environment* 120: 9–24. https://doi.org/10.1016/j.rse.2011.05.028

Verhelst K, Gou Y, Herold M, Reiche J (2021) Improving forest baseline maps in tropical wetlands using GEDI-based forest height information and Sentinel-1. *Forests* 12: 1374. https://doi.org/10.3390/f12101374

Wang L, Koike T, Yang K, Jackson TJ, Bindlish R, Yang D (2009a) Development of a distributed biosphere hydrological model and its evaluation with the Southern Great Plains Experiments (SGP97 and SGP99). *Journal of Geophysical Research* 114: D08107. https://doi.org/10.1029/2008JD010800

Wang L, Koike T, Yang K, Yeh PJ-F (2009b) Assessment of a distributed biosphere hydrological model against streamflow and MODIS land surface temperature in the upper Tone River Basin. *Journal of Hydrology* 377: 21–34. https://doi.org/10.1016/j.jhydrol.2009.08.005

Witharana C, Civco DL (2014) Optimizing multi-resolution segmentation scale using empirical methods: Exploring the sensitivity of the supervised discrepancy measure Euclidean distance 2 (ED2). *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 108–121. https://doi.org/10.1016/j.isprsjprs.2013.11.006

Xie Z, Phinn SR, Game ET, Pannell DJ, Hobbs RJ, Briggs PR, McDonald-Madden E (2019) Using Landsat observations (1988–2017) and Google Earth Engine to detect vegetation cover changes in rangelands - A first step towards identifying degraded lands for conservation. *Remote Sensing of Environment* 232: 111317. https://doi.org/10.1016/j.rse.2019.111317

Xu H (2006) Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing* 27: 3025–3033. https://doi.org/10.1080/01431160600589179

Xue J, Su B (2017) Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors* 2017: 1353691. https://doi.org/10.1155/2017/1353691

Yang L, Mansaray L, Huang J, Wang L (2019) Optimal segmentation scale parameter, feature subset and classification algorithm for geographic object-based crop recognition using multisource satellite imagery. *Remote Sensing* 11: 514. https://doi.org/10.3390/rs11050514

Ygorra B, Frappart F, Wigneron JP, Moisy C, Catry T, Baup F, Hamunyela E, Riazanoff S (2021) Monitoring loss of tropical forest cover from Sentinel-1 time-series: A CuSum-based approach. *International journal of applied earth observation and geoinformation* 103: 102532. https://doi.org/10.1016/j.jag.2021.102532

Zanaga D, van de Kerchove R, Keersmaecker W de, Souverijns N, Brockmann C, Quast R, Wevers J, Grosu A, Paccini A, Vergnaud S, Cartus O, Santoro M, Fritz S, Georgieva I, Lesiv M, Carter S, Herold M, Li L, Tsendbazar N-E, Ramoino F, Arino O (2021) ESA WorldCover 10 m 2020 v100. https://doi.org/10.5281/zenodo.5571936

Zheng W, Shao J (2018) Water body change detection based on Sentinel-1 and HJ-1A/B satellites data. Proceedings of the International Workshop on Environment and Geoscience – *IWEG*: 484–488. https://doi.org/10.5220/0007432204840488

Authors

Marcus Goebel
marcus.goebel@rub.de
Jun.-Prof. Dr. Andreas Rienow
ORCID: 0000-0003-3893-3298
andreas.rienow@rub.de
Ruhr University Bochum
Institute of Geography
Universitätsstraße 150
44780 Bochum
Germany

Dr. Kuria Thiong'o
kuria.thiongo@dkut.ac.ke
Dedan Kimathi University of Technology
Institute of Geomatics, GIS & Remote Sensing
Private Bag Dedan Kimathi, Nyeri
Kenya