Mapping Rangeland Health Indicators in East Africa from 2000 to 2022

Gerardo E. Soto[1,2], Steven Wilcox[3], Patrick E. Clark[4], Francesco P. Fava[5], Nathan D. Jensen[6], Njoki Kahiu[6,7], Chuan Liao[8], Benjamin Porter[9], Ying Sun[2] & Christopher B. Barrett[10,11].

Affiliations:

[1] Instituto de Estadística, Facultad de Ciencias Económicas y Administrativas, Universidad Austral de Chile, Valdivia, Chile.

[2] School of Integrative Plant Science, Soil and Crop Sciences Section, Cornell University, Ithaca, NY, USA.

[3] Department of Applied Economics, Utah State University, Logan, UT, USA.

[4] USDA-Agricultural Research Service, Boise, ID, USA.

[5] Department of Environmental Science and Policy, Università Degli Studi Di Milano, Milano, Italy.

[6] The Global Academy of Agriculture and Food Systems, University of Edinburgh, Scotland.

[7] Department of Plant and Environmental Sciences, New Mexico State University, Las Cruces, NM, USA.

[8] Department of Global Development, Cornell University, Ithaca, NY, USA.

[9] Forest Ecosystem Monitoring Cooperative, University of Vermont, Burlington, VT, USA.

[10] Charles H. Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY, USA.

[11] Jeb E. Brooks School of Public Policy, Cornell University, Ithaca, NY, USA

Correspondence to: Gerardo E. Soto (gerardo.soto@uach.cl)

Short summary. Using machine learning and linear unmixing, this paper produced rangeland health indicators: Landsat time series of land cover classes and vegetation fractional cover of photosynthetic vegetation, non-photosynthetic vegetation, and bare ground in arid and semi-arid Kenya, Ethiopia, and Somalia. This represents the first multi-decadal high-resolution dataset specifically designed for mapping and monitoring rangeland health in the arid and semi-arid rangelands of this portion of Eastern Africa.

Abstract. Tracking environmental change is important to ensure efficient and sustainable natural resources management. Eastern Africa is dominated by arid and semi-arid rangeland systems, where extensive grazing of livestock represents the primary livelihood for most people. Despite several mapping efforts, Eastern Africa lacks accurate and reliable high-resolution maps of rangeland health necessary for many management, policy, and research purposes. Earth Observation data offer the opportunity to assess spatiotemporal dynamics in rangeland health conditions at much higher spatial and temporal coverage than conventional approaches that rely on in-situ methods, while complementing their accuracy. Using machine learning classification and linear unmixing, we produced rangeland health indicators: Landsat-based time series from 2000 to 2022 at 30 m spatial resolution for mapping land cover classes (LCC) and vegetation fractional cover (VFC, including photosynthetic vegetation PV, non-photosynthetic vegetation NPV, and bare ground BG), two important data assets for deriving metrics of rangeland health in Eastern Africa. Due to scarcity of in-situ measurements in the large, remote, and highly heterogeneous landscape, an algorithm was developed to combine very high-resolution WorldView-2 and 3 satellite imagery at < 2 m resolutions with a limited set of ground observations to generate reference labels across the study region using visual photo-interpretation. The LCC algorithm yielded an overall accuracy of 0.856 when comparing predictions to our validation dataset comprised of a mixture of in-situ observations and visual photo-interpretation from very high-resolution imagery, with Kappa of 0.832; the VFC returned a $R^2 = 0.795$, $p < 2.2e\text{-}16$, and normalized root mean squared error (nRMSE) = 0.123 when comparing predicted bare-ground fractions to visual photo-interpreted very high-resolution imagery. Our products represent the first multi-decadal high-resolution dataset specifically

designed for mapping and monitoring rangelands health in Eastern Africa including Kenya, Ethiopia and Somalia, covering a total area of 745,840 km^2. These data can be valuable to a wide range of development, humanitarian, and ecological conservation efforts and are available at https://doi.org/10.5281/zenodo.7106166 (Soto et al., 2023) and Google Earth Engine (GEE; details in data availability section).

1. Introduction

Rangelands cover nearly half of the African continent land mass and support the livelihoods of tens of millions of households (Reid et al., 2008, Sayre et al., 2013). The productivity of these rangelands along with the human and livestock populations they sustain is significantly affected by land degradation due to soil erosion, cropland expansion, shrub encroachment resulting from heavy grazing and suppression of fires, as well as climate change (Barbier and Hochard 2018, Roques et al., 2001, Angassa and Oba, 2008, Wynants et al., 2019, Vetter 2005, Hoffman and Vogel 2008). Episodes of extreme climate events, in particular, drought, have led to emergency population migrations and humanitarian crises of historic proportions (Blackwell 2010). Improved understanding of the variation in rangeland health across space and over time is crucial for community development, ecological conservation, and humanitarian programming in the region.

The extensive development of Earth Observation (EO) platforms has largely improved our understanding of ecosystems (Giuliani et al., 2020, Sudmanns et al., 2020). Long-term EO systems, such as the Landsat constellation, have provided valuable data to assess and accurately detect multiple ecosystem functions and patterns (Wulder et al., 2012, Loveland and Dwyer 2012, Williams et al., 2006). Further development of EO and analytics has allowed the integration of multiple platforms into complex algorithms and workflows, benefiting from the ability of image data to scale at different spatial and temporal levels (e.g., AghaKouchak et al., 2015) and leading to paradigm shift from change detection to continuous monitoring at high resolution (Woodcock et al., 2020). These recent developments have led to much interest in applying EO and related analytics to rangeland ecology and management (e.g., Allred et al., 2021, Hill et al., 2020, Rigge et al., 2020, Fava and Vrieling 2021).

Rangeland health has been conceptualized as framework of three fundamental attributes reflecting soil/site stability, and hydrologic function (Pellant et al. 2020). Historically, associated assessments have largely relied on in-situ methods for assessment. Recent scientific advances create an opportunity to map rangelands health using satellite imagery to monitor changes in rangeland health at ecologically meaningful scales for landscape planning and management (Allred et al., 2022). EO in these often-remote, arid and semi-arid regions becomes extremely valuable for its capacity to enable measurements in areas where data have never or rarely been collected on the ground. In addition, high-resolution (HR) remote sensing datasets can capture the fine spatial heterogeneity and the temporal dynamics that are key to informing management decisions but are also exceedingly difficult to discern at scale using conventional, ground-based monitoring systems (Zhou et al., 2020).

EO-based data have been used to inform on rangeland health since the early days of EO programs (e.g., Landsat 1 program: Haas et al., 1975, Gaetz et al., 1976). Understanding of rangeland ecosystems relies on information about the specific composition of the various vegetation communities within these ecosystems, oftentimes over large spatial extents, such as the Great Plains in North America (Reeves and Baggett 2014). Composition changes over time are important to track trajectories such as bush encroachment and soil degradation, impacts on grazers, etc. (Ghafari et al., 2018, Liao et al., 2018). HR thematic mapping of rangeland ecosystem can help explain key interannual variability in ecological processes such as water changes (Cooley et al., 2017), terrestrial and aquatic vegetation phenology (Cheng et al., 2020, Coffer et al., 2020), and crop dynamics (Lin et al., 2021), as well as long-term effects, such as land use change, aboveground carbon, and sedimentation (Sankey et al., 2019, 2021).

The lower computational barriers from the continuous advancement of technology are promoting the shift from plot-based assessments to the integration of satellite-based maps into landscape management, improving broad-scale mapping of rangelands at higher spatial and temporal resolutions than ever before (Jones et al., 2020, Allred et al., 2022). Many recent contributions to this field have shown that even though moderate resolution datasets (from MODIS sensors at 250 m resolution) are able to detect short-term vegetation phenology and long-term demographic dynamics of herbaceous and woody species, they cannot detect changes at local scales, because the spatial patterns of herbaceous and woody species typically occur at such fine scales (Angassa, 2014, Browning et al., 2017, 2019, Matongera et al., 2021, Oba et al., 2003). Despite collecting data at lower temporal resolutions, the Landsat collection at 30 m spatial resolution has consistently played an important role in science for over fifty years due to continuous efforts in calibration and corrections (Wulder et al., 2012, 2022, Franks et al., 2016). The recent collection-based reprocessing that resulted in the Landsat collection 2 (Wulder et al., 2022) represents an important opportunity to build consistent time series for HR rangeland mapping. In addition, field studies have demonstrated that Landsat-scale sub-pixel estimation of fractional cover of rangeland functional types, such as herbaceous and shrub components, and especially bare ground, is crucial to overcome the difficulties of parsing out the underlying heterogeneity within thematic land cover classifications and in understanding ecological dynamics

(Jones et al., 2018, Rigge et al., 2019). As a result, land cover classification (LCC) and vegetation fractional cover (VFC, including photosynthetic vegetation PV, non-photosynthetic vegetation NPV, and bare ground BG) estimations have become the two building blocks of rangeland health assessment of today's EO-based rangeland management (Jones et al., 2020). However, HR land and fractional cover mapping (i.e., using Landsat) over large and remote regions is hampered by the difficulty of collecting ground truth data at fine resolution. This is especially true in East Africa, where limited infrastructure and physical insecurity make it very difficult to collect field data at scale.

In this study, we produced a unique and new dataset composed of high resolution (HR) LCC and VFC annual estimates of rangeland components for Eastern Africa based on the Landsat collection from 2000 to 2022. We used a LCC scheme to help identify rangeland vegetation transition pathways, and VFC to describe rangeland health condition trajectories within each class. To overcome the challenge of scarce ground data for training and validating our models over this vast and remote region, we used a large collection of very high-resolution satellite imagery (VHR), visual photo-interpretation and ad-hoc algorithms to generate a large sample of reference data to generate and validate our two products.

2. Data and Methods

The overall strategy of our methodological framework to generate the long-term time series of LCC and VFC for rangelands in Eastern Africa consists of three major steps: first, the development of a training/testing dataset from VHR imagery (section 2.3); second, the LCC classification (section 2.4); and third, the VFC classification (section 2.5). The detailed workflow is provided in Figure 1.

To integrate in-situ and VHR data to create reference data, specifically, we used ground reference data to inform a Visual Photo-Interpretation (VPI) protocol to create reference labels to train supervised classifications of VHR imagery. These VHR classifications were used to create a large amount of machine-generated reference data to train HR classifiers and to identify areas with large proportions of the focal rangeland components for VFC estimation. To generate the LCC reference data, we generated an algorithm that created reference points using a set of conditions with the proportions of reference compositional component (RCC). The RCCs within each of our LC class definitions includes vegetation functional groups and other important classes such as bare ground. The RCCs are then compared to the calculated proportion of pixels from the VHR classification within a moving window matching the 30 m spatial resolution of the HR data. We also generated VFC reference data by using image segmentation on the RCC classifications with the assistance of an application on GEE to identify homogeneous areas of rangeland components that could spatially allocate HR pixels to use them to calculate spectral endmembers and generate VFC estimations. Figure 1 shows our general workflow, including reference data partitions, remote sensing data and results, processing algorithms, and accuracy assessments.

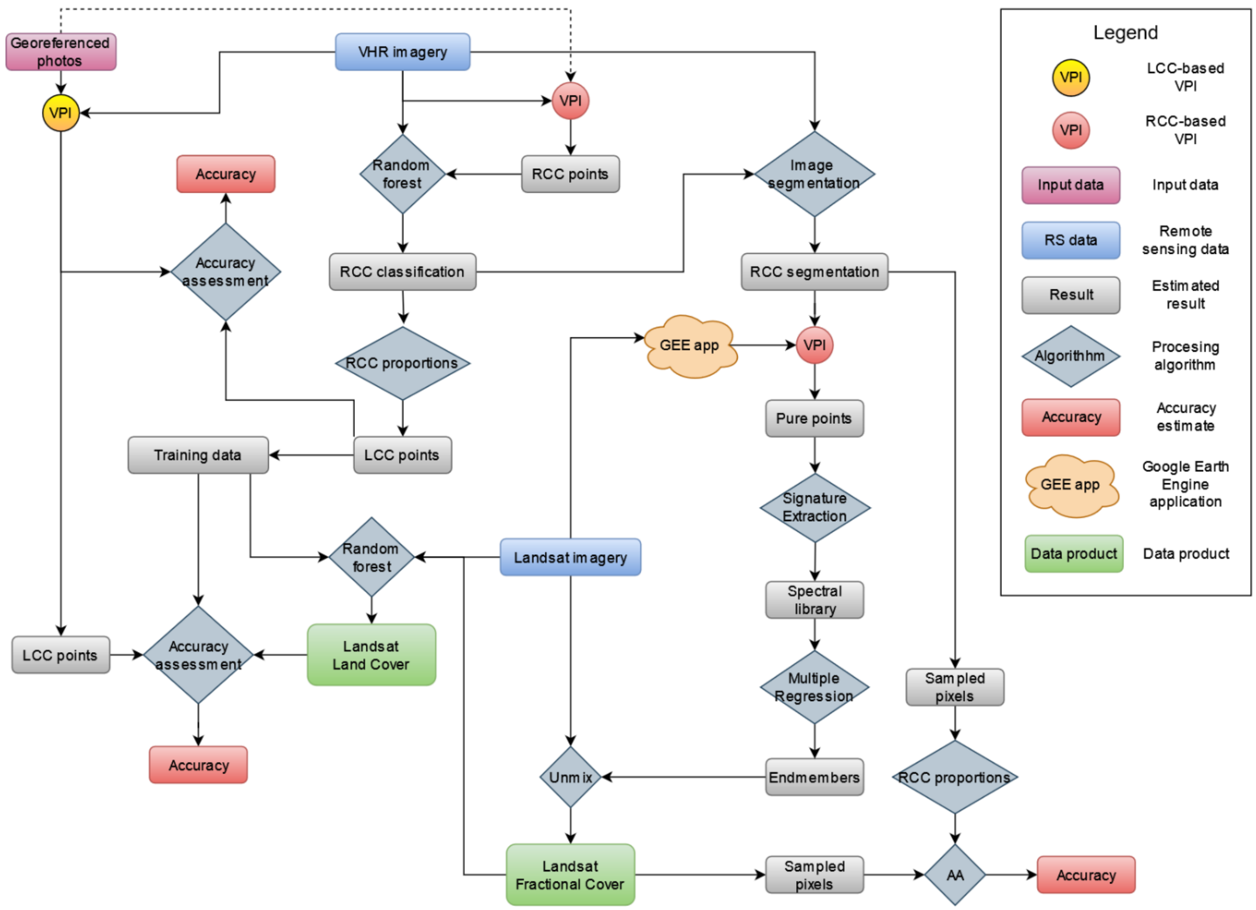

Figure 1: Schematic workflow of the process used in this work to generate Land Cover Classifications and Fractional Cover estimations based on Landsat imagery. The dashed line denotes the role of georeferenced pictures on informing the visual photo-interpretation (VPI) process. Short names correspond to: points generated with reference compositional component (RCC) work to train the RCC classification (RCC Points); collection of points with pure pixels (i.e., points with 100% of a single VFC type, Pure points); Library of spectral signatures of pure points (Spectral Library); Spectral endmembers for PV, NPV and BG (Endmembers); random sample of points with overlapping Landsat imagery to perform accuracy assessment on VFG estimations (Sampled pixels). See text for further details on the uses of data and processing.

Prior to the detailed technical description of this entire workflow (sections 2.4-2.6), we first described our study domain (section 2.1) and satellite datasets utilized in this study (section 2.2).

2.1 Study area

The study area is located in the semi-arid and arid regions centered on east and northern Kenya, western Somalia and southern Ethiopia (Figure 2). We chose this study region because it has been a geographic area with numerous development interventions on the ground in the past decades (Liao and Fei, 2017), but limited land cover datasets exist to evaluate the concurrent changes on the landscapes. In addition, this region suffers strongly from climate change extremes (e.g. droughts, floods, etc. IPCC, 2022) and their consequences on rangeland health, resilience, and well-being of pastoralists (Pricope et al., 2013, Beal et al., 2023). Covering a total of 745,840 km^2, it includes diverse types of rangelands, which represent hyper complex and rapid physiological and phenological dynamics in other regions of the world (ILRI, IUCN, FAO, UNEP and ILC, 2021, Adams et al., 2021, Nandintsetseg et al., 2024). Therefore, the study area demonstrates potential for broad generalization and sheds light for development efforts for stakeholders.

We used two main features to bound our study area. To the east and north, we used Landsat tiles, using PATH 164 and ROW 56 as limits, dropping tiles PATH 164, ROW 59 and 60 due to heavy cloud cover. To the west and south, we used a threshold value of mean annual precipitation of 700 mm using TerraClimate data (smoothed with a kernel convolution with Standard Deviation = 5 km; Abatzoglou et al., 2018), thus keeping the focus on the rangeland-dominated arid and semi-arid areas.

The study covers the epoch 2000 to 2022 to help capture decadal variation in ecosystem conditions and maximize Landsat data availability. Landsat imagery is limited in this area due high cloud cover often occurring during the two wet seasons observed in the region including the long rains (March to June) and short rains (October to December). In most cases, cloud free data was available during December through early March, which corresponds to the short dry (SD) season. Thus, we generated our datasets using imagery from a portion of the SD, from 15 December over 2000-2022 to 1 March over 2001-2023, which maximized the annual available data count per pixel and ensured even distribution of data over our period of study.

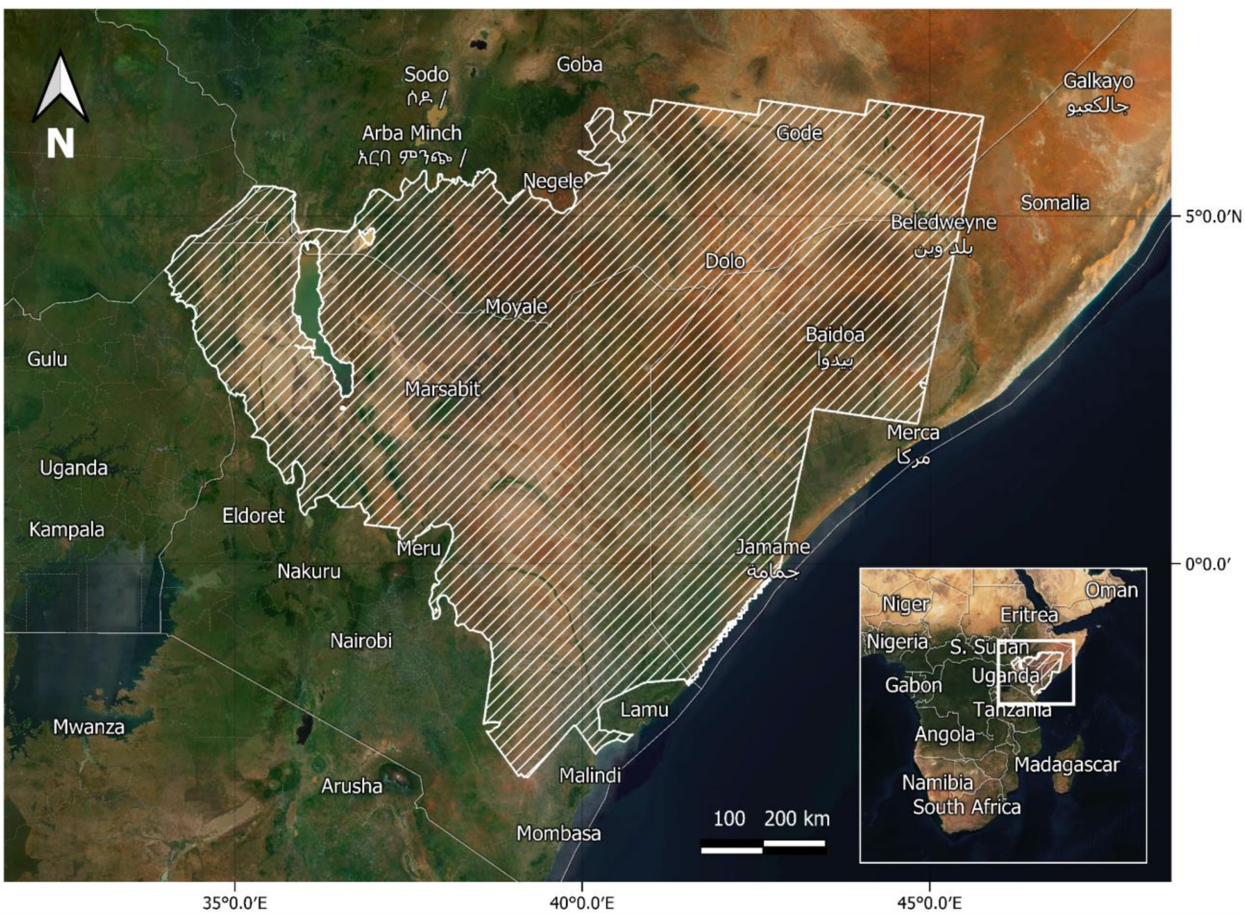

Figure 2: Map showing our study area in East Africa. Basemap: ©MapTiler, https://www.maptiler.com/copyright/.

2.2 Remote sensing data

2.2.1 VHR - Very high-resolution satellite imagery

To train our models and validate the results, we used VHR satellite imagery as little ground reference information exists in this vast and remote region. We obtained a large collection of imagery from Maxar Technologies via the United States National Geospatial-intelligence Agency (NGA): ordered with the following filtering parameters: sun elevation > 45°, off-nadir angle < 40°, and cloud cover < 50 %. The VHR collection was composed of 2,500 mosaicked strips of imagery scenes from Worldview-2 and -3 sensors (Figure 3). These mosaicked strips, typically 16.4 km in width, were delivered as orthorectified- and radiometrically-corrected bundles of eight bands including Coastal (400-450 nm), Blue (450-510 nm), Green (510-580 nm), Yellow (585-625 nm), Red (630-690 nm), Red Edge (705-745 nm), Near-InfraRed 1 (NIR1, 770-895 nm), and NIR2 (860-1040 nm) at a spatial resolution of 184 cm for WorldView-2 and 124 cm for WorldView-3, and a panchromatic band at a spatial resolution of 46 cm for WorldView-2 and 31 cm for WorldView-3. Shortwave Infrared (SWIR) imagery (1195 to 2365 nm) collected by Worldview-3 with a spatial resolution of ~3.7 m was also used in this study.

After subsetting to the short dry season, we manually selected 321 strips maximizing the spatial coverage and minimizing cloud cover, as most images with scattered clouds projected oblique shadows often resulting in < 10 % of pixels being usable for further analysis. These data corresponded to imagery acquired from 2016 to 2020. We considered using Quick Bird imagery from previous years, but data availability for our area of interest was minimal.

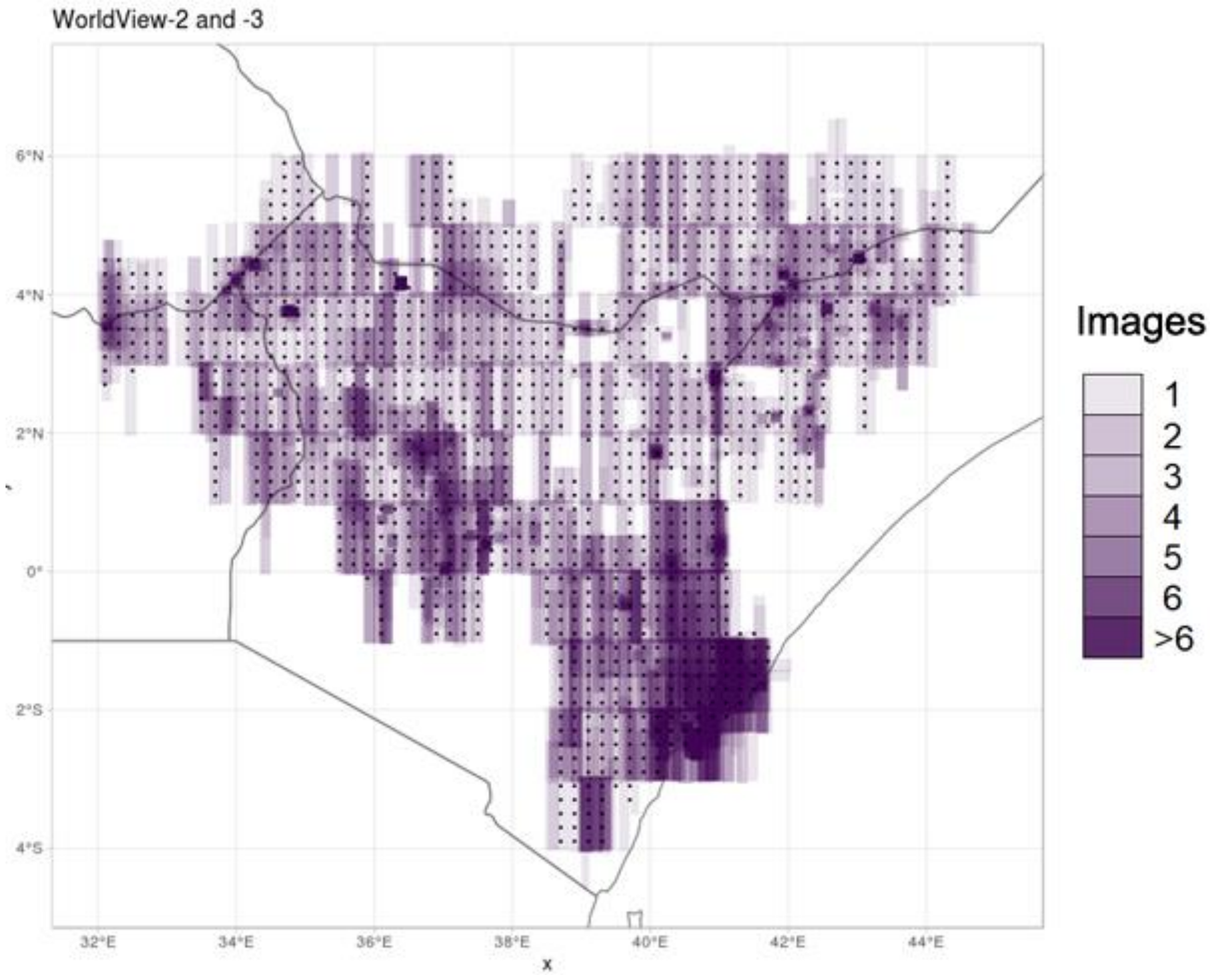

Figure 3: Spatial coverage of high-resolution imagery (polygons), and the spatial distribution of point grid (dots) used for generating reference data.

2.2.2 HR - High resolution Landsat collections

To capture historical changes in vegetation health in our area of study, we utilized Landsat dataset, which has been available for over four decades (1982-present; Wulder et al., 2012) and thus enables the development of long-term time series of land cover classes and vegetation fractional cover. While other studies have shown the value of higher resolution sensors such as Sentinel-2 to show the potential higher gain in accuracy compared to Landsat collection for the detection of invasive species in Eastern Africa (Duve et al., 2020), ESA's Sentinel mission only features a short history of imagery acquisition from 2015 (Drusch et al., 2013), which could bias our assessment towards the last decade, thus confusing the interpretation of our results.

Landsat data is readily and freely accessible for scientific purposes. It available at different processing levels, from raw images, to radiometrically-, geometrically-, and atmospherically-corrected scenes (Wulder, 2019). We used Google Earth Engine (GEE; Gorelick et al., 2017) to access and analyze atmospherically-corrected surface reflectance images for Landsat 5, 7 and 8 satellites from collection 2 (USGS, 2021), processed at the L1TP level (https://www.usgs.gov/core-science-systems/nli/landsat/landsat-levels-processing). Landsat data are packaged into overlapping "tiles", covering approximately 170 x 183 km each, using a standardized reference grid (USGS, 2019). In this study we used 42 of these tiles, totaling 1,192,654 km^2 (Figure 4). Differences in Landsat satellite sensors require different processing and correction techniques. We describe each sensor first and then outline our harmonization efforts.

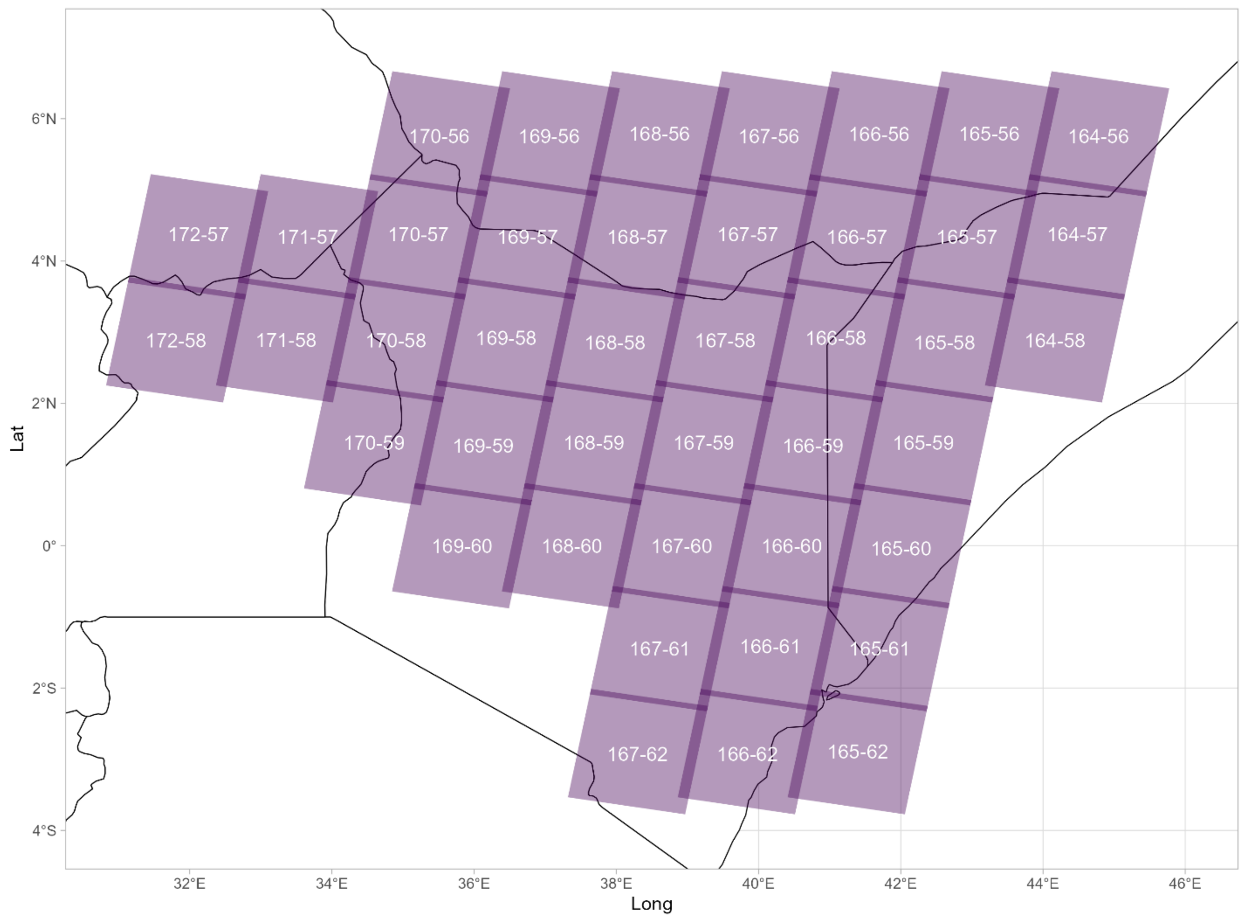

Figure 4: Spatial coverage of Landsat tiles used in this study spanning from 2000-2022. Numbers within each tile correspond to the PATH and ROW used in the Landsat data storage protocol.

Landsat 8 Operational Land Imager (OLI) uses data comprise of five visible and near-infrared bands: Coastal aerosol, Blue, Green, Red and Infrared (NIR), and two short-wave infrared (SWIR1 and 2). All bands were atmospherically corrected using the LaSRC (Land Surface Reflectance Code; USGS 2020). Other auxiliary data includes cloud, shadow, water, and snow mask layers generated with the C Function of Mask (CFMask) algorithm version 3.3.1 and stored in the Pixel Quality Assessment Band (QA_PIXEL; Foga et al., 2017, USGS 2022), as well as a saturation mask band in the Radiometric Saturation Quality Assessment Band (QA_RADSAT).

Landsat 5 (TM) and 7 Enhanced Thematic Mapper Plus (ETM+) data also contains different types of observation bands according to their position in the electromagnetic spectrum. Visible, near-infrared and SWIR bands: Blue, Green, Red, Infrared (NIR), and SWIR1 and SWIR2 bands processed to convert raw values to orthorectified surface reflectance values. All bands have a resolution of 30 m / pixel. All bands were atmospherically corrected using LEDAPS (Schmidt et al., 2013). Other auxiliary data includes cloud, shadow, water, and snow mask layers generated with the CFMask algorithm and stored in the QA_PIXEL band, as well as a saturation mask band in the QA_RADSAT band.

Landsat 7 has the potential to help fill the gaps between Landsat 5 and 8, being available from the year 1999 to date. However, the failure of the Scan Line Corrector (SLC) of Landsat 7 in 2003 somewhat limits its utility (Markham et al., 2004). This failure resulted in areas that are not imaged (~22 % of each tile), otherwise, data are valid for work and analysis. These data show similar distribution of cloud cover and revisiting times as Landsat 8 collection. Hereafter, we refer to data pixels as any pixel where no masking occurred, and valid and usable data was available.

2.2.3 Landsat collection harmonization

We used reduced major axis regression to harmonize the surface reflectance values from Landsat 5 and 7 to match the spectral information of Landsat 8 following Roy et al., (2016) on each Landsat data tile. These transformations are performed to improve temporal continuity between Landsat sensors (TM, ETM+ and OLI). After harmonization, the collections were merged and annual composites from December 15th to March 1st were generated using the median value of available data pixels. We used the median value, as the mean often gets biased with cloud contaminated pixels that were not included in the Level-1 QA_PIXEL Band used for cloud masking. In this study, the year of the annual composites correspond to the calendar year where the composite starts (i.e., December 15th). We selected this time interval as it was when imagery was mostly available, thus minimizing temporal imbalances among annual estimations. However, a large proportion of pixels were masked as a result of heavy cloud cover, with more than 5 % of masked pixels in 4 out of 21 different years. 2006 was a particularly problematic year, where the cloud component resulted in 10 % of pixels being masked (Figure 5). The launch of Landsat 8 in 2013 not only implied an improvement in the sensor characteristics, but also increased data collection capacity, thus reducing the likelihood of acquiring cloud-covered imagery as is evident in our study area (Figure 5).

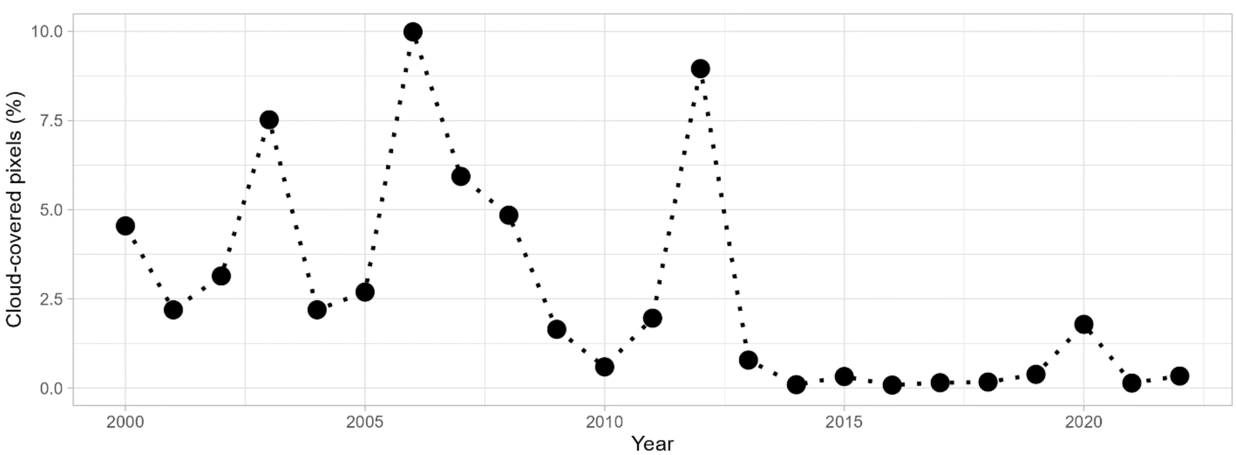

Figure 5: Cloud covered pixels present on the short dry (SD) season composites of Landsat imagery used in this study.

2.3 Development of training/testing datasets by integration of in-situ and VHR data

2.3.1 VPI - Reference dataset by visual photo-interpretation of VHR imagery

We applied this classification scheme using VPI methods to develop training data for the classification algorithms for both LCC and VFC. We started first at the Borana Zone in southern

Ethiopia, in the northern portion of our AOI where a rich source of georeferenced, ground-based photography (N = 1419 photos) was available for both a dry season 28 June – 26 August 2013 and wet season 6-31 May 2014 (Liao et al., 2018). In this VPI work, we leveraged this photography with VHR satellite imagery of the same locations and approximate time frames to capitalize on the differing contextual strengths of each data source. The photography provided a low-angle oblique view of vegetation functional groups and canopy layers for better class identification. The VHR imagery, viewed via Google Earth (GE) or via United States National Geospatial-intelligence Agency's (NGA) Global Enhanced GEOINT Delivery (G-EGD), provided a broader, nadir-oriented view of differing vegetation stands in context with one another, allowing more confident class separation.

Specifically, a team of four VPI analysts was trained to identify eight land cover classes following those employed by Liao and Clark (2018). We made additional refinements to these classes as detailed in Table 1. A detailed protocol was developed to ensure effective quality control. Training materials included reference flash card sets (see Appendix B) created for each of our land cover classes depicting a ground-based oblique view of a stand of representative vegetation in addition to a nadir VHR satellite view of that same stand in context with other surrounding vegetation in the locale. Canopy cover flash card sets were also created for 2-m, 4-m, and 8-m shrub and tree crown diameters to aid in visually estimating cover percentages relative to the thresholds separating each land cover class. The VPI classification was calibrated using the reference card sets and a standardized set of VPI points and associated photographs and imagery. Upon implementation, periodic spot checks of each analyst's VPI classifications were conducted to affirm consistency and accuracy.

Table 1: Land cover classes used for the Landsat land cover mapping (modified from Liao and Clark, 2018).

Land Cover Class	Code	Description
Closed Canopy Woodland	CCW	Areas vegetated by a stand of trees with an interlaced canopy. Shrubs are usually present and interspersed within the woodland. Tree canopy cover is > 50 %. The CCW class represents a state that is usually distributed in relatively humid areas at upland elevations, on ridge crowns, or within riparian corridors where favorable edaphic and/or climatic conditions facilitate relatively dense tree growth.
Dense Scrubland	DS	Areas vegetated by an abundance of shrubs with a low to moderately productive herbaceous component. Shrub canopy cover is > 50 %. Herbaceous cover generally decreases with increasing shrub cover due to competitive relationships. Trees, if present are sparsely to moderately conspicuous with canopy cover typically < 10 %. The DS class represents a state to which the Sparse Scrubland (SS) state can transition to via shrub recruitment. The DS state can itself, transition to the Closed Canopy Woodland (CCW) state via tree recruitment under favorable edaphic and/or climatic conditions.
Bushland	BU	Areas sparsely to abundantly vegetated almost exclusively by shrubs. This class is largely limited to arid lowland areas where climatic and edaphic conditions severely limit herbaceous presence. Woody plant cover ranges 10-100 %. Although bushland thickets can and do form, this class represents a state that is separated from the Sparse Scrubland (SS) and Dense Scrubland (DS) states by its severe, site-based limitation on herbaceous presence.
Open Canopy Woodland	OCW	Areas vegetated by an open stand of trees with a sparse to abundant herbaceous or herbaceous/shrub component. Trees are always conspicuous occurring as scattered individuals or clumps of a few individuals with a canopy cover of 10-50 %. A woodland aspect is always retained. If shrubs are present, their occurrence ranges from sparse to common but shrub canopy cover is < 50 %. The OCW class represents a state which can transition to Sparse Scrubland (SS) by tree loss and shrub encroachment or, under favorable climatic and/or edaphic conditions, to Closed Canopy Woodland (CCW) by tree recruitment.

Sparse Scrubland	SS	Areas vegetated by scattered shrubs with a sparse to very abundant, productive herbaceous component. Shrubs are always conspicuous. Shrub canopy cover ranges from 10-50 %. Herbaceous cover generally decreases with increasing shrub cover due to competitive relationships. Trees, if present, are sparsely to moderately conspicuous with canopy cover of < 10 %. This vegetation class represents a state that lies between the Grassland (GR) and Dense Scrubland (DS) states.
Cultivated Land	CL	Areas currently being used for crop cultivation or, in cases of field abandonment, cropping disturbance is still visually evident (bare soil, tillage boundaries, etc.). Seasonally fallow fields (bare) are included in this class as well as those with growing crops. Common crops include *Zea s*p., *Sorghum sp.* and *Erogrostis tef*. CL areas usually fenced and located in places of relatively deep and moist soils (e.g., near seasonal river or stream courses).
Grassland	GR	Areas where the vegetation cover is dominated by grasses and occasionally other herbs. Herbaceous canopy cover ranges from 10 to 100 %. Widely scattered trees and shrubs may be present but woody canopy cover is < 10 %. The vegetation state represented by this class can transition to the Open Canopy Wood (OCW) state by tree recruitment or to the Sparse Scrubland (SS) state by shrub recruitment.
Sparsely Vegetated Land	SV	Areas poorly covered by vascular herbaceous or woody plants. Plant cover is < 10 %. SV typically represents areas where vegetation presence is severely limited by soil chemical (e.g., hypersalinity) or physical conditions (very shallow depth). Rock outcrops are included in this class. SV can also occur in areas which have suffered topsoil loss due to heavy disturbance (e.g., recursive, heavy grazing and/or trampling) and subsequent wind and/or water erosion.

VPI classification took place as follows. The VPI point set from the georeferenced photograph locations were randomly subset into equal partitions, and each partition was assigned to a VPI analyst. The software package, Nikon View NXi ™ was used to view the photographs and mapped camera location and oblique view direction on a satellite imagery background provided by the software. The camera location coordinates were then plotted in GE and vegetation at the location was evaluated using VHR imagery that was concurrent or nearly concurrent with that of the photograph. Where concurrent imagery was missing from GE, imagery from the NGA archive was ordered and viewed via G-EGD and a VPI-based classification was made for the camera location. Where the camera location occurred in a mixed or ecotonal area, a new point in a nearby, more representative location (i.e., more homogenous vegetation structure, cover, and composition) was selected by the analyst and classified to a land cover class. Upon completion of the VPI classification, a random sample of 10 % of the 1,419 VPI points was spot checked to confirm overall consistency and accuracy across analysts. Where consistent bias or misclassification was found, additional training was provided, and the analyst(s) re-visited all assigned points for the troublesome class or classes and re-classified these points as necessary.

As the extent of this dataset was limited to the north area of our AOI, we extended the use of this dataset as reference to inform recognition of the vegetation functional group components of each land cover class used here. Vegetation functional groups generally refer to different types of vegetation that are functionally and structurally different. In our setting, the primary groups are trees, shrubs, and grasses. Using pan-sharpened VHR imagery, we then performed independent VPI classifications of VFGs within classes to develop and refine a supervised machine classifier and to support fractional cover analyses which are described in the next sections. This additional VPI work followed a procedure to spatially label the key components within each of the land cover classes and was focused on a grid of 8 x 8 km squares centered in a regular point pattern where VHR imagery was available (see Figure 3). These reference compositional components (RCC) included the vegetation functional groups (trees, shrubs, grass) as well as bare ground, water, cultivated land and impervious surfaces. We leveraged the combination of nadir views from VHR satellite imagery and the large set of available landscape photographs from the northern portions of our AOI to recognize visible characteristics of each sub-class component and apply these characteristics in VPI classification of the entire study area.

2.3.2 RCC - Reference compositional component classification of VHR imagery

To create the reference dataset for calibration and validation of LCC and VFC estimations for our entire study area, we relied on RCC data generated from the classification of VHR imagery. RCC represents the basis of LCC as our land cover scheme (see below) follows a compositional combination of them. In addition, RCCs are an important input for VFC estimation, which needs to be complemented with non-photosynthetic vegetation reference points, created with a different approach (see below).

We calculated the normalized difference vegetation index (NDVI) from the red and NIR-1 bands of the VHR imagery and then added the NDVI as a new band to the VHR dataset. Spectral signals were then extracted and assigned to the points generated in the VHR VPI work with each assigned RCC class and a random forest classification was performed to predict RCCs using the spectral information as covariates. The number of trees was set to 1000, with two variables tried at each split. After model fitting, we used a graph showing the out of bag error of each class versus the number of trees in the classification to explore the effects of sample sizes on the accuracy of the method and increase it when needed. Classification of VHR imagery focused on classifying RCCs; trees, shrubs, grasses, bare ground, water, cultivated land and impervious surfaces (e.g., Figure 6). After training our classification algorithms on 90 % of the generated labels, we then used the remaining 10 % to compare the (out of sample, OOS) prediction of the classifier against the actual reference labels using confusion matrices. We set a threshold minimum value of 85 % overall accuracy for using the resulting classifications in the following analysis steps. A random sample of VHR classified imagery with accuracies above the threshold was selected and visually inspected to understand misclassifications and their potential drivers. We increased our RCC-oriented VPI effort if threshold levels were not met until accuracy met our threshold value. Despite our efforts and due to cloud cover and other factors such as cropland misclassifications in humid areas, only 44.5 % (n = 143) of the total RCC classifications were retained using the 85 % accuracy threshold. Lower accuracy classifications occurred in areas of highlands on the west and southeast portions of our study area, characterized by higher precipitation. After contrasting classification predictions against pan-sharpened images, we recognized that most of the misclassifications corresponded to classes including green vegetation such as grass, crops and trees. Other sources of error included areas with cloud shadows and impervious surfaces.

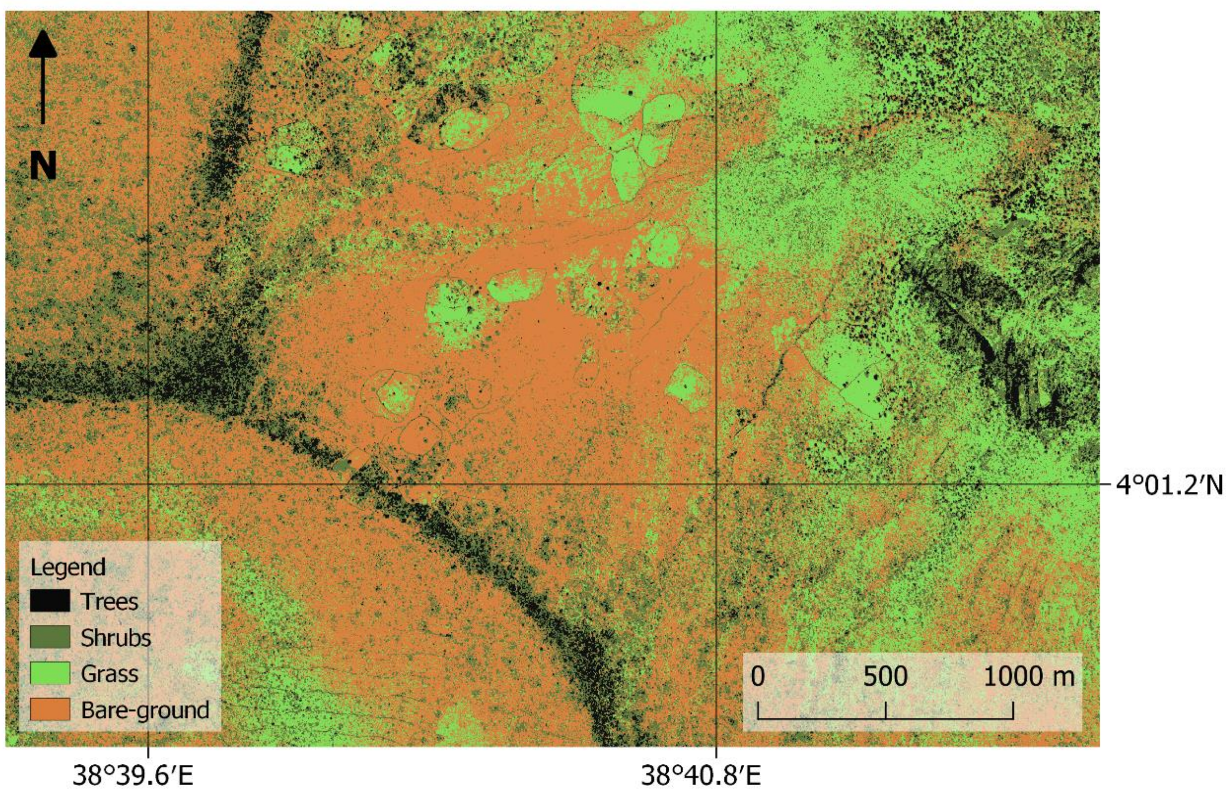

Figure 6: Example of a RCC classification result using a Worldview-3 image.

2.3.3 Composition-based algorithm for HR reference data creation

After classifying VHR strips and selecting those with higher accuracy, we applied a custom-made algorithm that uses a squared moving window of the size of a Landsat pixel (30 x 30 m) and calculates the proportion of VHR pixels, representing the area in the window covered by each of the RCC classes from the predicted VHR classification. Using the proportion of VHR pixels for each RCC class allowed us to use both Worldview datasets, as they have different spatial resolution. Then, using the list of defined threshold compositional percentages of RCC classes per land cover class in Table 1, we built code to meet the criteria for each land cover class. We then selected a stratified random sample of 80,000 points to be used as training points for the Landsat classification, described next. Points retained the date of the VHR strip used to generate them. Due to misclassifications associated with scattered cloud cover in some imagery, we further applied a buffer of 500 m around areas where more than 100 pixels of cloud or shadows were detected inside the moving window described above and excluded these from the RCC proportion calculation and class assignment.

2.4 Land cover classification

2.4.1 Land cover classification model

Our LCC scheme is based on the State Transition Model (STM; Bestelmeyer et al., 2017, Steele et al., 2012, Blanco et al., 2014) developed for this region by Liao and Clark (2018), with adjustments based on contributions from Pratt et al., (1966) and Liao et al., (2018) (Figure 7). Specific changes included the addition of classes not included in Liao and Clark (2018) and more precise definitions of the characteristics of each class and the trajectories between them, given the extension of our study area. The scheme includes eight land cover classes, each representing a vegetation state defined by structure, cover, and functional group composition. The potential transitions among these states or classes are described in the mapping legend provided in Table 1, which adapts Table 1 from Liao and Clark (2018). However, tree, shrub, and herbaceous cover thresholds have been further refined to better define class separations. The bushland class was also more clearly defined as a state where herbaceous presence was severely limited by climatic and/or edaphic factors rather than interspecific competition with shrubs and/or trees for resources. Transitional pathways associated with wild or prescribed fires have been excluded from Figure 7 and the legend (Table 1) to simplify description and presentation given the complexities associated with fire-tolerant versus fire-intolerant woody species, wildfire control, and past prohibitions on prescribed fire.

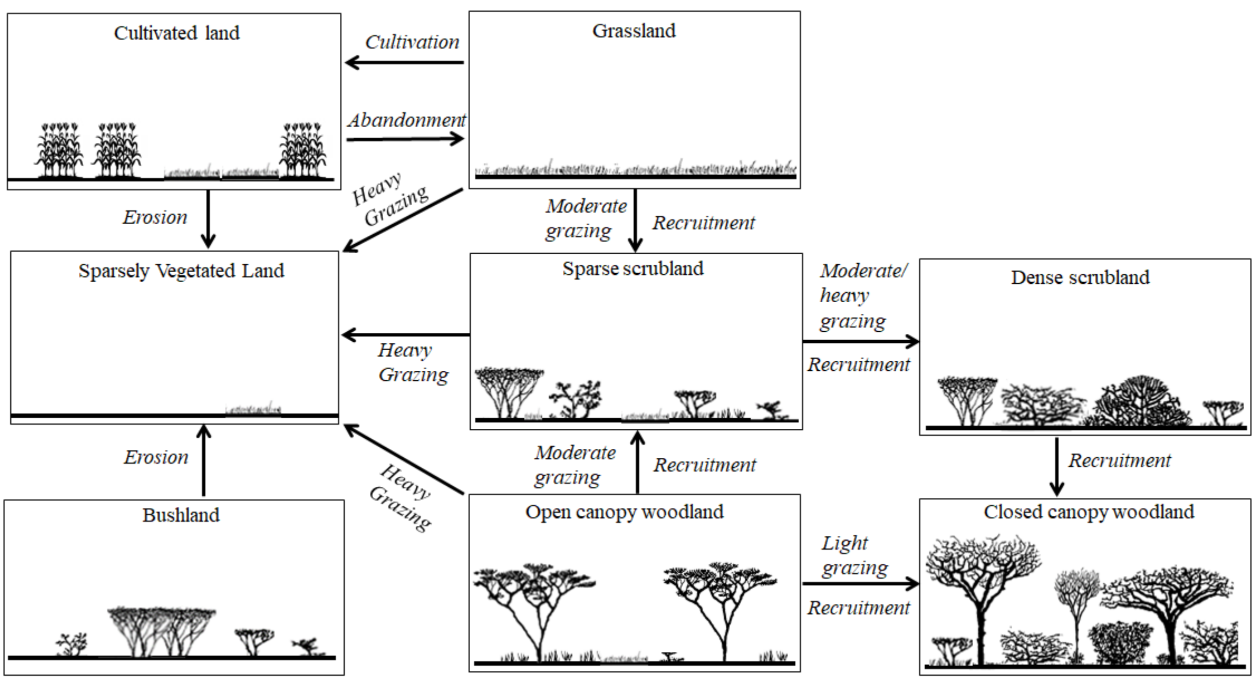

Figure 7: States and transition pathways among eight land cover classes.

2.4.2 Land cover classification algorithm

The land cover classification consists of two general steps. First, the VHR imagery was classified using the combination of RCC labels generated from VPI work (described in section 2.3.2) and random forest classifiers (Belgiu and Dragut 2016), producing RCC classifications. Second, an automatic algorithm, based on conditionals and the percentage thresholds of RCC defining each LC class (described in Table 1) was run over the RCC classifications to generate new training labels for the classification of the Landsat collections with a second random forest classifier. Here, we describe the HR Landsat classifications.

Landsat collections were classified using the random points generated from the RCC classifications (see section 2.3.3). We reserved 1,419 in-situ points from Liao and Clark (2018), so we could later use this dataset with VHR ground reference data to independently assess the accuracy of our results. We first masked all Landsat images using the SR_CLOUD_QA band generated from the CFMASK algorithm of Surface Reflectance Landsat data. To eliminate water bodies and rivers in our AOI, we applied a normalized difference water index (NDWI) mask, whereby pixels with values > 0.2 were removed (Gao 1996). We also calculated and added enhanced vegetation index (EVI), modified soil adjusted vegetation index 2 (MSAVI2), and Normalized Difference Water Index (NDWI) bands to the collections (Qi et al., 1994, Liu and Huete et al., 1995, McFeeters 1996). We also used CGIAR SRTM 90m Digital Elevation Database version 4 to include elevation and derived slope and horizontal curvature (Jarvis et al., 2008, Safanelli et al, 2020). Last, we included the bare ground and photosynthetic vegetation fractions from our fractional cover results (see Figure 1) as covariates, which were found to increase accuracy during our testing/tuning stage. We used the 80,000 algorithm-generated training points through the RCC classification protocol explained in section 2.3, and randomly partitioned them into 90 % training and 10 % for accuracy assessment. We then extracted the spectral information from the Landsat composite corresponding to the year of the date of each VHR image used to generate the training points through VPI work (see section 2.3.1). With these points, we trained a random forest algorithm to predict the vegetation classes of the entire collection. Thus, a single multi-year random forest classifier was used for prediction on the harmonized Landsat collection. After initial tuning of the classifier, we used 20 trees and a maximum number of 50 nodes. The resulting classified collection includes images with pixel values associated with our main land cover classes and masked pixels of cloud cover, shadows, and water.

2.4.3 Accuracy assessment of land cover classification

We used multiple reference year calibration to generate a classification model dependent on the surface reflectance data (Gomez et al., 2016). Based on the standard assumption that surface reflectance data represent the true ground response of features to sunlight, the classification model is then used to predict past and future time steps in the RS time series. Often, these data are referred to as absolute-normalized data (radiometrically and atmospherically corrected and orthorectified, Thenkabail et al., 2015). After generating reference labels through the combination of VHR imagery classification and an area-proportional classifier to upscale VFGs to land cover classes, we randomly partitioned this reference data set into training (90 %) and validation (10 %). We used the validation partition with the addition of the 1,419 points from Liao and Clark (2018) to create confusion matrices to assess the accuracy of the predictions. The percentage of classes in the random validation sample relative to the total amount of reference data was 9.3% for CCW, 14.5% for DS, 13.6% for BU, 7.7% for OCW, 9.6% for SS, 1.7% for CL, 8.8% for GR, and 16.7% for SV (see Table 1 for class names).

2.5 Fractional cover classification

We used bilinear unmixing to estimate fractional cover (Quintano et al., 2012) of three components of rangeland: bare ground (BG), photosynthetic vegetation (PV), and non-photosynthetic vegetation (NPV). We combined VHR and Landsat imagery to identify homogeneous areas where the spatial footprint of Landsat pixels could capture pure spectral signals for the three components of fractional cover. In this context, pure refers to pixels with 100 % cover of one of our three main components (Boardman et al. 1995). Given the heterogeneity of soil types in our study area, we allocated special effort on finding as many BG pixels as possible. To find these, we used the resulting RCC classification of VHR imagery (see 2.2.2) and performed image segmentation to identify homogeneous areas covered by bare ground. Because Landsat pixels are 30 by 30 m and their footprints could change with each revisit, we built an algorithm to scan the classifications to find homogeneous areas larger than 50 by 50 m, in order to allocate Landsat pixels with a margin of 10 m in both spatial axes.

We used GEE to manually create a sample of pure pixels, by mapping different Landsat color composites and creating graphs of 10-year-long NDVI and MSAVI2 time series and spectral profiles (i.e., spectral signatures) including all bands from the Landsat imagery for visually selected locations in the map. Using these visualizations, we checked that Landsat pixels corresponding to BG always covered the extent of the focal area and were not contaminated by vegetation or other features such as litter or impervious surfaces. To identify PV, we checked NDVI and MSAVI2 time series and natural color composites and selected a given acquisition time for a Landsat image containing green vegetation. Finally, to identify NPV, we used the reflectance profiles, NDVI and MSAVI2 time series and natural color composites to identify senescent vegetation and pixels where and when crops were harvested and dead vegetation was left behind.

After a sample of 108 locations for bare ground, 900 locations for NPV, and 900 locations for PV were established, the spectral information of the temporally closest Landsat image was extracted for its use in the endmember estimation. We estimated the endmembers from the spectral signatures of the sampled pure points using an R-based function for modeling of endmember compositions based on bilinear unmixing (Seidel and Hlawitschka 2015, Weltje 1997). We used "Blue", "Green", "Red", "NIR", "SWIR1", "SWIR2" bands as input spectral data for each point and established a convexity threshold of -6 and 10000 iterations with a standard weighting exponent of 1, as suggested by Weltje (1997).

We used a pseudo-inverse unmixing algorithm on GEE with two constraints to calculate fractional covers. The first constraint forces the fractions to sum to one, so that each fraction represents an actual percentage of each class. The second constraint forces all fractional values to be non-negative. The resulting maps include three bands corresponding to each of the three calculated fractions.

2.5.1 Accuracy assessment of fractional cover

We used RCC classifications to assess the performance of our fractional cover estimations, as the RCC classifications provide very accurate measures of class fractions at the Landsat pixel scale. Using the results from the classifications performed over VHR imagery, we aggregated the classified classes into vegetation, BG, and other (including impervious surfaces, water, and cloud classes). Since NPV is difficult to detect with available VHR datasets, this aggregation permits a separation between vegetation classes (which logically include PV and NPV) and BG, since BG is the complementary proportion of vegetation when just the two classes occur (i.e., where there is no cloud obstruction, water or impervious surfaces, or: 1 - BG = PV + NPV). Second, we selected the temporally closest Landsat-based fractional cover layer to a subset of 10 RCC classifications.

Third, we generated a layer of the centroids of pixels for these fractional cover estimates and randomly selected 5000 centroids. Fourth, we generated circles of 15m radius (approximate size of Landsat pixels) at the locations of the sampled centroids and clipped the aggregated classification. From this sample, we only selected the circles fully overlapping vegetation and BG pixels. Fifth, we calculated the proportion of pixels of vegetation and BG within each circle. Finally, after completion of this process, we compared the values of these proportions to the Landsat-derived fractional cover by using regression statistics: R^2, normalized root-mean-squared error (nRMSE) in units of percent cover, and *p*-values.

3. Analysis

3.1 Land cover classification

Overall, the LCC procedure resulted in an overall accuracy of 85.57 %, with Kappa of 0.832, which is above the recommended threshold of 85 % for LCC predictions and remarkable for such a large area as our study area (Foody 2002, see Figure 8). The resulting confusion matrix from the accuracy testing partition of the 8,191 randomly selected points is presented in Table 2. The random forest model using all bands was more accurate than those using subsets of input bands. In decreasing order, variable importance derived from the random forest classifier for every band was elevation, Green, EVI, Red, SWIR2, Blue, Slope, Photosynthetic vegetation, SWIR1, MSAVI2, horizontal curvature, NIR, and bare ground (Figure 9). The proportion of importance on the elevation covariate is almost double the next most important variable, the green band. Figure 10 shows the proportion of reference data, including training and validation sets, showing the apparent elevation segregation of the samples.

Table 2: Confusion matrix of the random forest classifier using multi-year validation samples. Class codes are presented in Table 1.

Class		Reference Data								Sum	User's Accuracy (%)
		CCW	DS	BU	OCW	SS	CL	GR	SV		
Predicted LCC	CCW	918	6	0	13	4	3	1	0	945	97.1
	DS	13	969	15	109	29	6	2	0	1143	84.8
	BU	0	26	1064	138	7	5	3	0	1243	85.6
	OCW	81	18	138	1080	7	8	4	0	1336	80.8
	SS	0	9	110	98	914	14	82	5	1232	74.2
	CL	2	3	4	0	4	23	31	0	67	34.3
	GR	4	3	1	2	109	26	823	9	977	84.2
	SV	0	0	9	7	3	8	3	1218	1248	97.6
Sum		1018	1034	1341	1447	1077	93	949	1232	8191	
Producer's Accuracy (%)		90.2	93.7	79.3	74.6	84.9	24.7	86.7	98.9		

CCW: Closed canopy woodland, DS: Dense scrubland, BU: Bushland, OCW: Open canopy woodland, SS: Sparse scrubland, CL: Cultivated land, GR: Grassland, SV: Sparsely vegetated land.

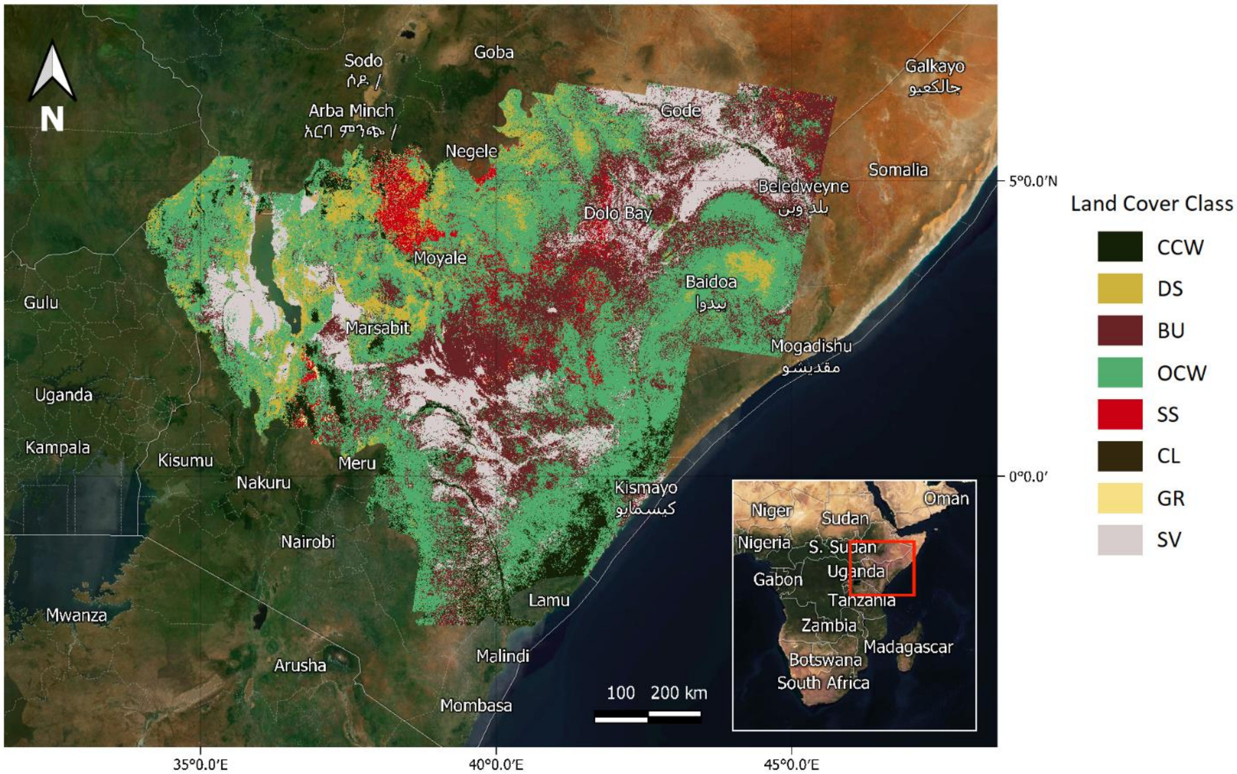

Figure 8: 30 m resolution predicted land cover classification for 2015. Class codes and descriptions are presented in Table 1. Basemap: ©MapTiler, https://www.maptiler.com/copyright/. Class codes correspond to: CCW: Closed canopy woodland, DS: Dense scrubland, BU: Bushland, OCW: Open canopy woodland, SS: Sparse scrubland, CL: Cultivated land, GR: Grassland, SV: Sparsely vegetated land.

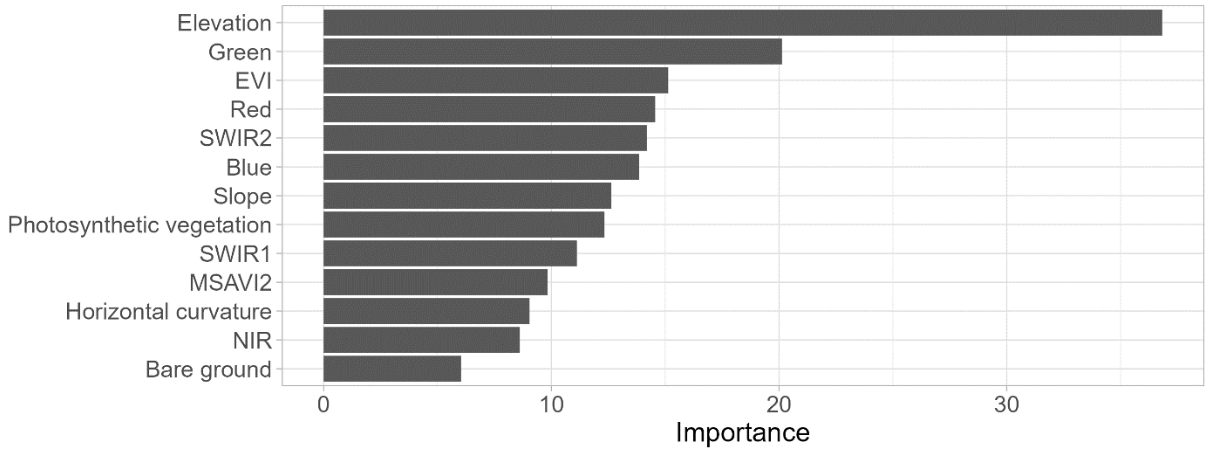

Figure 9: Variable importance derived from the best random forest classifier (see description of variables in section 2.4.2).

Figure 10: Proportion of reference data (81,419 total pixels) for each land cover class and 400 m elevation interval in our study area. Class codes correspond to: CCW: Closed canopy woodland, DS: Dense scrubland, BU: Bushland, OCW: Open canopy woodland, SS: Sparse scrubland, CL: Cultivated land, GR: Grassland, SV: Sparsely vegetated land.

The annual time-series of the total proportion of each land cover class in our study area, shows variations in the proportion of SV, SS, and OCW classes around the same years within the studied time frame (Figure 11). To understand the source of such variation, Figure 12 presents the proportion of inter-annual transitions of each pixel from class to class for the study period. Potentially valid transitions are defined in our state transition model, presented in Figure 7. Using this model, we can use the potentially valid inter-annual transitions and compare them with all inter-annual transitions in each pair of subsequent years (only using unmasked pixels with class values in both years). Our expected, potentially valid, inter-annual state transitions between land cover classes (Figure 7) were above 62.30 % in all yearly transitions (Figure 12) with a mean of 75.20 % and a maximum of 83.20 %. The number of unmasked paired pixels as a proportion of the total Landsat-based pixels used for the calculation of land cover had a minimum of 86.60 %, with a mean of 95.30 %. Three drops in the number of valid transitions are visible in Figure 12, which correspond to three drought events followed by rains and a greening effect on the landscape (Okal et al., 2020). This effect becomes evident while looking at the changing proportions of Closed Canopy Woodland (CCW) and Sparse Vegetation (SV) for 2005-2006, 2010-2011 and 2017-2019 in Figure 11.

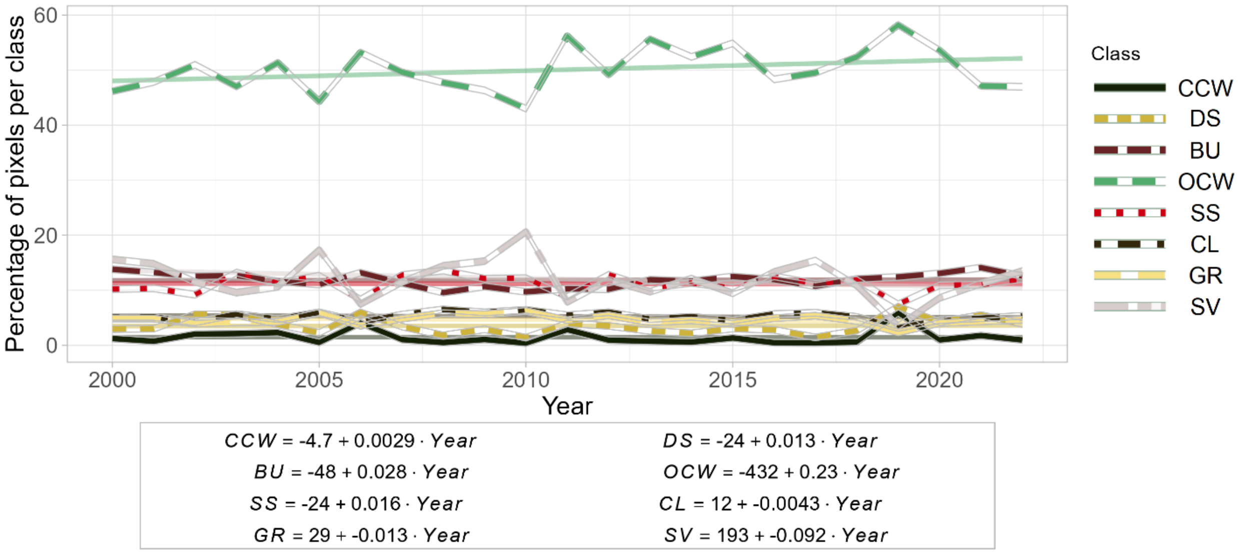

Figure 11: Annual time series of proportion of pixels of land cover classes for the entire study area (total 30 m pixel count = 858,780,117). Colored lines correspond to linear trends for each class over the study period.

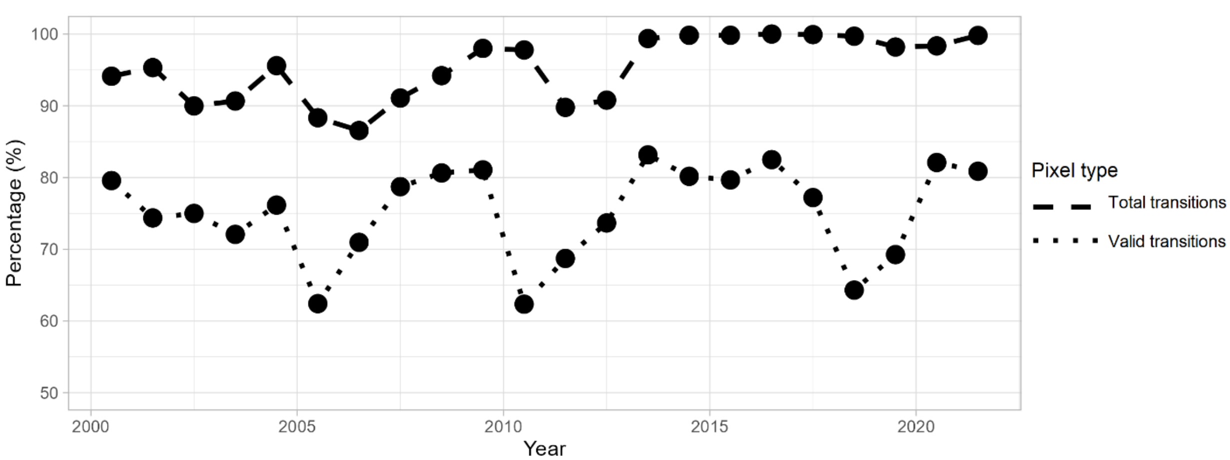

Figure 12: Proportion of pixels with potentially valid yearly transitions. Dashed and dotted lines show the total amount of paired unmasked land cover classes, and the total amount of potentially valid transitions as per our state transition model presented in Figure 7.

Filtering out pixels with unlikely transitions as defined in our state transition model, allows to reconstruct the history of individual pixels and help understand their change through time. The alluvial chart is a useful visualization to track such transitions through time by presenting the frequency distributions of classes in different time periods, aggregating the change of pixels with the same transitions between classes into individual ribbons. Figure 13 shows the decadal change of 48,280 randomly selected pixels with potential valid transitions and no missing data in our study area from 2000-2020. By assigning colors to the last year in the sequence, it is possible to visually track changes, evidenced by the width of the lines moving from one class to another between periods. The largest change of classes in this sample corresponds to 1.75 % of pixels (n = 845) staying as OCW in 2000 and 2010 but changing to CCW by the year 2020 (see dark ribbon going from OCW to CCW between 2010 to 2020). This is followed by 1.37 % of BU pixels (n = 661) turning into SV by the year 2010 and staying in that class until 2020 (see dark ribbon going from BU to SV between 2000 and 2010). Other classes present changes less than 1 %.

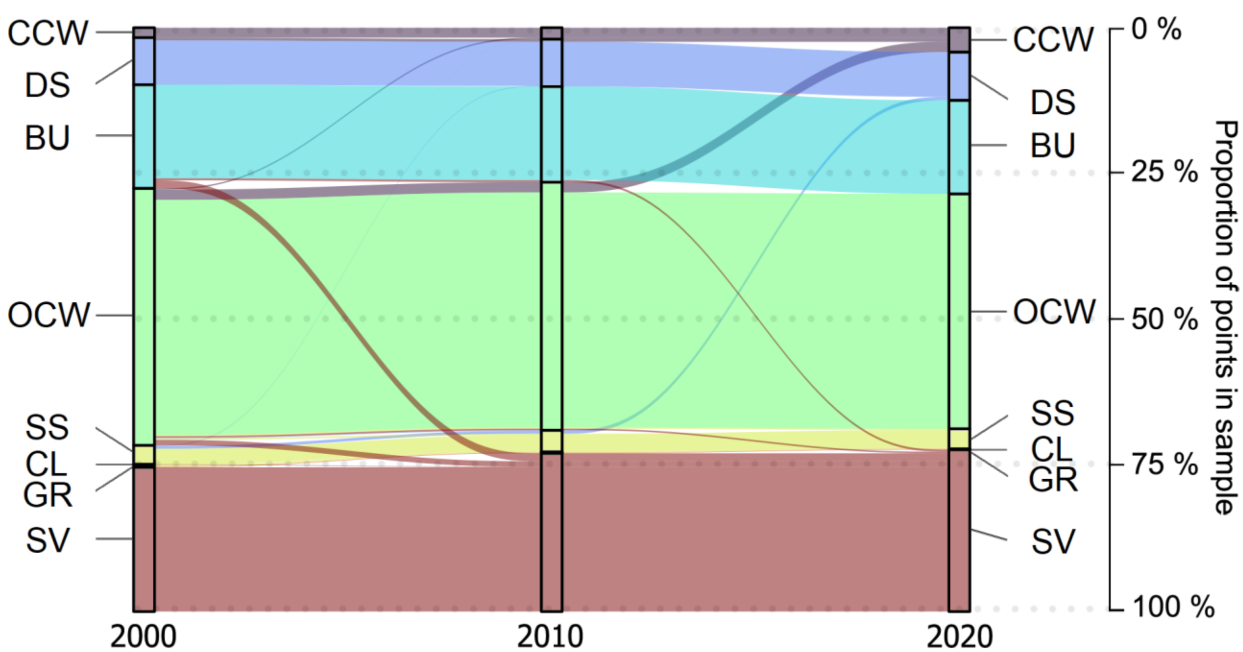

Figure 13: Decadal vegetation transition between 2000-2020 of 48,280 random pixels with potentially valid land cover transitions as defined in our state transition model for the three selected years. Land cover classes are presented in Table 1. Color codes were assigned to land cover classes present in the locations in year 2020 in order to track changes between decades.

3.2 Vegetation fractional cover estimation

Endmember estimation reached the threshold convexity error of -6 after 3,265 iterations, with total negative values representing just 0.026 % of the sample, reflecting excellent model fit and a very small proportion of sample points falling off the multidimensional space between endmembers (Weltje 1997). Figure 14 shows the estimated spectral signatures of endmembers, where a large spike in NIR is visible for PV and high values of reflectance at the SWIR bands are also discernible for BG. Regression results from the comparison between bare ground estimations from HR imagery and Landsat-based predictions yielded $R^2 = 0.795$, $p < 2.2e\text{-}16$, normalized root mean squared error nRMSE = 0.123, with equation $y = 0.959$ (SE = 0.010) $x + 5.768$ (SE = 0.843), F = 9201.1 on 1 and 2152 DF with p-value: < 2.2e-16 (Figure 15).

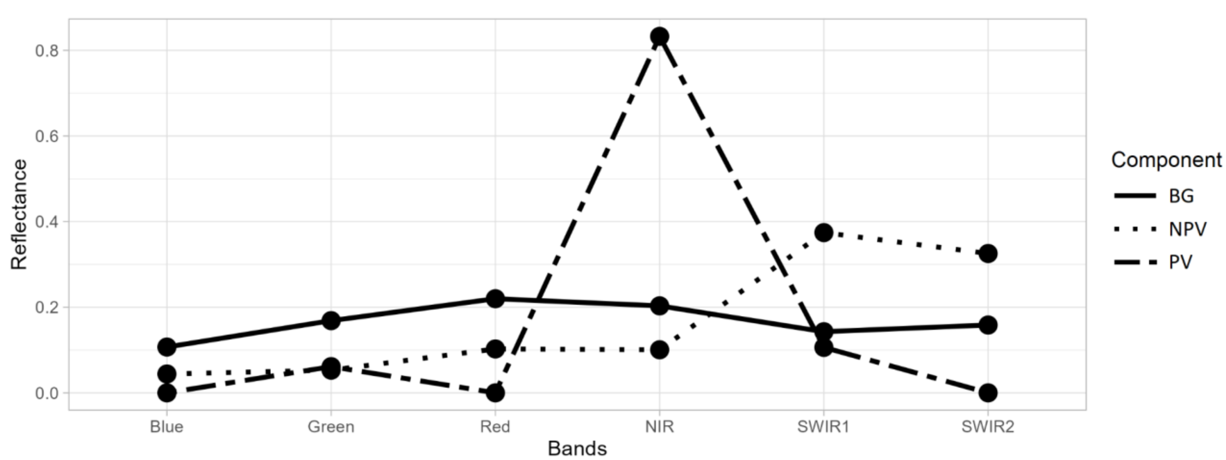

Figure 14: Estimated spectral endmembers for fractional cover estimation.

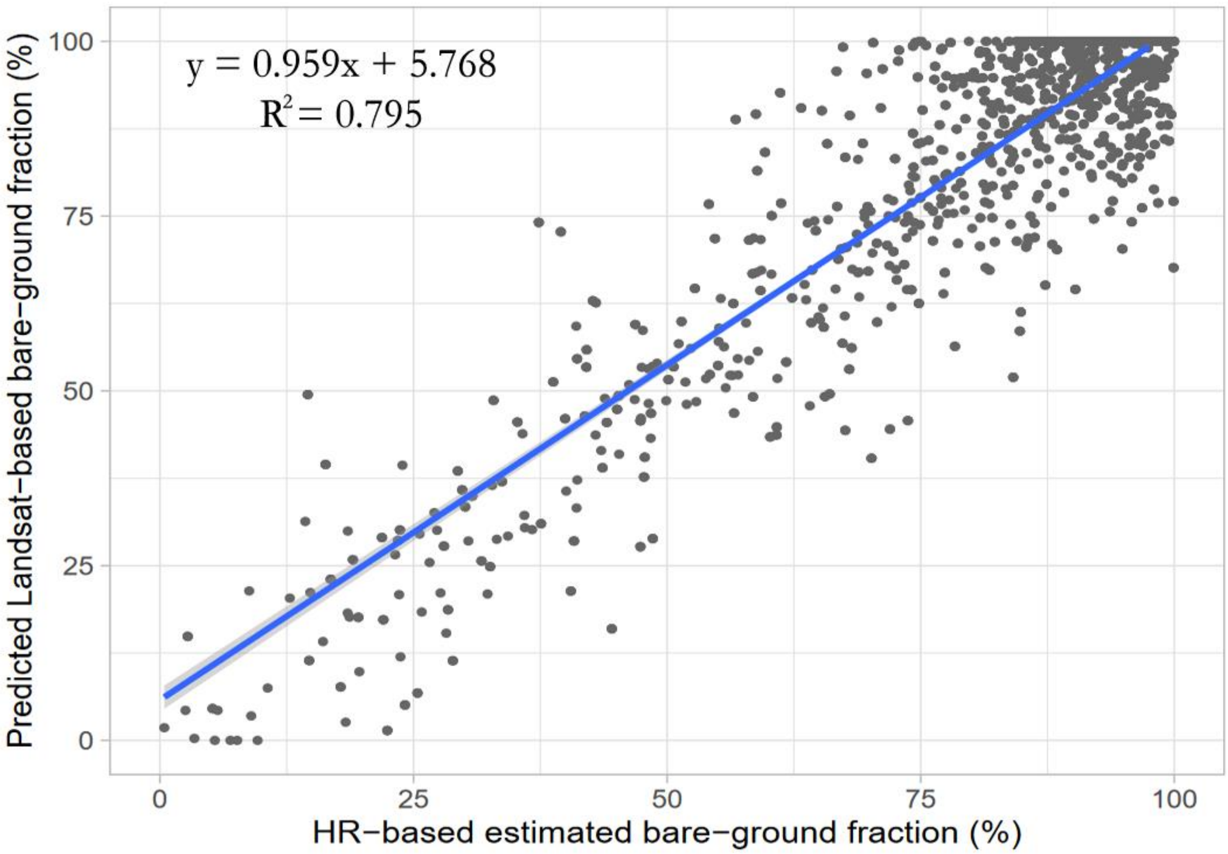

Figure 15: Spatial-temporal correlation between HR imagery and Landsat-based predictions of bare ground fractional cover (n = 2,190) at Landsat scale of 30 m from 2016 to 2020. F = 9,201.1 on 1 and 2152 DF with p-value: < 2.2e-16.

Final products consisted of yearly short dry season estimations of fractional cover for our entire AOI with a total of 858,780,117 pixels (Figure 16). Further qualitative assessment of fractional cover predictions against natural color Landsat images and compositions, confirmed accurate representations of the ground conditions. The most readily identifiable components BG and PV, show regional accordance with very dry and forested areas, respectively, within our AOI

(Figure 16). Similar to the LCC time series, fractional cover showed distinct variations in three different periods (Figure 17).

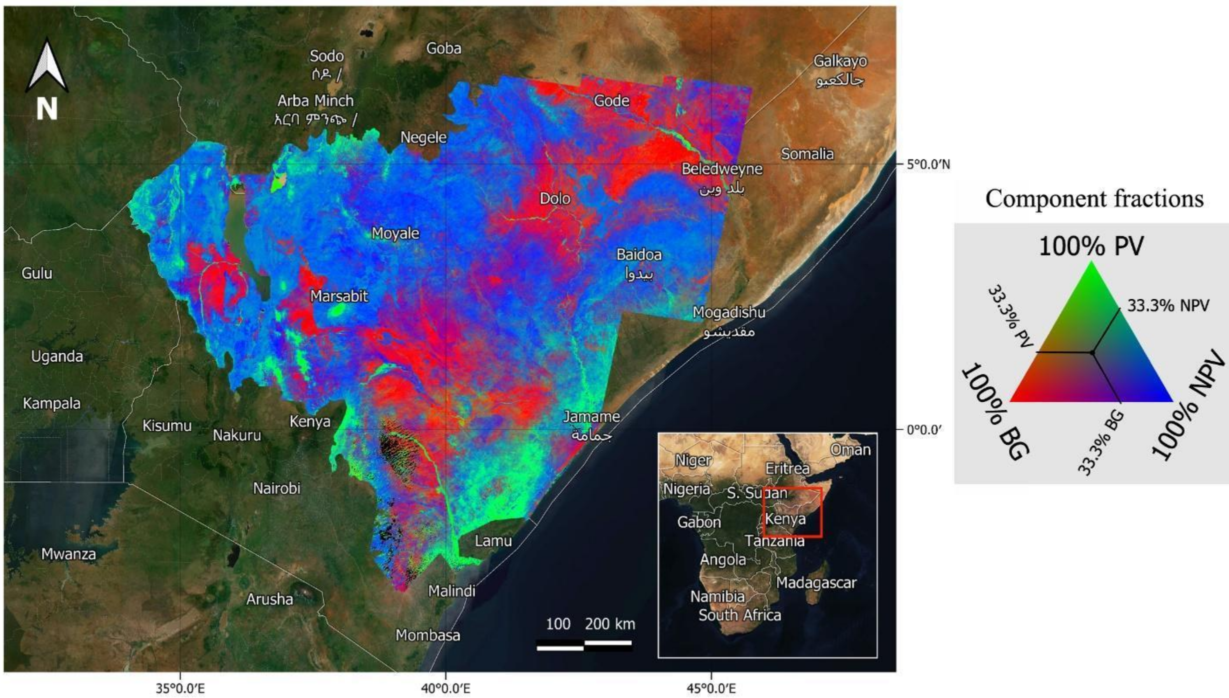

Figure 16: Landsat derived 30 m resolution fractional cover estimations for the short dry season of 2020, with mixtures of PV: Photosynthetic vegetation, NPV: non-photosynthetic vegetation, and BG: bare ground for our entire AOI (see legend on figure). Basemap: ©MapTiler, https://www.maptiler.com/copyright/.

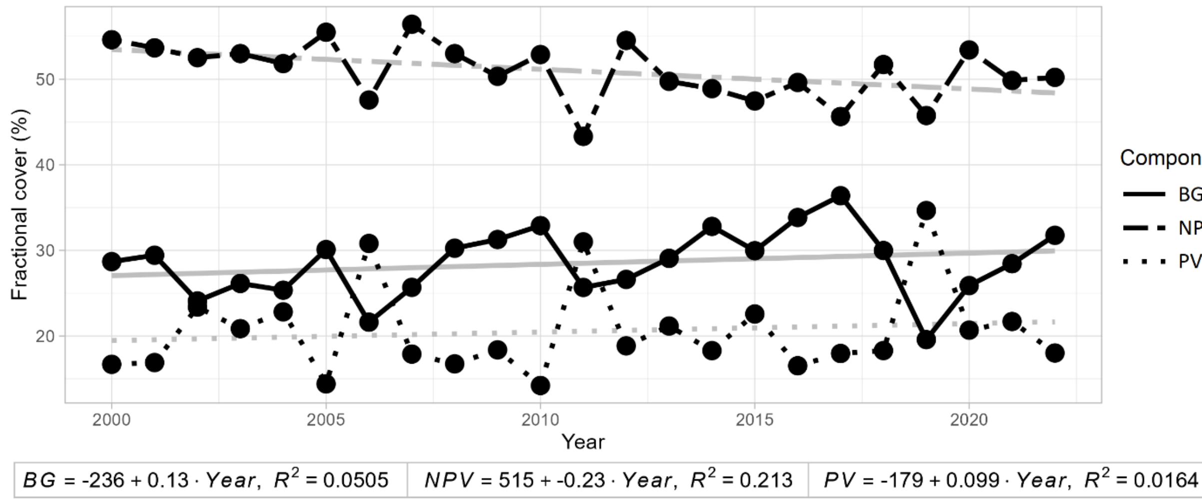

Figure 17: Annual time series of average fractional cover values for BG, NPV and PV for the entire study area (pixel count = 858,780,117). Straight lines correspond to linear trends for each component over the study period.

4. Discussion

The dataset generated in this study represents a substantial improvement over previously available data to assess rangeland health in the region, such as plain NDVI from Landsat and MODIS products. These improvements are the result of a high spatial resolution, a long temporal extent, and use of land and fractional cover metrics expressly designed to inform monitoring and assessment of East African rangeland systems (e.g. Hill and Guerschman 2022, Sexton et al., 2013, Buchhorn et al., 2021).

Our land cover classification scheme allowed us to reach acceptable per-class accuracy levels, using 85 % as a reference value for most of our land cover classes (Mundia and Aniya 2005, Rogan et al., 2003, Treitz and Rogan 2004, Weng 2002, Yang and Lo 2002), considering the limitations of both the availability of ground reference data and Landsat imagery. Our proposed method that used VHR imagery to generate training and validation data for the Landsat-based classification has proven to be key to reaching these accuracy levels, enabling us to increase the amplitude of spectral information of the different features found across such a large and heterogeneous area. VHR imagery also allowed us to have homogeneous spatial representation in ground-reference data as shown in Figure 2, thus reducing biases from imbalanced sampling (Carlotto 2009, Elmes et al., 2020). We also included a minimum threshold value for VHR classifications and applied a ruled-based algorithm to generate training data, therefore helping to reduce and control our training data error (Elmes et al., 2020, Padial-Iglesias et al., 2021). Homogenization of the VPI process also helped standardize training data generation, accounting for the arising inconsistencies that might impact the Landsat LCC estimations (Elmes et al., 2020, Foody 2009).

One limitation of our product is its comparatively lower classification accuracy for cultivated land areas. The close spectral correspondence between the dominant cultivated grain crops in the region (e.g., teff, maize and sorghum in Ethiopia) and wild grasses makes separation of the two challenging. In addition, other land classes such as sparse shrub could also be difficult to separate from cultivated land (Hansen et al 2005, Sexton et al 2013), because they are dominated by either PV or NPV during the short dry season where our Landsat compositions were compiled. These two factors limit the applicability of the proposed approach to extensive rangeland areas. We encourage users of this dataset to explore the behavior of the CL class within their study areas before carrying out further analyses. In addition, cloud cover in this region implies that other tools such as dynamic time warping (Muller 2007) might not improve land cover estimations, as this technique requires the extraction of temporal features from time series that are not possible to generate using Landsat imagery in our defined temporal extent. As with virtually all visible-light satellite-based remote sensing, cloud cover limits our analysis, both reducing the amount of per-pixel available imagery, and also the proportion of pixels with available data over our study area. Other factors such as precipitation resulted in a > 30 % drop in accuracy due to increases in annual accumulated precipitation, as found in our preliminary classifications.

In addition to class-specific issues, the multi-year classification scheme used here, has limitations and possible effects on the classification results on years without reference data, which can include misrepresentation of the real patterns. This study does not explore this effect due to the lack of in-situ reference data for the total length of the studied period. However, other studies in similar ecosystems where reference data is available, can help improve the products presented here or to find the possible biases they might have. Current research on the use of transfer learning, with the use of pre-trained models and fine-tuning with limited data provides very good opportunities for further improvement of remote sensing products and possible bias exploration (e.g. Li et al. 2023, Račič et al. 2024, Weikmann et al. 2021).

As shown in Figure 13, this dataset can not only provide descriptions of all the LC pixel transitions of a given study area but has the potential value of providing a foundation for assessments of long-term change trajectories that likely will extend beyond the time scope of the current study. Ecological studies on ecosystem and community dynamics require long-term ecological datasets (Ellis et al., 2006, Magurran et al., 2010, Ott et al., 2019). Further use of these products should demonstrate its usefulness as a monitoring, prioritization and inventory tool for planning and decision-making (Allred et al., 2022). Land cover mapping will enable isolating signals from rangelands and incorporate heterogeneity into management frameworks, providing foundations for assessments of long-term change trajectories that likely will extend beyond the time scope of the current study in this specific geographical region (Fuhlendorf et al., 2012).

Vegetation fractional cover estimates showed high accuracy. This accuracy is likely aided by the availability of VHR imagery (Brandt et al., 2020) used for generation of ground reference data for training and validation. Even under our limitations on ground reference data, bare ground, a key indicator of rangeland health conditions for monitoring and management (Pellant et al., 2020, Rigge et al., 2019, 2020), was accurately identified over a relatively large area of more than 4.6 million hectares. Figure 17 shows the potential value of this dataset by presenting a summarization of the annual trend of all three fractional components, which can be reconstructed from different spatiotemporal aggregations, down to the pixel level. Such trajectories will likely help understand the contributing factors for observed and unobserved patterns in the past two decades (Rigge et al., 2021). While further exploration of the spatial and temporal distribution of these trends is needed, this overall assessment might reflect a slow degradation of rangeland condition as bare ground fraction gradually increases (Figure 17).

Here, we used intensive algorithms on VHR satellite imagery to allow training and assessment of the performance of our proposed methods, as little ground reference information exists in this vast and remote region. This approach helps to maintain enough detail on the land cover classes and allowed the creation of a relevant VFC estimation. Our maps could help generate new threads of rangeland maps for East Africa, especially to improve community development, ecological conservation, and humanitarian programming. As the lack of ground reference data has been a bottleneck to empirical rangelands research in this part of the world, our VHR-based estimations can help develop and improve assessments of rangeland health trajectories. The increasing availability of remote sensing imagery and the application and development of new machine learning algorithms will certainly help develop better management tools. Relatively recent collections such as Sentinel-2 and its harmonization with Landsat imagery (Claverie et al.,

2018) will need to be tested for its advantages and disadvantages for its use in long-term time series in this geographic area.

The framework proposed here of harnessing VHR images to generate training labels in a semi-automatic procedure, including manual VPI and RCC to automatically create reference data based on class proportions will become highly relevant considering recent technological advancements. Modern tools such as large language models (LLMs) and foundation models carry a huge promise to improve generalizability of this approach and classification accuracies in complex landscapes. With the future use of these new tools and fine-tuning, we expect our models specifically trained in our study domain to be generalizable to other dryland/rangeland regions in the whole of Sub-Saharan Africa or other continents (e.g., Australia, parts of Central Asia), where ecosystems, land cover compositions, herding intensities, and other similar features exist.

Overall, this dataset will be useful to monitor the impacts of different rangeland management practices or test the impact of development programs. The open access to sophisticated cloud computing platforms, such as GEE (Gorelick et al., 2017), will contribute to practical use and further assessment of this dataset. To accomplish this, have made these two products available in GEE (see Data availability).

5. Data availability

Our 30 m resolution annual land cover classification and fractional cover data are publicly available at https://doi.org/10.5281/zenodo.7106166 (Soto et al., 2023) and Google Earth Engine (see Appendix A).

6. Competing interests

The contact author has declared that none of the authors has any competing interests.

7. Acknowledgements

This project is supported by Biodiversity International and the CGIAR Standing Panel on Impact Assessment (SPIA) for financial support under LoA L20HQ130. YS also acknowledges support from NASA-CMS award (80NSSC21K1058).

8. Author contributions

CB, PC, FF, NK, SW, NJ, CL, BP, and YS conceived the study. PC and CL conducted fieldwork. SW, PC, CL, BP, and GES conducted visual photo-interpretation work. GES performed the remote sensing analyses and wrote the first draft of the manuscript. All authors contributed to discussions and writing of the manuscript.

9. References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C.: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015, Sci. data, 5(1), 1-12, https://doi.org/10.1038/sdata.2017.191, 2018.

Adams, E. C., Parache, H. B., Cherrington, E., Ellenburg, W. L., Mishra, V., Lucey, R., & Nakalembe, C.: Limitations of remote sensing in assessing vegetation damage due to the 2019–2021 desert locust upsurge. Frontiers in Climate, 3, 714273, https://doi.org/10.3389/fclim.2021.714273, 2021.

AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., and Hain, C. R.: Remote sensing of drought: Progress, challenges and opportunities, Rev. Geophys., 53(2), 452-480, https://doi.org/10.1002/2014RG000456, 2015.

Allred, B. W., Bestelmeyer, B. T., Boyd, C. S., Brown, C., Davies, K. W., Duniway, M. C., Ellsworth, L. M., Erickson, T. A., Fuhlendorf, S. D., Griffiths, T. V., Jansen, V., Jones, M. O., Karl, J., Knight, A., Maestas, J. D., Maynard, J. J., McCord, S. E., Naugle, D. E., Starns,

H. D., Twidwell, D., and Uden, D. R.: Improving Landsat predictions of rangeland fractional cover with multitask learning and uncertainty, Methods Ecol. Evol., 12(5), 841-849, https://doi.org/10.1111/2041-210X.13564, 2021.

Allred, B. W., Creutzburg, M. K., Carlson, J. C., Cole, C. J., Dovichin, C. M., Duniway, M. C., Jones, M. O., Maestas, J. D., Naugle, D. E., Nauman, T. W., Okin, G. S., Reeves, M. C., Rigge, M., Savage, S. L., Twidwell, D., Uden, D. R., and Zhou, B.: Guiding principles for using satellite-derived maps in rangeland management, Rangelands, 44(1), 78-86, https://doi.org/10.1016/j.rala.2021.09.004, 2022.

Angassa, A.: Effects of grazing intensity and bush encroachment on herbaceous species and rangeland condition in southern Ethiopia, Land. Degrad. Dev., 25(5), 438-451, https://doi.org/10.1002/ldr.2160, 2014.

Angassa, A., and Oba, G.: Herder perceptions on impacts of range enclosures, crop farming, fire ban and bush encroachment on the rangelands of Borana, Southern Ethiopia, Hum. Ecol., 36(2), 201-215, https://doi.org/10.1007/s10745-007-9156-z, 2008.

Barbier, E. B., and Hochard, J. P.: Land degradation and poverty, Nat. Sustain., 1(11), 623-631, https://doi.org/10.1038/s41893-018-0155-4, 2018.

Beal, T., Gardner, C. D., Herrero, M., Iannotti, L. L., Merbold, L., Nordhagen, S., & Mottet, A.: Friend or foe? The role of animal-source foods in healthy and environmentally sustainable diets. The Journal of nutrition, 153(2), 409-425, https://doi.org/10.1016/j.tjnut.2022.10.016, 2023.

Belgiu, M., and Csillik, O.: Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis, Remote Sens. Env., 204, 509-523, https://doi.org/10.1016/j.rse.2017.10.005, 2018.

Bestelmeyer, B. T., Ash, A., Brown, J. R., Densambuu, B., Fernández-Giménez, M., Johanson, J., Levi, M., Lopez, D., Peinetti, R., Rumpff, L., and Shaver, P.: State and transition models: theory, applications, and challenges. Rangeland systems: Processes, management and challenges. In: Rangeland Systems. Springer Series on Environmental Management, edited by: Briske, D., Springer, Cham, 303-345, https://doi.org/10.1007/978-3-319-46709-2_9, 2017.

Blackwell, P. J.: East Africa's Pastoralist Emergency: is climate change the straw that breaks the camel's back?, Third World Q., 31(8), 1321-1338, https://doi.org/10.1080/01436597.2010.541085, 2010.

Boardman, J. W., Kruse, F.A., and Green, R.O. Mapping target signatures via partial unmixing of AVIRIS data. In Proceedings of the Summaries 5th JPL Airborne Earth Science Workshop, Pasadena, CA, USA, Volume 1, pp. 23-26, January 1995.

Brandt, M., Tucker, C. J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., Chave, J., Rasmussen, L. V., Hiernaux, P., Diouf, A. A., Kergoat, L., Mertz, O., Igel, C., Gieseke, F., Schöning, J., Li, S., Melocik, K., Meyer, J., Sinno, S., Romero, E., Glennie, E., Montagu, A., Dendoncker, M, and Fensholt, R.: An unexpectedly large count of trees in the West African Sahara and Sahel, Nature, 587(7832), 78-82, https://doi.org/10.1038/s41586-020-2824-5, 2020.

Browning, D. M., Maynard, J. J., Karl, J. W., and Peters, D. C.: Breaks in MODIS time series portend vegetation change: verification using long-term data in an arid grassland ecosystem, Ecol. Appl., 27(5), 1677-1693, https://doi.org/10.1002/eap.1561, 2017.

Browning, D. M., Snyder, K. A., and Herrick, J. E.: Plant phenology: Taking the pulse of rangelands, Rangelands, 41(3), 129-134, https://doi.org/10.1016/j.rala.2019.02.001, 2019.

Buchhorn, M., Smets, B., Bertels, L., De Roo, B., Lesiv, M., Tsendbazar, N.E., Linlin, L., Tarko, A.: Copernicus Global Land Service: Land Cover 100m: Version 3 Globe 2015-2019: Product User Manual, Zenodo, Geneve, Switzerland, 22 pp., https://doi.org/10.5281/zenodo.3938963, 2020.

Carlotto, M. J.: Effect of errors in ground truth on classification accuracy, Int. J. Remote Sens., 30(18), 4831-4849, https://doi.org/10.1080/01431160802672864, 2009.

Chantarat, S., Mude, A. G., Barrett, C. B., and Carter, M. R.: Designing index-based livestock insurance for managing asset risk in northern Kenya, J. Risk Insur., 80(1), 205-237, https://doi.org/10.1111/j.1539-6975.2012.01463.x, 2013.

Cheng, Y., Vrieling, A., Fava, F., Meroni, M., Marshall, M., and Gachoki, S.: Phenology of short vegetation cycles in a Kenyan rangeland from PlanetScope and Sentinel-2, Remote Sens. Environ., 248, 112004, https://doi.org/10.1016/j.rse.2020.112004, 2020.

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J. C., Skakun, S. V., and Justice, C.: The Harmonized Landsat and Sentinel-2 surface reflectance data set, Remote Sens. Environ., 219, 145-161, https://doi.org/10.1016/j.rse.2018.09.002, 2018.

Cooley, S. W., Smith, L. C., Stepan, L., and Mascaro, J.: Tracking dynamic northern surface water changes with high-frequency planet CubeSat imagery, Remote Sens., 9(12), 1306, https://doi.org/10.3390/rs9121306, 2017.

Dube, T., Shoko, C., Sibanda, M., Madileng, P., Maluleke, X. G., Mokwatedi, V. R., Tibane, L., and Tshebesebe, T.: Remote sensing of invasive *Lantana camara* (*verbenaceae*) in semiarid savanna rangeland ecosystems of south africa, Rangeland Ecol. Manag., 73(3), 411-419, https://doi.org/10.1016/j.rama.2020.01.003, 2020.

Ellis, E. C., Wang, H., Xiao, H. S., Peng, K., Liu, X. P., Li, S. C., Ouyang, H., Cheng, X., and Yang, L. Z.: Measuring long-term ecological changes in densely populated landscapes using current and historical high resolution imagery, Remote Sens. Environ., 100(4), 457-473, https://doi.org/10.1016/j.rse.2005.11.002, 2006.

Elmes, A., Alemohammad, H., Avery, R., Caylor, K., Eastman, J. R., Fishgold, L., Friedl, M. A., Jain, M., Kohli, D., Laso Bayas, J. C., Lunga, D., McCarty, J. L., Pontius, R. G., Reinmann, A. B., Rogan, J., Song, L., Stoynova, H., Ye, S., Yi, Z. F., and Estes, L.: Accounting for training data error in machine learning applied to Earth observations, Remote Sens., 12(6), 1034, https://doi.org/10.3390/rs12061034, 2020.

Fava, F. P., Jensen, N. D., Sina, J., Mude, A. G., and Maher, B.: Building financial resilience in pastoral communities in Africa: Lessons learned from implementing the Kenya Livestock Insurance Program (KLIP), World Bank, Washington DC, 32 pp., https://hdl.handle.net/10568/111681, 2021.

Fava, F., and Vrieling, A.: Earth observation for drought risk financing in pastoral systems of sub-Saharan Africa, Curr. Opin. Env. Sust., 48, 44-52, https://doi.org/10.1016/j.cosust.2020.09.006, 2021.

Foga, S., Scaramuzza, P. L., Guo, S., Zhu, Z., Dilley Jr, R. D., Beckmann, T., Schmidt, G. L., Dwyer, J. L., Hughes, M. J., and Laue, B.: Cloud detection algorithm comparison and validation for operational Landsat data products, Remote Sens. Environ., 194, 379-390, https://doi.org/10.1016/j.rse.2017.03.026, 2017.

Foody, G. M.: Status of land cover classification accuracy assessment, Remote Sens. Environ., 80(1), 185-201, https://doi.org/10.1016/S0034-4257(01)00295-4, 2002.

Foody, G. M.: The impact of imperfect ground reference data on the accuracy of land cover change estimation, Int. J. Remote Sens., 30(12), 3275-3281, https://doi.org/10.1080/01431160902755346, 2009.

Franks, S., Storey, J., and Rengarajan, R.: The new Landsat collection-2 digital elevation model, Remote Sens., 12(23), 3909, https://doi.org/10.3390/rs12233909, 2020.

Fuhlendorf, S. D., Engle, D. M., Elmore, R. D., Limb, R. F., and Bidwell, T. G.: Conservation of pattern and process: developing an alternative paradigm of rangeland management, Rangeland Ecol. Manag., 65(6), 579-589, https://doi.org/10.2111/REM-D-11-00109.1, 2012.

Gao, B. C.: NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space, Remote Sens. Environ., 58(3), 257-266, https://doi.org/10.1016/S0034-4257(96)00067-3, 1996.

Ghafari, S., Ghorbani, A., Moameri, M., Mostafazadeh, R., and Bidarlord, M.: Composition and structure of species along altitude gradient in Moghan-Sabalan rangelands, Iran, J. Mt. Sci., 15(6), 1209-1228, https://doi.org/10.1007/s11629-017-4820-2, 2018.

Giuliani, G., Mazzetti, P., Santoro, M., Nativi, S., Van Bemmelen, J., Colangeli, G., and Lehmann, A.: Knowledge generation using satellite earth observations to support sustainable development goals (SDG): A use case on Land degradation, Int. J. Appl. Earth Obs., 88, 102068, https://doi.org/10.1016/j.jag.2020.102068, 2020.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sens. Environ., 202, 18-27, https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Graetz, R. D., Carneggie, D. M., Hacker, R., Lendon, C., and Wilcox, D. G.: A quantitative evaluation of Landsat imagery of Australian rangelands, Rangel. J., 1(1), 53-59, https://doi.org/10.1071/RJ9760053, 1976.

Haas, R. H., Deering, D. W., Rouse Jr, J. W., and Schell, J. A.: Monitoring vegetation conditions from Landsat for use in range management. In NASA Earth Resources Survey Symposium Proc., Houston, Texas, Vol. 1, pp. 43-52, 1975.

Hill, M. J., and Guerschman, J. P.: Global trends in vegetation fractional cover: Hotspots for change in bare soil and non-photosynthetic vegetation. Agric. Ecosyst. Environ., 324, 107719, https://doi.org/10.1016/j.agee.2021.107719, 2022.

Hill, M. J., and Guerschman, J. P.: The MODIS Global Vegetation Fractional Cover Product 2001–2018: Characteristics of Vegetation Fractional Cover in Grasslands and Savanna Woodlands, Remote Sens., 12, 406, https://doi.org/10.3390/rs12030406, 2020.

Hoffman, T., and Vogel, C.: Climate change impacts on African rangelands, Rangel., 30(3), 12-17, https://doi.org/10.2111/1551-501X(2008)30[12:CCIOAR]2.0.CO;2, 2008.

ILRI, IUCN, FAO, WWF, UNEP and ILC.: Rangelands Atlas. Nairobi Kenya: ILRI, 2021.

IPCC: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., https://doi.org/10.1017/9781009325844, 2022.

Jarvis, A., Reuter, H. I., Nelson, A., Guevara, E.: Hole-filled seamless SRTM data Version 4, available from the CGIAR-CSI SRTM 90m, International Centre for Tropical Agriculture CIAT, Available from: https://srtm.csi.cgiar.org, 2008.

Jensen, N. D., Barrett, C. B., and Mude, A. G. (2017). Cash transfers and index insurance: A comparative impact analysis from northern Kenya, JJ. Dev. Econ., 129, 14-28, https://doi.org/10.1016/j.jdeveco.2017.08.002, 2017.

Jensen, N., Sheahan, M., Barrett, C. B., and Mude, A.: Hunger Safety Net Program (HSNP) and Index Based Livestock Insurance (IBLI) baseline comparison, Available from: http://dyson.cornell.edu/faculty_sites/cbb2/research/datasets/HSNP_IBLI_Comparison.pdf, 2014.

Johnson, L., Wandera, B., Jensen, N., and Banerjee, R.: Competing expectations in an index-based livestock insurance project, J. Dev. Stud., 55(6), 1221-1239, https://doi.org/10.1080/00220388.2018.1453603, 2019.

Jones, M. O., Allred, B. W., Naugle, D. E., Maestas, J. D., Donnelly, P., Metz, L. J., Karl, J., Smith, R., Bestelmeyer, B., Boyd, C., Kerby, J. D., and McIver, J. D.: Innovation in rangeland monitoring: annual, 30 m, plant functional type percent cover maps for US rangelands, 1984–2017, Ecosphere, 9(9), e02430, https://doi.org/10.1002/ecs2.2430, 2018.

Jones, M. O., Naugle, D. E., Twidwell, D., Uden, D. R., Maestas, J. D., and Allred, B. W.: Beyond inventories: Emergence of a new era in rangeland monitoring, Rangeland Ecol. Manag., 73(5), 577-583, https://doi.org/10.1016/j.rama.2020.06.009, 2020.

Li, G., Han, W., Dong, Y., Zhai, X., Huang, S., Ma, W., Cui, X. & Wang, Y.: Multi-Year Crop Type Mapping Using Sentinel-2 Imagery and Deep Semantic Segmentation Algorithm in the

Hetao Irrigation District in China. Remote Sensing, 15(4), 875, https://doi.org/10.3390/rs15040875, 2023.

Liao, C., & Fei, D.: Pastoralist Adaptation Practices under Non-Governmental Development Interventions in Southern Ethiopia. The Rangeland Journal, 39 (2): 189-200, https://doi.org/10.1071/RJ16015, 2017.

Liao, C., and Clark, P. E.: Rangeland vegetation diversity and transition pathways under indigenous pastoralist management regimes in southern Ethiopia, Agric. Ecosyst. Environ., 252, 105-113, https://doi.org/10.1016/j.agee.2017.10.009, 2018.

Liao, C., Clark, P. E., and DeGloria, S. D.: Bush encroachment dynamics and rangeland management implications in southern Ethiopia, Ecol. Evol., 8(23), 11694-11703, https://doi.org/10.1002/ece3.4621, 2018.

Lin, C., Jin, Z., Mulla, D., Ghosh, R., Guan, K., Kumar, V., and Cai, Y.: Toward Large-Scale Mapping of Tree Crops with High-Resolution Satellite Imagery and Deep Learning Algorithms: A Case Study of Olive Orchards in Morocco, Remote Sens., 13(9), 1740, https://doi.org/10.3390/rs13091740, 2021.

Liu, H. Q., and Huete, A.: A feedback based modification of the NDVI to minimize canopy background and atmospheric noise, IEEE T. Geosci. Remote., 33(2), 457-465, https://doi.org/10.1109/TGRS.1995.8746027, 1995.

Loveland, T. R., and Dwyer, J. L.: Landsat: Building a strong future, Remote Sens. Environ., 122, 22-29, https://doi.org/10.1016/j.rse.2011.09.022, 2012.

Magurran, A. E., Baillie, S. R., Buckland, S. T., Dick, J. M., Elston, D. A., Scott, E. M., Smith, R. I., Somerfield, P. J., and Watt, A. D.: Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time, Trends Ecol. Evol., 25:574–582, https://doi.org/10.1016/j.tree.2010.06.016, 2010.

Markham, B. L., Storey, J. C., Williams, D. L., and Irons, J. R. Landsat sensor performance: history and current status. IEEE Trans Geosci Remote Sens, 42(12), 2691-2694, https://doi.org/10.1109/TGRS.2004.840720, 2004.

Matongera, T. N., Mutanga, O., Sibanda, M., and Odindi, J.: Estimating and Monitoring Land Surface Phenology in Rangelands: A Review of Progress and Challenges, Remote Sens., 13(11), 2060, https://doi.org/10.3390/rs13112060, 2021.

McFeeters, S. K.: The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, Int. J. Remote Sens., 17(7), 1425-1432, https://doi.org/10.1080/01431169608948714, 1996.

Müller, M.: Information retrieval for music and motion, Springer, Berlin, 313 pp., ISBN 978-3-540-74048-3, https://doi.org/10.1007/978-3-540-74048-3, 2007.

Mundia, C. N., and Aniya, M.: Analysis of land use/cover changes and urban expansion of Nairobi city using remote sensing and GIS, Int. J. Remote Sens., 26(13), 2831-2849, https://doi.org/10.1080/01431160500117865, 2005.

Nandintsetseg, B., Chang, J., Sen, O. L., Reyer, C. P., Kong, K., Yetemen, O., Ciais, P. & Davaadalai, J.: Future drought risk and adaptation of pastoralism in Eurasian rangelands. npj climate and atmospheric science, 7(1), 82, https://doi.org/10.1038/s41612-024-00624-2, 2024.

Oba, G., Weladji, R. B., Lusigi, W. J., and Stenseth, N. C.: Scale-dependent effects of grazing on rangeland degradation in northern Kenya: a test of equilibrium and non-equilibrium hypotheses, Land Degrad. Dev., 14(1), 83-94, https://doi.org/10.1002/ldr.524, 2003.

Okal, H. A., Ngetich, F. K., & Okeyo, J. M.: Spatio-temporal characterization of droughts using selected indices in Upper Tana River watershed, Kenya. Scientific African, 7, e00275, https://doi.org/10.1016/j.sciaf.2020.e00275, 2020.

Ott, J. E., Kilkenny, F. F., Summers, D. D., and Thompson, T. W.: Long-term vegetation recovery and invasive annual suppression in native and introduced postfire seeding treatments, Rangeland Ecol. Manag., 72, 640–653, https://doi.org/10.1016/j.rama.2019.02.001, 2019.

Padial-Iglesias, M., Serra, P., Ninyerola, M., and Pons, X.: A Framework of Filtering Rules over Ground Truth Samples to Achieve Higher Accuracy in Land Cover Maps, Remote Sens., 13(14), 2662, https://doi.org/10.3390/rs13142662, 2021.

Pellant, M., Shaver, P.L., Pyke, D.A., Herrick, J.E., Lepak, N., Riegel, G., Kachergis, E., Newingham, B.A., Toledo, D., and Busby, F.E.: Interpreting Indicators of Rangeland Health, Version 5, Tech Ref 1734-6. U.S. Department of the Interior, Bureau of Land Management, National Operations Center, Denver, CO, Available from: https://pubs.er.usgs.gov/publication/70215720, 2020.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J., and Stenseth, N. C.: Using the satellite-derived NDVI to assess ecological responses to environmental change, Trends Eco. Evol., 20(9), 503-510, https://doi.org/10.1016/j.tree.2005.05.011, 2005.

Pratt, D. J., Greenway, P. J., and Gwynne, M. D.: A classification of East African rangeland, with an appendix on terminology, J. Appl. Ecol., 369-382, https://doi.org/10.2307/2401259, 1966.

Pricope, N. G., Husak, G., Lopez-Carr, D., Funk, C., & Michaelsen, J.: The climate-population nexus in the East African Horn: Emerging degradation trends in rangeland and pastoral livelihood zones. Global environmental change, 23(6), 1525-1541, https://doi.org/10.1016/j.gloenvcha.2013.10.002, 2013.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., and Sorooshian, S.: A modified soil adjusted vegetation index. Remote sensing of environment, 48(2), 119-126, https://doi.org/10.1016/0034-4257(94)90134-1, 1994.

Quintano, C., Fernández-Manso, A., Shimabukuro, Y. E., and Pereira, G.: Spectral unmixing, Int. J. Remote Sens., 33(17), 5307-5340, https://doi.org/10.1080/01431161.2012.661095, 2012.

Račič, M., Oštir, K., Zupanc, A., & Čehovin Zajc, L.: Multi-Year Time Series Transfer Learning: Application of Early Crop Classification. Remote Sensing, 16(2), 270, https://doi.org/10.3390/rs16020270, 2024.

Reeves, M. C., and Baggett, L. S.: A remote sensing protocol for identifying rangelands with degraded productive capacity, Ecol. Indic., 43, 172-182, https://doi.org/10.1016/j.ecolind.2014.02.009, 2014.

Reid, R. S., Galvin, K. A., and Kruska, R. S.: Global significance of extensive grazing lands and pastoral societies: an introduction, in: Fragmentation in semi-arid and arid landscapes, edited by: Galvin, K. A., Reid, R. S, Behnke Jr., R. H., and Hobbs, N. T., Springer, Dordrecht, 1-24, https://doi.org/10.1007/978-1-4020-4906-4, 2008.

Rigge, M., Homer, C., Cleeves, L., Meyer, D. K., Bunde, B., Shi, H., Xian, G., Schell, S. and Bobo, M.: Quantifying western US rangelands as fractional components with multi-resolution remote sensing and in situ data, Remote Sens., 12(3), 412, https://doi.org/10.3390/rs12030412, 2020.

Rigge, M., Homer, C., Shi, H., Meyer, D., Bunde, B., Granneman, B., Postma, K., Danielson, P., Case, A., and Xian, G.: Rangeland fractional components across the Western United States from 1985 to 2018, Remote Sens., 13(4), 813, https://doi.org/10.3390/rs13040813, 2021.

Rigge, M., Shi, H., Homer, C., Danielson, P., and Granneman, B.: Long-term trajectories of fractional component change in the Northern Great Basin, USA, Ecosphere, 10(6), e02762, https://doi.org/10.1002/ecs2.2762, 2019.

Rogan, J., Miller, J., Stow, D., Franklin, J., Levien, L., and Fischer, C.: Land-cover change monitoring with classification trees using Landsat TM and ancillary data, Photogramm. Eng. Rem. Sens., 69(7), 793-804, https://doi.org/10.14358/PERS.69.7.793, 2003.

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W.: Monitoring vegetation systems in the Great Plains with ERTS, in: NASA special publication Third Earth Resoources Technology Satellite – 1 symposium, edited by: Freden, M. A., Mercanti, E. P., and Becker, M. A., NASA, Washington DC., 309-317, 1974.

Roy, D. P., Kovalskyy, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., and Egorov, A.: Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity, Remote Sens. Environ., 185, 57-70, https://doi.org/10.1016/j.rse.2015.12.024, 2016.

Safanelli, J. L., Poppiel, R. R., Ruiz, L. F. C., Bonfatti, B. R., Oliveira Mello, F. A. D., Rizzo, R., and Demattê, J. A.: Terrain Analysis in Google Earth Engine: A Method Adapted for High-Performance Global-Scale Analysis, ISPRS Int. Geo-Inf., 9(6), 400, https://doi.org/10.3390/ijgi9060400, 2020.

Sankey, T. T., Leonard, J. M., and Moore, M. M.: Unmanned aerial vehicle−based rangeland monitoring: examining a century of vegetation changes, Rangeland Ecol. Manag., 72(5), 858-863, https://doi.org/10.1016/j.rama.2019.04.002, 2019.

Sankey, T. T., Leonard, J., Moore, M. M., Sankey, J. B., and Belmonte, A.: Carbon and ecohydrological priorities in managing woody encroachment: UAV perspective 63 years after a control treatment, Environ. Res. Lett., 16(12), 124053, https://doi.org/10.1088/1748-9326/ac3796, 2021.

Sayre, N. F., McAllister, R. R., Bestelmeyer, B. T., Moritz, M., and Turner, M. D.: Earth stewardship of rangelands: coping with ecological, economic, and political marginality, Front. Ecol. Environ., 11(7), 348-354, https://doi.org/10.1890/120333, 2013.

Schmidt, G. L., Jenkerson, C., Masek, J. G., Vermote, E., and Gao, F.: Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description, 17 pp., Available from: https://pubs.usgs.gov/of/2013/1057/, 2013.

Seidel, M., and Hlawitschka, M.: An R-Based function for modeling of end member compositions, Math. Geosci., 47(8), 995-1007, https://doi.org/10.1007/s11004-015-9609-7, 2015.

Sexton, J. O., Song, X. P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D. H., Collins, K. M., Channan, S., DiMiceli, C., and Townshend, J. R.: Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. International Journal of Digital Earth, 6(5), 427-448, https://doi.org/10.1080/17538947.2013.786146, 2013.

Soto, G. E., Wilcox, S., Clark, P. E., Fava, F. P., Jensen, N. M., Kahiu, N., Liao, C., Porter, B., Sun, Y., and Barrett, C. B.: Mapping Rangeland Health Indicators in East Africa from 2000 to 2022 [data set], https://doi.org/10.5281/zenodo.7106166, 2023.

Sudmanns, M., Tiede, D., Lang, S., Bergstedt, H., Trost, G., Augustin, H., Baraldi, A., and Blaschke, T.: Big Earth data: disruptive changes in Earth observation data management and analysis?, Int. J. Digit. Earth, 13(7), 832-850, https://doi.org/10.1080/17538947.2019.1585976, 2020.

Treitz, P., and Rogan, J.: Remote sensing for mapping and monitoring land-cover and land-use change-an introduction, Prog. Plann., 61(4), 269-279, https://doi.org/10.1016/S0305-9006(03)00066-7, 2004.

U.S. Geological Survey: Landsat 8 Data Users Handbook, 114 pp., Sioux falls, S.D., Available from: https://www.usgs.gov/media/files/landsat-8-data-users-handbook, 2019.

U.S. Geological Survey.: Landsat Collection 2 (ver. 1.1, January 15, 2021): U.S. Geological Survey Fact Sheet 2021–3002, 4 pp., Sioux falls, S.D., https://doi.org/10.3133/fs20213002, 2021.

U.S. Geological Survey.: Landsat 8-9 Collection 2 (C2) Level 2 Science Product (L2SP) Guide. Version 5.0., 43 pp., Sioux falls, S.D., Available from : https://www.usgs.gov/media/files/landsat-8-9-collection-2-level-2-science-product-guide, 2022.

Vetter, S.: Rangelands at equilibrium and non-equilibrium: recent developments in the debate, J. Arid Environ., 62(2), 321-341, https://doi.org/10.1016/j.jaridenv.2004.11.015, 2005.

Weikmann, G., Paris, C., & Bruzzone, L.: Multi-year crop type mapping using pre-trained deep long-short term memory and Sentinel 2 image time series. In Image and Signal Processing for Remote Sensing XXVII, Vol. 11862, pp. 171-181, SPIE, 2021.

Weltje, G. J.: End-member modeling of compositional data: Numerical-statistical algorithms for solving the explicit mixing problem, Math. Geol., 29(4), 503-549, https://doi.org/10.1007/BF02775085, 1997.

Weng, Q.: Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling, J. Environ. Manag., 64(3), 273-284, https://doi.org/10.1006/jema.2001.0509, 2002.

Williams, D. L., Goward, S., and Arvidson, T.: Landsat: Yesterday, today and tomorrow, Photogramm. Eng. Rem. Sens., 72(10), 1171-1178, https://doi.org/10.14358/PERS.72.10.1171, 2006.

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Alle, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J. D., Hostert, P., Hughes, M. J., Huntington, J., Johnson, D. M., Kennedy, R., Kilic, A., Li, Z., Lymburner, L., McCorkel, J., Pahlevan, N., Scambos, T. A., Schaaf, C., Schott, J. R., Sheng, Y., Storey, J., Vermotev, E., Vogelmann, JJ., White, J. C., Wynne, R. H., and Zhu, Z.: Current status of Landsat program, science, and applications, Remote Sens. Environ., 225, 127-147, https://doi.org/10.1016/j.rse.2019.02.015, 2019.

Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., and Woodcock, C. E.: Opening the archive: How free data has enabled the science and monitoring promise of Landsat, Remote Sens. Environ., 122, 2-10, https://doi.org/10.1016/j.rse.2012.01.010, 2012.

Wulder, M. A., Roy, D. P., Radeloff, V. C., Loveland, T. R., Anderson, M. C., Johnson, D. M., Healey, S., Zhu, Z., Scambos, T. A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C. J., Masek, J. G., Hermosilla, T., White, J. C., Belward, A. S., Schaaf, C., Woodcock, C. E., Huntington, J. L., Lymburner, L., Hostert, P., Gao, F., Lyapustin, A., Pekel, J. F., Strobl, P., and Cook, B. D.: Fifty years of Landsat science and impacts, Remote Sens. Environ., 280, 113195, https://doi.org/10.1016/j.rse.2022.113195, 2022.

Wynants, M., Kelly, C., Mtei, K., Munishi, L., Patrick, A., Rabinovich, A., Nasseri, M., Gilvear, D., Roberts, N., Boeckx, P., Wilson, G., Blake, W. H., and Ndakidemi, P.: Drivers of increased soil erosion in East Africa’s agro-pastoral systems: changing interactions between the social, economic and natural domains, Reg. Environ. Change, 19(7), 1909-1921, https://doi.org/10.1007/s10113-019-01520-9, 2019.

Yang, X., and Lo, C. P.: Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area, Int. J. Remote Sens., 23(9), 1775-1798, https://doi.org/10.1080/01431160110075802, 2002.

Zabel, F., Delzeit, R., Schneider, J. M., Seppelt, R., Mauser, W., and Václavík, T.: Global impacts of future cropland expansion and intensification on agricultural markets and biodiversity, Nat. Comm., 10(1), 1-10, https://doi.org/10.1038/s41467-019-10775-z, 2019.

Zhou, B., Okin, G. S., and Zhang, J.: Leveraging Google Earth Engine (GEE) and machine learning algorithms to incorporate in situ measurement from different times for rangelands monitoring, Remote Sens. Environ., 236, 111521, https://doi.org/10.1016/j.rse.2019.111521, 2020.

10. Appendices

Appendix A. Description of access to Google Earth Engine (GEE) data.

Land Cover Classification data can be accessed using GEE's asset ids with the following structure:

- projects/ee-gerardosoto/assets/lcClass<YEAR>
- For example, for year 2000, use: "projects/ee-gerardosoto/assets/lcClass2000"

Alternatively, use the GEE's links as follows:
https://code.earthengine.google.com/?asset=projects/ee-gerardosoto/assets/lcClass2000

Vegetation Fractional Cover data can be accessed using GEE's asset ids with the following structure:

- projects/ee-gerardosoto/assets/fracCov<YEAR>_int16
- For example, for year 2000, use: "projects/ee-gerardosoto/assets/fracCov2000_int16"

Alternatively, use the GEE's links as follows:
https://code.earthengine.google.com/?asset=projects/ee-gerardosoto/assets/fracCov2000_int16

Appendix B. Reference flash card sets.

The following pages include the flash cards used to reference land cover types and canopy cover.

Figure B1. Flashcard for land cover type "Closed Canopy Woodland".

Figure B2. Flashcard for land cover type “Dense Scrubland”.

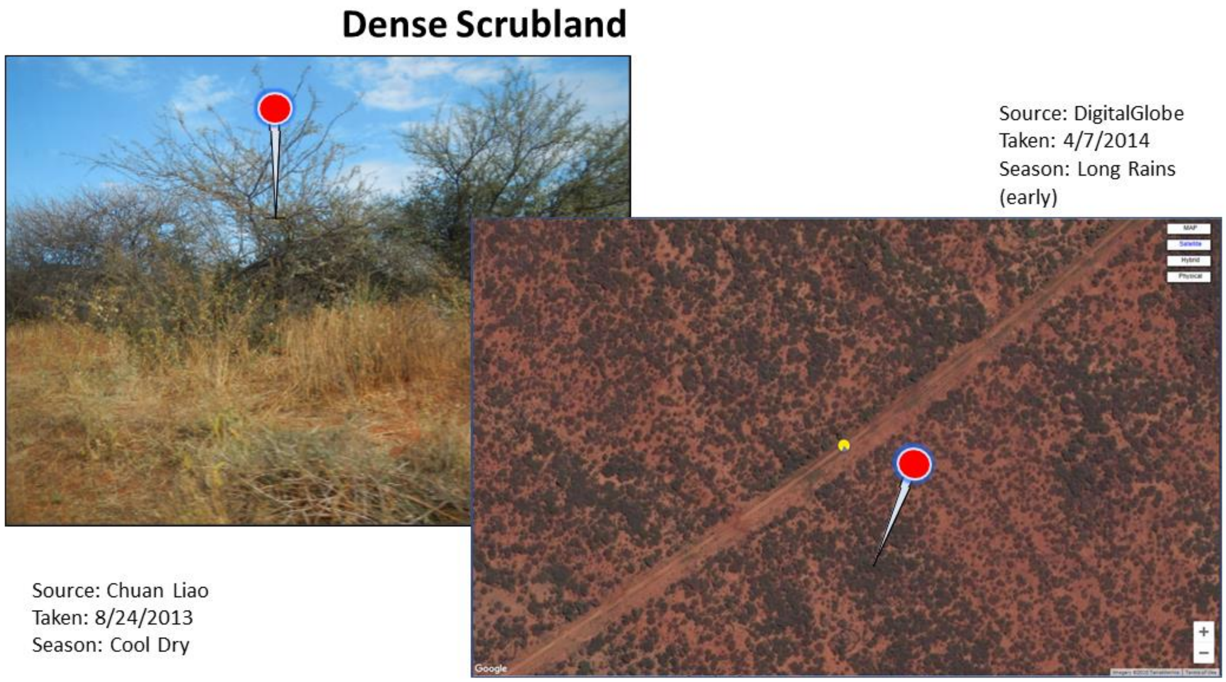

Figure B3. Flashcard for land cover type "Bushland".

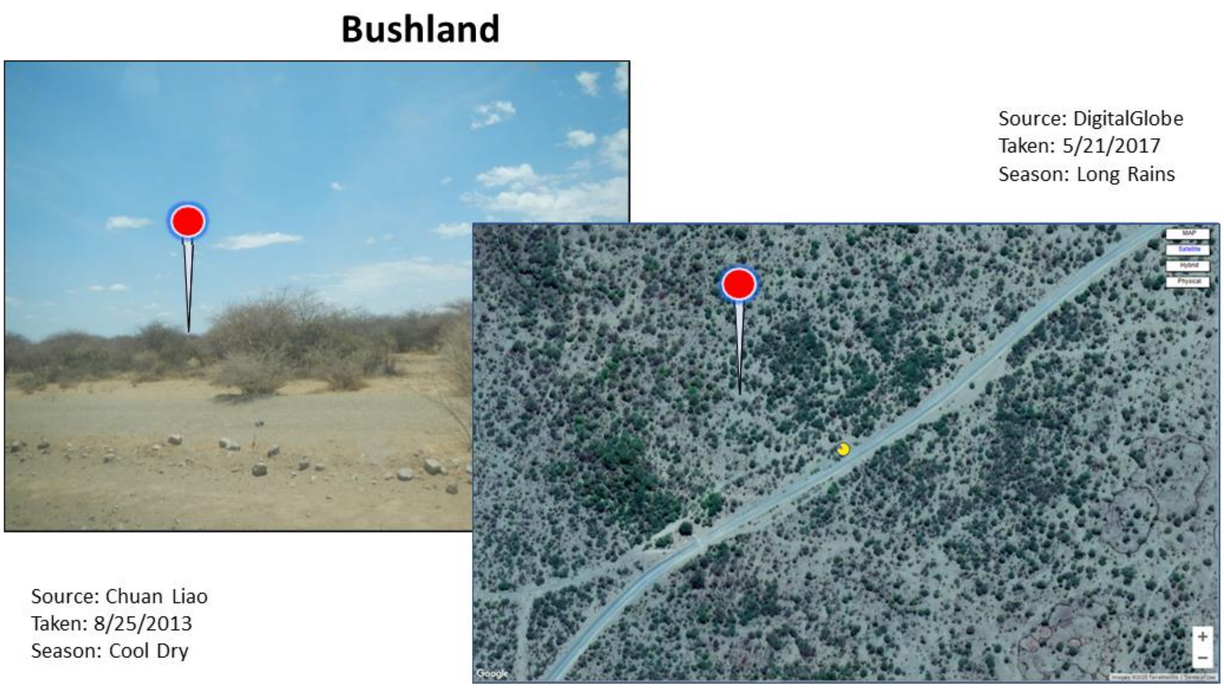

Figure B4. Flashcard for land cover type "Open Canopy Woodland".

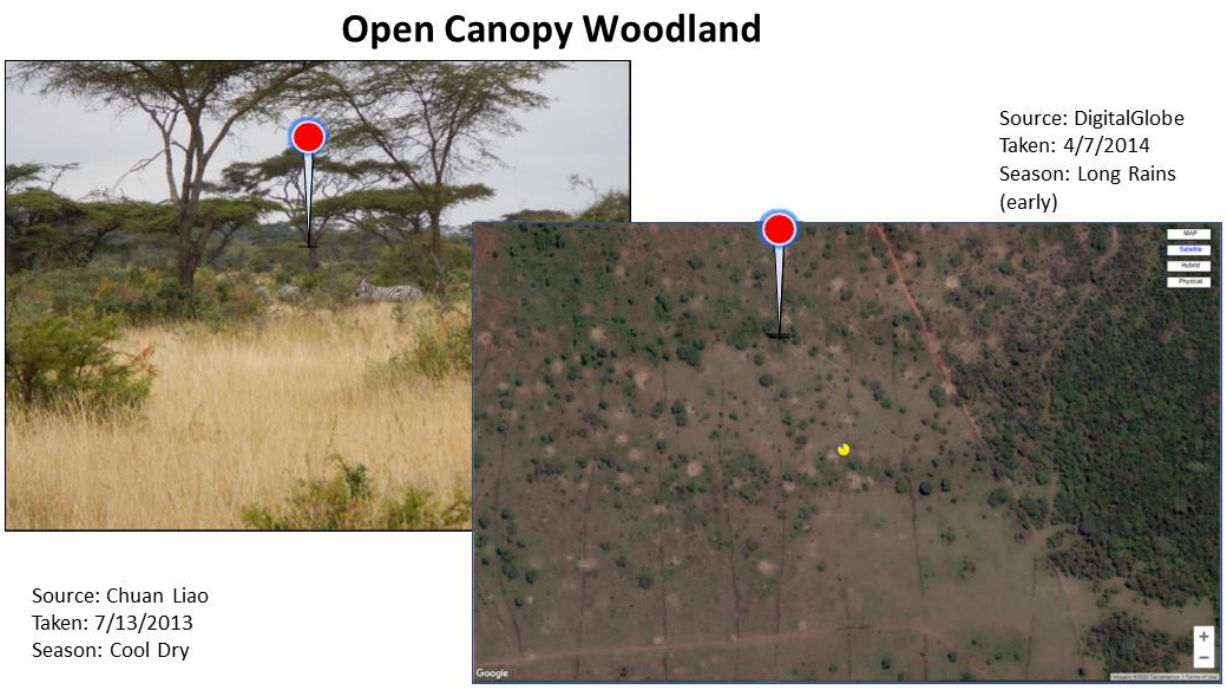

Figure B5. Flashcard for land cover type “Sparse Scrubland”.

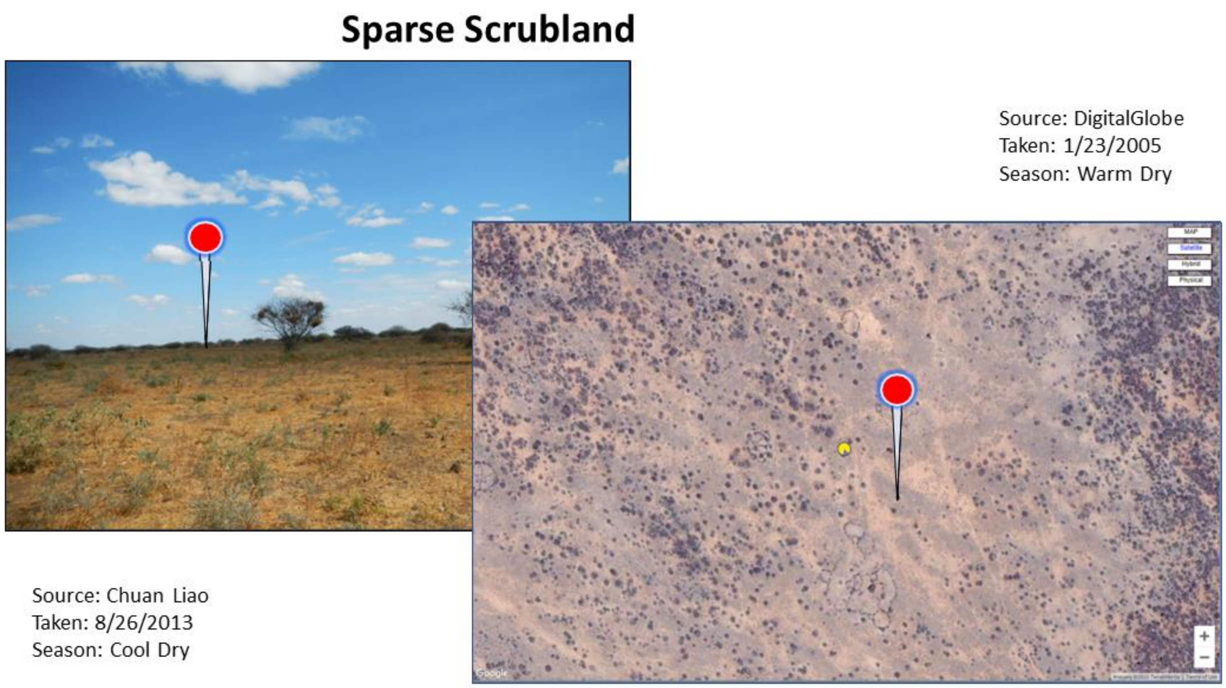

Figure B6. Flashcard for land cover type “Cultivated Land”, maize crop.

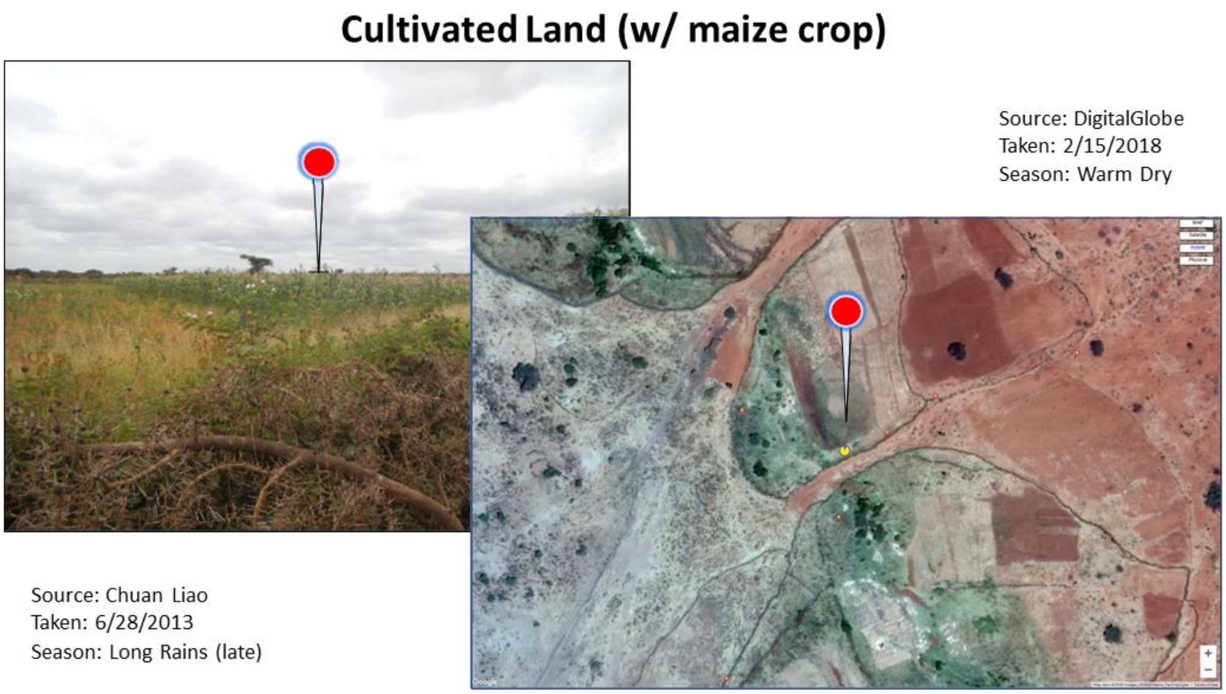

Figure B7. Flashcard for land cover type “Cultivated Land”, cropped versus fallow.

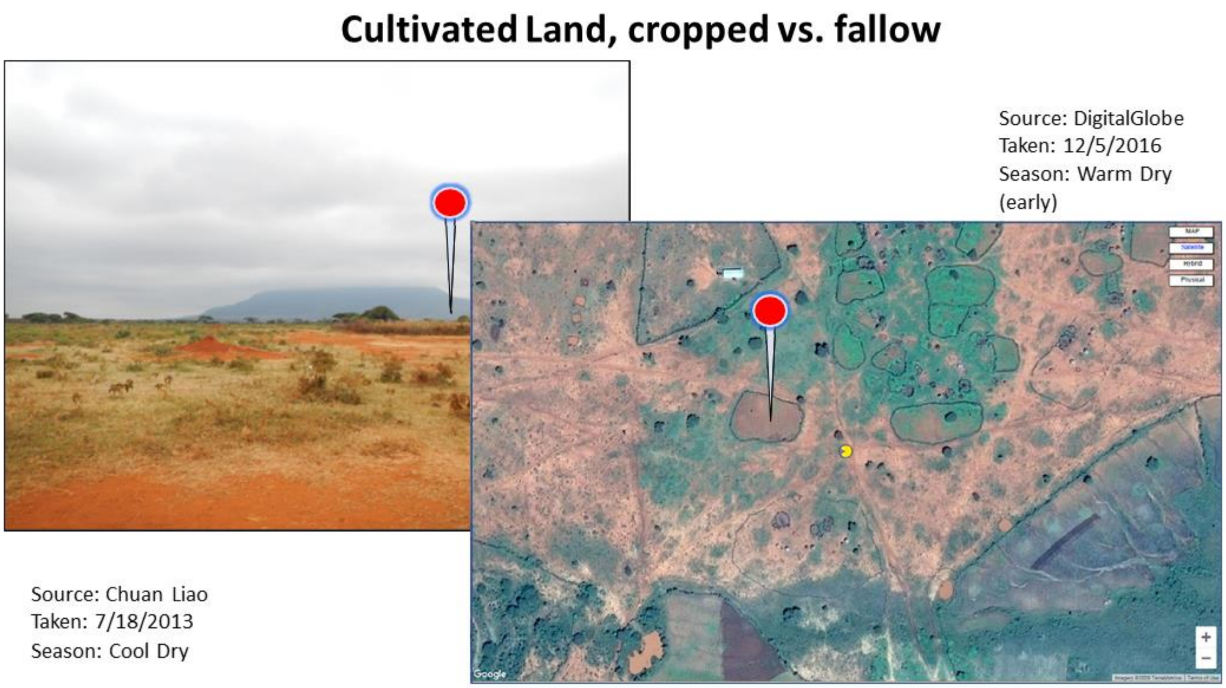

Figure B8. Flashcard for land cover type “Cultivated Land”, teff crop.

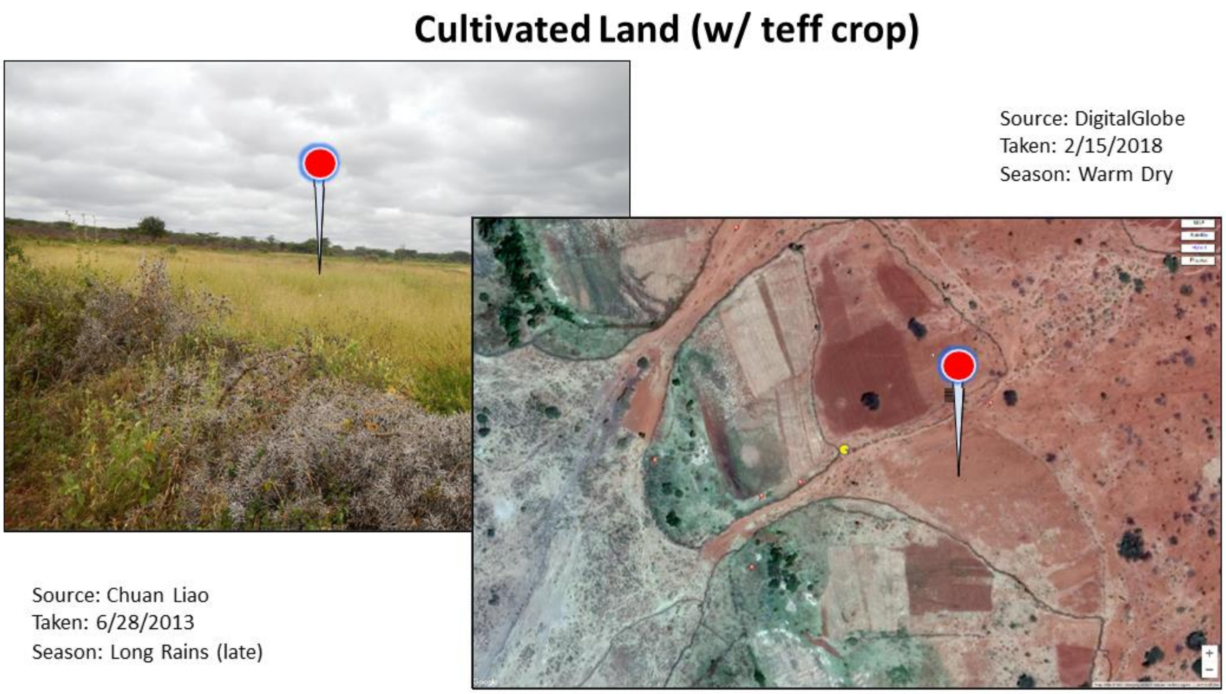

Figure B9. Flashcard for land cover type “Grassland”.

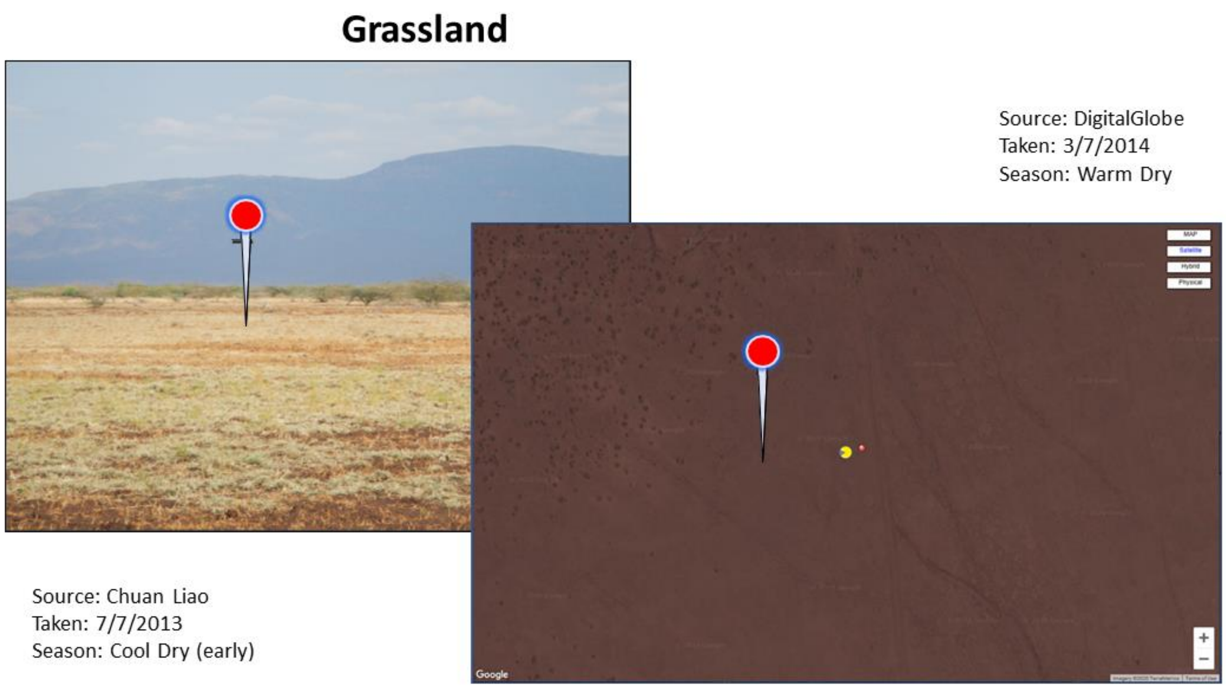

Figure B10. Flashcard for land cover type “Sparsely Vegetated Land”.

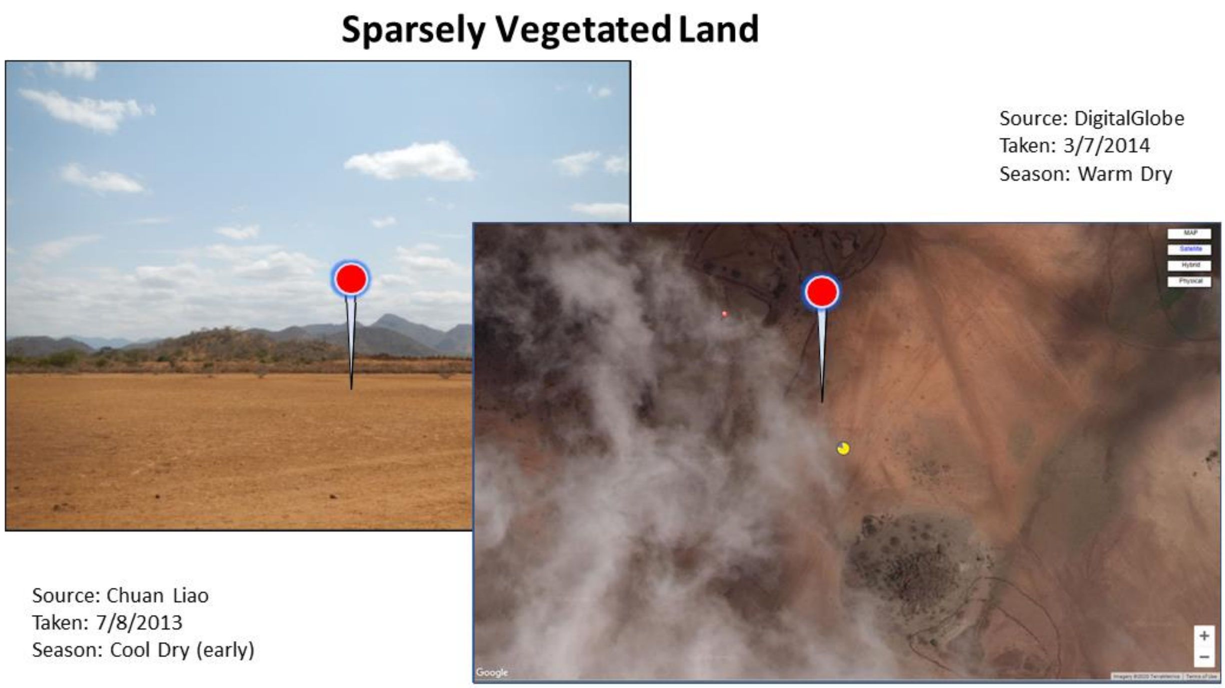

Figure B11. Flashcard for canopy cover level "2m diameter in a 30 by 30 m plot".

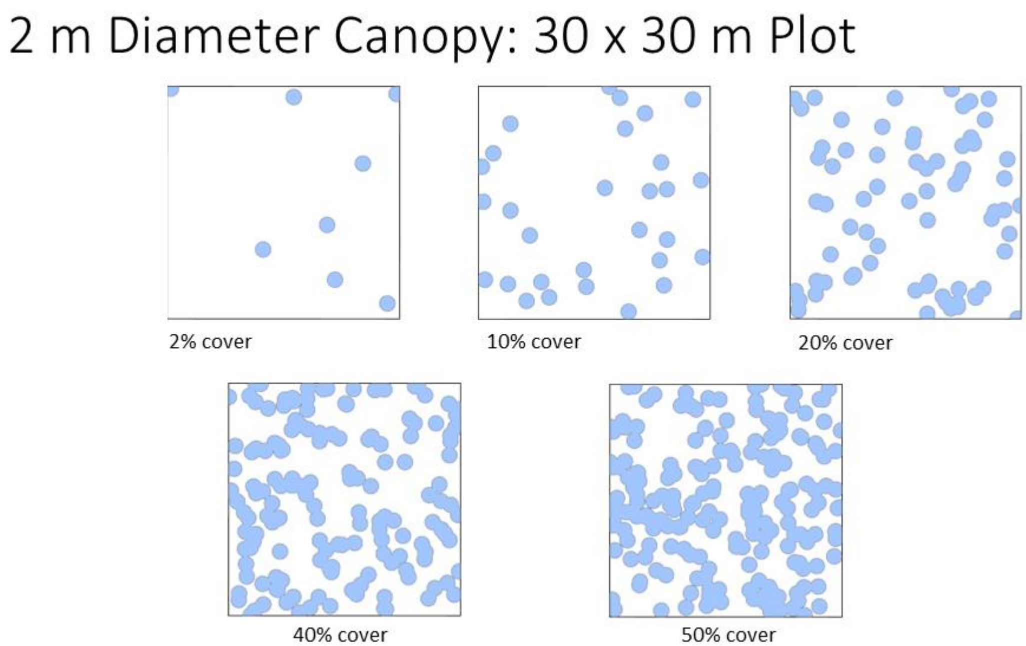

Figure B12. Flashcard for canopy cover level “4m diameter in a 30 by 30 m plot”.

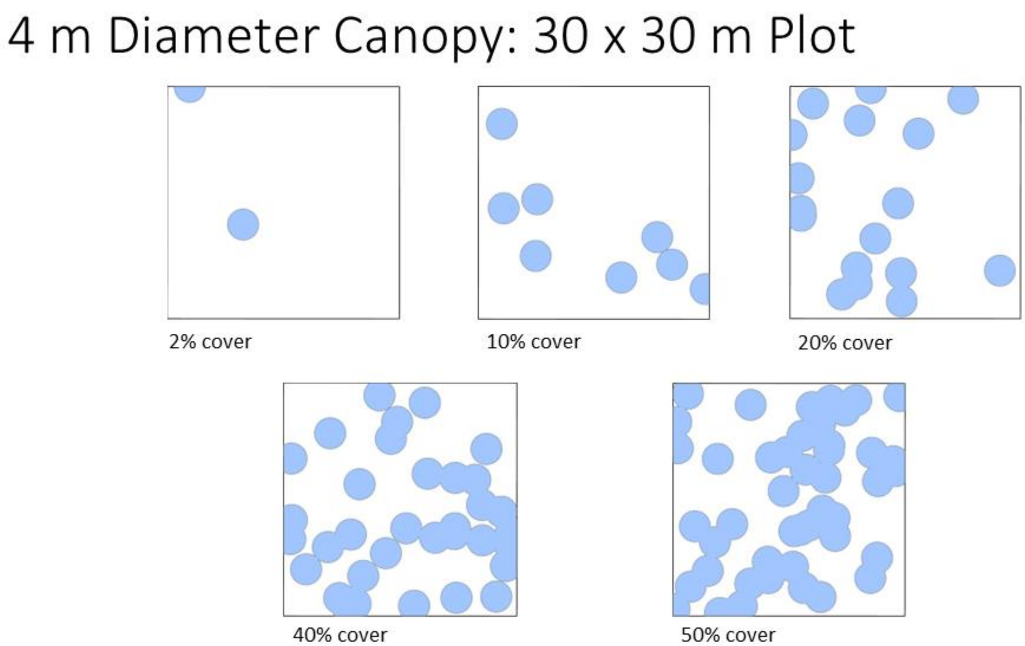

Figure B13. Flashcard for canopy cover level “8m diameter in a 30 by 30 m plot”.

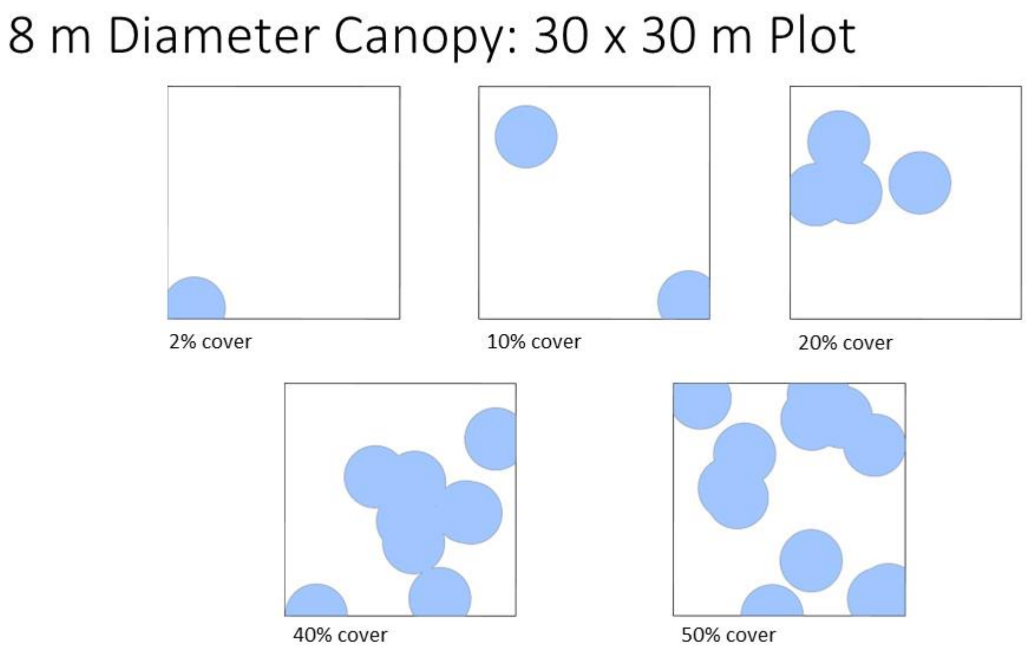

Figure B14. Flashcard for canopy cover level “2m diameter in a 10 by 10 m plot”.

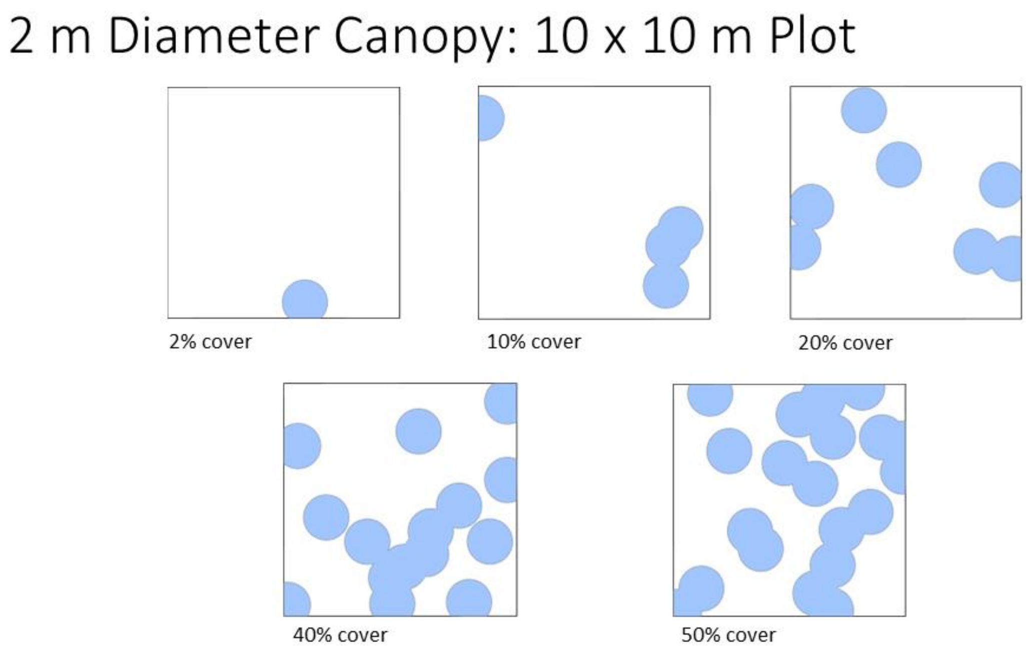

Figure B15. Flashcard for canopy cover level “4m diameter in a 10 by 10 m plot”.

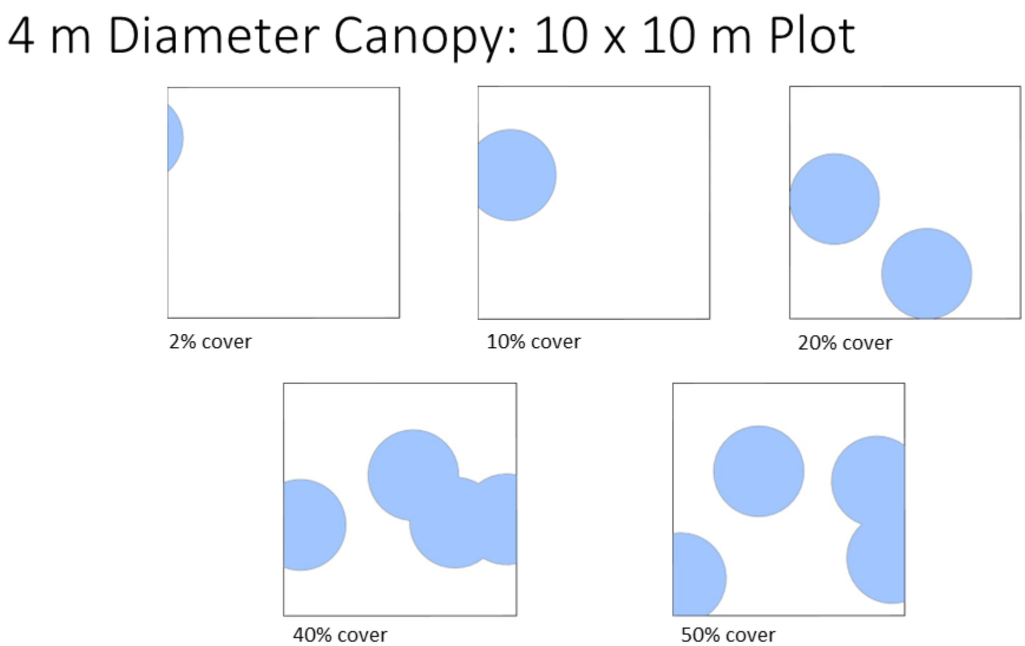

Figure B16. Flashcard for canopy cover level “8m diameter in a 10 by 10 m plot”.