Landscape fragmentation in coffee agroecological subzones in central Kenya: a multiscale remote sensing approach

Gladys Mosomtai[a,b,*], **John Odindi**[b], **Elfatih M. Abdel-Rahman**[a,c], **Régis Babin**[e,f], **Pinard Fabrice**[a,d,f], **Onisimo Mutanga**[b], **Henri E. Z. Tonnang**[a], **Guillaume David**[a,d,f] and **Tobias Landmann**[a,g]

[a]International Centre of Insect Physiology and Ecology, Nairobi, Kenya
[b]University of KwaZulu-Natal, School of Agricultural, Earth, and Environmental Sciences, Pietermaritzburg, South Africa
[c]University of Khartoum, Department of Agronomy, Faculty of Agriculture, Khartoum, North Sudan
[d]Centre de Coopération Internationale en Recherche Agronomique pour le Développement, UPR Bioagresseurs, Nairobi, Kenya
[e]Centre de Coopération Internationale en Recherche Agronomique pour le Développement, UPR Bioagresseurs, Abidjan, Côte d'Ivoire
[f]Bioagresseurs, Univ Montpellier, Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, France
[g]RSS-Remote Sensing Solutions GmBH, Munich, Germany

Abstract. Smallholder agroecological subzones (AEsZs) produce an array of crops occupying large areas throughout Africa but remain largely unmapped. We explored multisource satellite datasets to produce a seamless land-use and land-cover (LULC) and fragmentation dataset for upper midland (UM1 to UM4) AEsZs in central Kenya. Specifically, the utility of PlanetScope, Sentinel 2, and Landsat 8 images for mapping coffee-based landscape were tested using a random forest (RF) classifier. Vegetation indices, texture variables, and wavelength bands from all satellite data were used as inputs in generating four RF models. A LULC baseline map was produced that was further analyzed using FRAGSTAT to generate landscape metrics for each AEsZs. Wavelength bands model from Sentinel 2 had the highest overall accuracy with shortwave near-infrared and green bands as the most important variables. In UM1 and UM2, coffee was the dominant cover type, whereas annual and other perennial crops dominated the landscape in UM3 and UM4. The patch density for coffee was five times higher in UM4 than in UM1. Since Sentinel 2 is freely available, the approach used in our study can be adopted to support land-use planning in smallholder agroecosystems. [DOI: 10.1117/1.JRS.14.044513]

Keywords: agroecosystems; remote sensing; machine learning; *Coffea Arabica*; landscape fragmentation.

Paper 200561 received Jul. 28, 2020; accepted for publication Nov. 5, 2020; published online Dec. 1, 2020.

1 Introduction

The fast-paced conversion of global terrestrial land into croplands, mainly attributed to the growing human population, continuously exerts pressure on flora and fauna and results in habitat loss and disturbance of species communities and their interactions.[1] In Africa, agricultural landscapes typically vary from extensive monocrops with fragments of isolated natural vegetation to mixed crops interspersed with remnants of seminatural vegetation that form a matrix that can impede or facilitate species interactions.[2] Unlike in Europe, where land-use policies have been developed and implemented, Africa's agricultural landscapes remain largely unplanned with no baseline data that can guide sustainable development.[3]

*Address all correspondence to Gladys Mosomtai, gmosomtai@icipe.org

1931-3195/2020/$28.00

Smallholder farmers in Africa practice ~13 general farming systems. Typically, these farms form complex and heterogeneous landscapes with farms commonly small and intercropped, particularly in populated regions.[4] Capturing the structure of these landscapes for ecological applications requires accurate land-use and land-cover (LULC) classifications generated from high-resolution satellite imagery with sufficient reference data. Ecological processes such as spillover edge effects of biodiversity across adjacent LULC types,[5] landscape connectivity effects on species flow[6] or the effects of fragmentation on patch size[7] in sustaining a viable species population can be estimated using LULC information.[8] Thus accurate and up-to-date LULC maps are needed to capture the dynamics and better represent the heterogeneities that characterize specific agroecological setups.

Although global initiatives such as the 300-m GlobCover 2009 and 1-km Global Land Cover 2000 have generated LULC baseline datasets, they remain insufficient in providing accurate maps at regional to subnational scales.[9] LULC legends in these global datasets are generated to estimate global biogeochemical processes such as carbon sequestration, which have limited application to local scale dynamics.[10] Data gaps, inconsistent acquisition periods, cloud cover, and insufficient validation data, especially in the context of Africa, increase the error margins in using the existing global maps.[11] According to Saah et al.,[3] policymakers in developing countries often use outdated maps or opt for global datasets that do not meet their specific needs. Furthermore, their unwillingness to share the existing data across government agencies, creating maps in silos that cannot be harmonized, low-budget allocation, and inadequate human resources slow the creation of useful baseline datasets.

Remote sensing data are the major source of LULC information and the existing satellite datasets have different spatial, spectral, and temporal characteristics with differing cost implications that require users to make trade-offs in their utilization as per their objectives.[12] In the tropics, continuous cloud cover most of the year limits the use of optical satellite datasets; conversely, the cost of using below cloud options such as drones or flight campaigns is too expensive for extensive and wide-area mapping. A critical agroecological zone (AEZ) in the tropical regions is the coffee-based landscapes. Coffee is grown by 25 million smallholders in over 60 countries in the tropics and is a significant source of gross domestic product (GDP) in many developing countries. In Kenya, for instance, coffee is produced by 700,000 smallholder farmers and 3000 large estates, contributing ~230 Million of the GDP annually.[13] Smallholder farmers grow coffee on <2-hectare farms, either as agroforestry systems (i.e., shade coffee) intercropped with mainly subsistence crops, or monocropping systems (i.e., full-sun coffee).[14] Currently, there is no spatially explicit map for these coffee systems because it is often generalized either as croplands (intercropped coffee), shrublands (full-sun), or forest land (agroforest) in many tropical countries.

Existing coffee maps at the global scale are probability distributions generated from ecological niche models generated from climate variables, environmental layers, and presence-only data.[15] These maps limit further analysis of landscape composition and configuration. LULC maps from satellite imageries are the primary baseline data for analysis in landscape ecology. Furthermore, LULC types and their spatial patterns vary according to different landscapes such as AEZ and landforms (e.g., rivers, mountains, cliffs, coasts, and plateaus). They are often generalized in probability distribution maps, yet the subtle dynamics in LULC types influence ecological processes at varying scales.[16]

Herein, we hypothesize that the landscape structure varies across the agroecological subzones (AEsZs) in the coffee-based landscape. Using a random forest (RF) classifier and FRAGSTAT, this study aims to characterize the landscape setup in AEsZs of a coffee-based landscape in central Kenya using an optimal satellite dataset. Both tools provide a unique opportunity to determine landscape patterns in small-scale farming areas. RF classifier[17] can handle nonlinear effects in complex datasets with high accuracy and speed, whereas FRAGSTAT is an efficient tool that has become a reference in landscape ecological studies that involve highly complex agroecological systems.[18]

2 Study Location

This study was conducted in Murang'a County, a major coffee and tea growing region in central Kenya. The county borders Nyeri and Kirinyaga counties in the north, Machakos and Embu counties in the east, and Kiambu county in the south (inset of Fig. 1). Within the coffee growing

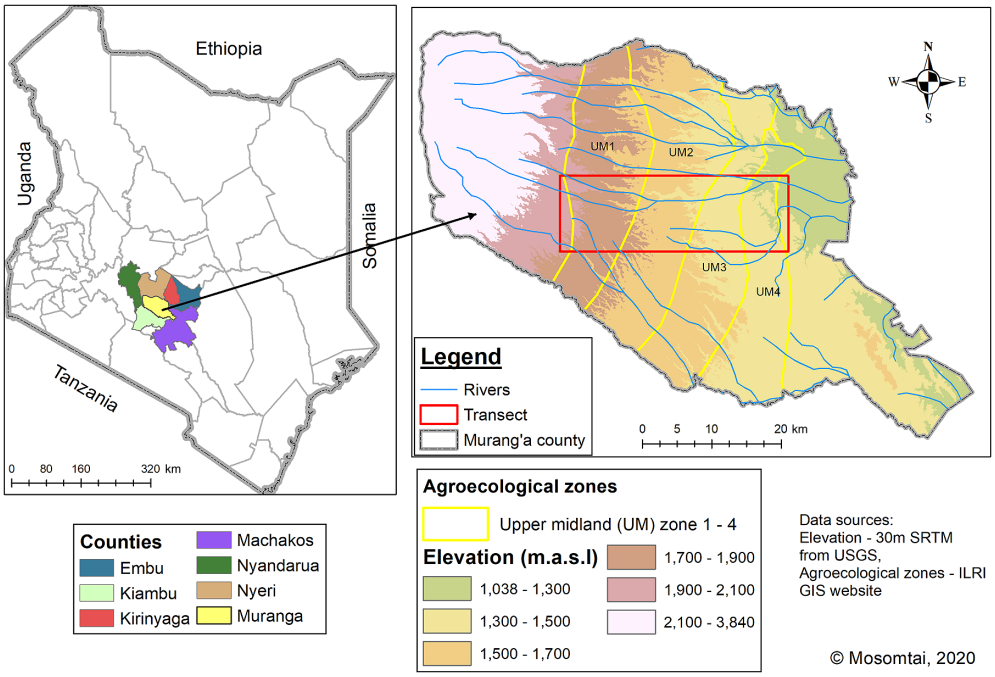

Fig. 1 Map of AEZs of Murang'a county, Kenya, and position of the study transect.

zone, the mean annual temperature and rainfall range from 18°C to 21°C and 1000 to 1500 mm, respectively. The rainfall pattern is bimodal, where long rains occur from March to May, whereas short rains occur from October to December.[19] Consequently, this coincides with the coffee planting, management schedules (e.g., pruning, fertilizer, and pesticides application), and harvesting season of the first and the main crop, respectively.[13] However, the erratic rains and prolonged droughts due to climate variability have resulted in an inconsistency in planting seasons as well as interfered with coffee tree physiology, which makes it more susceptible to low yields and pest and disease infestation.[20]

We selected a transect of 20,240 ha that lies at latitude S 0.8295 deg and S 0.7538 deg and longitude E 36.9472 deg and E 37.1647 deg as a representative of the entire coffee belt (Fig. 1). Our study was conducted in the context of a bigger project that aimed to improve the coffee value chain for smallholder farmers. This informed the choice of the study transect, which covered all the coffee AEsZs (i.e., upper midland; UM1 to UM4). Specifically, coffee grows in four subzones that cut across an elevation gradient of 1300 to 2000 m a.s.l. (above sea level).[19] UM1 is the transition zone for growing tea and coffee. UM2 and UM3 are the primary and marginal coffee-growing zones, respectively. Unlike UM1 to UM3, where coffee is rainfed, at UM4, coffee is grown under irrigation.[19] In the study area, coffee is grown either as full-sun, intercrop, or under shade. Common shade trees include macadamia (*Macadamia integrifolia*), avocado (*Persea americana*), mango (*Mangifera indica*), and hedgerow trees like grevillea (*Grevillea robusta*), whereas intercrops include banana, maize, bean, and sweet potato on an average farm size of 0.5 ha.[21]

The topography of the region is undulating with dissected hills sloping from northwest to southeast.[21] Soils on the hills and minor escarpments are cambisols and rigosols formed on the homogenous basement system of gneiss rocks. In contrast, soils on the plateaus and foot ridges such as nitisols developed on tertiary igneous rocks. Nitisols contain high nutrients from their primary minerals and montmorillonite clay, which make them suitable for coffee and tea plantation.

3 Methodology

Figure 2 presents the datasets and summarizes the methods used in the study. The initial stage involved preprocessing of the satellite images and deriving vegetation indices (VIs) and texture variables. The second stage involved running the RF model to generate LULC maps and the final step was the analysis of the LULC map to generate landscape fragmentation metrics for each AEsZ.

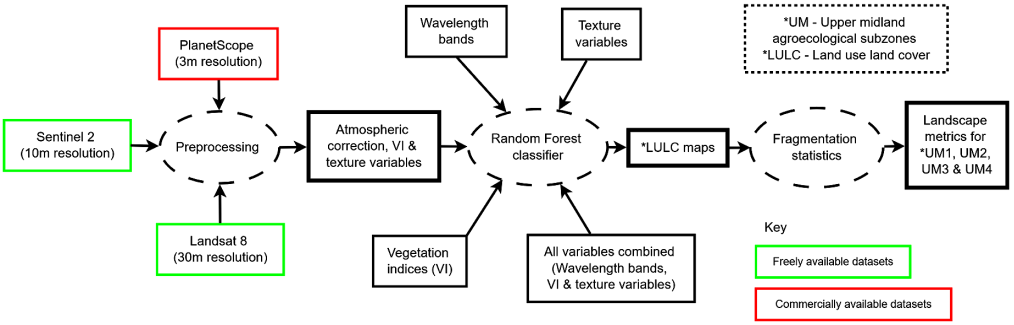

Fig. 2 Flowchart of the datasets and methods used in the study.

3.1 *Datasets*

3.1.1 *Satellite imagery used*

PlanetScope (PS) is a high-resolution satellite dataset from Planet labs that is available commercially (Table 1). In contrast, Sentinel 2 (S2) (high to medium resolution) and Landsat 8 (L8) (medium-resolution dataset) are freely available from the European Space Agency and the United States Geological Survey, respectively. PS is a constellation of 130 CubeSat satellites with a daily revisit time of the entire land surface, which measures the reflected energy within the blue (B), green (G), red (R), and near-infrared (NIR) wavelengths at 3-m spectral resolution.[22] S2 measures a broader range of the electromagnetic spectrum ranging from visible, NIR, and shortwave using the MultiSpectral Instrument sensor at 10-, 20-, and 60-m spatial resolution, with a 5-day revisit time.[23] L8 measures similar wavelengths as S2 at 30-m spatial resolution (Table 1); however, S2 has additional red edge bands, which have proven useful in various vegetation, agriculture, and LULC monitoring studies.[24] Due to cost implications of high-resolution images, PS was used in this study to test the utility of commercial satellites over freely available satellites and the expected trade-offs in the overall accuracy (OA) of LULC mapping.

For this study, we selected good quality images with <2% cloud cover. S2, PS, and L8 images were obtained in August, October, and December 2017, respectively. Table 7 provides detailed descriptions of the image scenes used in this study. The images were already orthorectified to remove the topographic effects. Due to the bimodal rainfall patterns, the study area has two annual growing seasons that overlap.[25] Hence, the landscape has continuous cover crops all year round. In the preprocessing stage, we converted the images into surface reflectance values for further analysis. L8 bands were pansharpened to 15-m resolution using the Brovey transform method in QGIS,[26] whereas S2 bands of 20 m were resampled to 10-m spatial resolution (Fig. 2). Out of 43 VIs, the variance inflation factor was used to select 11 uncorrelated VIs and five biophysical variables presented in Table 2. For texture analysis, we generated 10 indices that represent contrast, statistic, and orderliness features from the NDVI band of each satellite dataset, as outlined in Table 2. VIs have been shown to be sensitive to chemical and morphological aspects of the leaf organs, which are used to estimate water content, plant types, nutrients content, pigmentation, and others.

3.1.2 *Classification reference data*

We used very high-resolution images from Google Earth Pro acquired in July 2017 to obtain reference data for training and testing our classification models. These are high-spatial resolution images (<1 m) obtained from different platforms with acquisition dates indicated, and studies have shown that they can be used to obtain reference data.[37] Although there was no field reference data to distinguish crop types, especially annual crops, prior knowledge of the study area, interpretation of Google Earth Pro image texture, shape, canopy size, and literature on the crop types in the study area were used to generate the LULC classes. We considered the following LULC classes: coffee, tea, other perennial crops (herein referred to as perennials), banana, annual crops, grassland, agroforestry, bareland, settlements, and waterbodies. The perennial

Table 1 Summary of spatial and spectral characteristics of Sentinel 2, Landsat 8, and PlanetScope satellite imagery. NIR = near infrared and SWIF = shortwave infrared

Band	Description	Wavelength range (μm)	Spatial resolution (m)
Sentinel 2			
B2	Blue	0.439 to 0.535	10
B3	Green	0.537 to 0.582	10
B4	Red	0.646 to 0.685	10
B5	Red edge1	0.694 to 0.714	20
B6	Red edge2	0.731 –to 0.749	20
B7	Red edge3	0.768 to 0.796	20
B8	NIR	0.767 to 0.908	10
B8a	Narrow NIR	0.848 to 0.881	20
B11	SWIR1	1.539 to 1.681	20
B12	SWIR2	2.072 to 2.312	20
Landsat 8			
B2	Blue	0.452 to 0.512	30
B3	Green	0.532 to 0.590	30
B4	Red	0.639 to 0.673	30
B5	NIR	0.851 to 0.879	30
B6	SWIR1	1.567 to 1.6511	30
B7	SWIR2	2.107 to 2.294	30
PlanetScope			
B1	Blue	0.455 to 0.515	3
B2	Green	0.500 to 0.590	3
B3	Red	0.590 to 0.670	3
B4	NIR	0.780 to 0.860	3

crops comprised avocado, mango, and macadamia, which are plantations in the lower subzone that also exist in the other subzones often as shade trees, whereas coffee and tea were treated as independent perennial classes. The annual crops are mainly for subsistence and include maize, bean, arrowroot (mostly grown along the rivers), and sweet potato. On the other hand, bareland comprised exposed soils from quarries and unfallowed land, whereas agroforest constituted mainly shade coffee cropping systems and clusters of woodlots. The coffee class consisted of full-sun coffee plots, which were visible in Google Earth Pro image with no shade.

3.2 *Data Analysis*

3.2.1 *Random forest classification algorithm*

We used the RF classifier by Breiman (2001) to assess the robustness of the satellite datasets for classifying the coffee-based landscape in the study transect (Fig. 1). RF is a collection of decision trees, i.e., classification and regression trees that learn the characteristics of the training samples

Table 2 Summary of VIs, biophysical, and texture variables used in the study

Variable	Description	References
Vegetation index		
RI	Redness index	27
NDPI	Normalized difference pond index	28
MSAVI	Modified soil adjusted vegetation index	29
GEMI	Global environmental monitoring index	30
BI2	Second brightness index	31
BI	Brightness index	31
MTCI	Modified chlorophyll absorption ratio index	32
S2REP	Sentinel-2 red-edge position index	33
GNDVI	Green normalized difference vegetation index	30
REIP	Red-edge inflection point index	34
MCARI	Meris terrestrial chlorophyll index	30
Biophysical		
LAI	Leaf area index	35
LAI_CW	Canopy water content	
LAI_CAB	Chlorophyll content in the leaf	
FCOVER	Fraction of vegetation cover	
FAPAR	Fraction of absorbed photosynthetically active radiation	
Texture		
Contrast features	Contrast	36
	Dissimilarity	
	Homogeneity	
Statistics features	Gray level co-occurrence matrix (GLCM) variance	
	GLCM mean	
	GLCM correlation	
Orderliness features	Maximum probability (MAX)	
	Entropy	
	Energy	
	Angular second moment (ASM)	

and predict similar characteristics in an unclassified dataset.[38] Compared to other machine learning algorithms such as support vector machine, artificial neural network, and boosted regression trees, RF was found to produce robust mapping results in fragmented small scale farming areas in Africa than other methods.[24] RF can handle nonlinear effects in complex datasets or few and imbalanced training samples with high accuracy and speed better than most other algorithms. Furthermore, the algorithm ranks the essential predictor variables negating users' selection errors and subjectivity.[38,39] The algorithm splits the training samples into twofold, approximately

two-thirds for training the model, also known as in-bag samples, and one-third for testing the accuracy of the model, also known as out-of-bag (OOB) samples.[17] The algorithm internally assesses the accuracy of the model based on the OOB error, which averages the error frequency of the decision trees built using in-bag samples. The OOB error is also used in ranking variable importance based on mean decrease accuracy (MDA).[40] Apart from MDA, RF also uses mean decrease gini (MDG) to assign variable importance based on decrease in node impurity of a variable at split node.[41]

In this study, we used 70% of the classification reference data to train the RF model. We used four sets of variables for each satellite dataset to build the models: (i) wavelength bands only, (ii) vegetation indices only, (iii) texture variables only, and (iv) combined wavelength bands with VI and texture variables (Fig. 2). Both MDA and MDG were used to determine the important variables for LULC classification. We implemented the model using the "randomForest" package[42] in R software.[43] The default RF settings, which have been proven to be optimal for building accurate models, were used.[24] The remaining 30% of the reference data were used for model evaluation. We generated classification confusion matrices and calculated the following accuracy assessment metrics; OA, user's accuracy (UA), producer's accuracy (PA), $F1$ score, and kappa coefficient (K).[24] $F1$ score is a mean metric of precision (PA) and recall (UA), where a value of 100% indicates that the model achieved perfect precision and recall of all the test data and the inverse is true.[44] The equations for calculating the accuracy metrics are as follows:

$$\mathrm{OA} = (1/N)\sum_{j=1}^{r} n_j,$$

$$\mathrm{PA} = \frac{n_j}{n_{i\mathrm{col}}},$$

$$\mathrm{UA} = \frac{n_j}{n_{i\mathrm{row}}},$$

$$F1\ \mathrm{score} = 2 \times \frac{\mathrm{PA} \times \mathrm{UA}}{\mathrm{PA} + \mathrm{UA}},$$

where N is the total number of samples in the image, r is the number of rows, n_j are the samples that are correctly classified, and $n_{i\mathrm{col}}$ and $n_{i\mathrm{row}}$ are the total number of samples in columns and rows, respectively.[45]

3.2.2 *Landscape metrics*

LULC map with the highest accuracy was further analyzed using the FRAGSTAT tool[46] to quantify the composition of LULC types and landscape fragmentation across the four AEsZ in the study transect (i.e., UM1, UM2, UM3, and UM4). FRAGSTAT computes several metrics measured at patch, class, and landscape levels that describe, among others, area, edge, shape, contagion, contrast, and aggregation from LULC maps. In this study, we assessed the mean patch, largest patch index (LPI), patch density (PD), splitting index (SPLIT), contagion, and landscape percentage occupied by each LULC type across the four AEsZs at class and landscape levels. A patch defines a homogenous area that is different from its surroundings, herein referred to as the LULC type. Patches are computed based on the pixel size of the satellite image used to map the LULC; hence they are subjective to scale variability or specification of minimum patch size.[46] Furthermore, patch size holds ecological significance compared to all metrics. It is shown that patch sizes can influence species richness in semiforested coffee systems,[47] bird species abundance in naturally heterogeneous landscapes,[48] and insect pollinators in forest fragments in shaded coffee agrosystems.[49] Table 3 summarizes the description of the FRAGSTAT metrics used in this study.

Table 3 A description of class and landscape metrics used in the present study as defined by the FRAGSTAT user manual.[46]

Level	Metric	Description	Unit
Class	Mean patch area (MPA)	The average-weighted mean of the number of patches in the class and total class area	ha
	LPI	The largest patch of the corresponding patch type divided by total landscape area	Percent
	Percentange of landscape (PLAND)	Proportional abundance of each patch type in the landscape	Percent
	PD	Number of patches in the landscape, divided by total landscape area	Number of patches/100 ha
Landscape	SPLIT	Number of patches with a constant patch size when the landscape is subdivided into equal sizes	None
	Contagion (CONTAG)	A measure of both intermixing of patch types and spatial distribution of a patch type	Percent
	Euclidean nearest neighbor distance (ENN)	Shortest straight-line distance between the focal patch and its nearest neighbor of the same class	Meter

4 Results

4.1 *RF Model Accuracy Assessments*

Table 4 shows the OAs obtained using the RF classifier and the various satellite image datasets. Wavelength bands were better predictors of LULC in all the satellite datasets compared to using either VI or texture variables. For bands only, S2 had the highest OA (95%) compared to L8 (90%) and PS (83%). Combined with VI and texture variables, the OA for PS increased considerably compared to using only wavelength bands (by 3%) or texture variables (by

Table 4 The OA, OOB error, and kappa coefficient (K) for LULC maps of coffee-based landscape in Murang'a, Kenya, using different satellite datasets and the RF classification algorithm.

Model	OA (%)	OOB error (%)	Kappa (K) (%)
PS bands	83	19	80
PS vegetation indices	81	20	77
PS texture variables	60	39	51
PS bands, VI, and texture	86	13	83
S2 bands	95	5	93
S2 vegetation indices	91	9	89
S2 texture variables	79	23	75
S2 bands, VI, and texture	96	4	95
L8 bands	90	9	88
L8 vegetation indices	88	12	86
L8 texture variables	79	25	74
L8 bands, VI, and texture	86	17	84

26%). For S2, there was no substantial change when using combined variables in comparison to wavelength bands only (OA increased by 1%). For L8, the OA dropped by 4% when using combined variables compared to the wavelength bands only. VIs were the second best predictors in all the satellite datasets, whereas texture variables were the least. VIs from S2 had the highest OA (91%), whereas VIs from PS had the least (81%). Similar to the OA, the OOB error estimate showed that the texture variables from PS had the highest error rate (39%) in all the models, indicating its poor predictive ability of the internal OOB samples.

S2 variables showed better mapping results for all the classes with more than 90% PA, UA, and $F1$ scores than L8 and PS datasets (Table 5). The banana class was generally poorly mapped in all the satellite datasets (the least PA = 41% using PS, highest PA = 83% using S2), whereas the waterbodies had the highest accuracies except in L8, which had the least PA of 38%. Coffee and agroforest classes had the highest accuracies ($F1$ score = 94% and 97%, respectively) when mapped using S2 datasets. However, there was an increase in efficiency for PS (by 6% for coffee and 4% for agroforest) when the wavelength bands were combined with VI and texture variables.

4.2 *Variable Importance*

In the wavelength bands model, SWIR1, SWIR2, and green bands from S2 and L8 were the most important variables contributing to a total of 36% and 53% MDA and 44% and 57% MDG, respectively (Fig. 3). Additionally, the NIR bands from S2 and L8 contributed 10% and 18% MDA to the model accuracy, respectively. The red-edge band from S2 and the red band from L8 contributed 13% and 16% MDG in decreasing node impurity, respectively. For VI and texture variable models, BI, NDPI, RI, GNDVI, and additional LAI_CW and CAB from S2 were the most important variables. At the same time, contrast, GLCM mean, variance, and correlation were the most important texture variables in all the satellite datasets (Table 8) for mapping our landscape classes. When all the variables were combined, the same important variables identified in the individual models contributed more to the model (Table 9).

4.3 *Landscape Fragmentation in each AEsZ*

Figure 4 shows the LULC maps from the PS wavelength bands (map A), S2 (map C), and L8 (map B). Visually, PS and S2, unlike L8, mapped similar landscape structures, but with varying levels of accuracy. L8 overestimated annual crops and grassland at the expense of coffee, whereas agroforest appears in larger patches than in PS and S2. Across the AEsZ, the primary coffee-growing zone is at UM2 and the lower region of UM1. At UM3, coffee is interspersed with annual crops, which form the matrix of the landscape, whereas grasslands and other perennials dominate UM4. Pockets of agroforests are evenly distributed in UM1 and UM2, but in UM3, patches take a more linear shape. In UM2, annual crops are grown along the riverine, whereas settlements appear in linear patches with one major town situated at UM4. L8 mapped the extent of the town in the transect poorly compared to S2 and PS. We adopted the S2 map for further analysis of the landscape structure given its high accuracy, which captured the landscape physiognomy at 10-m spatial resolution better than PS, which had the highest spatial resolution of all the datasets used at 3 m.

4.3.1 *Class level metrics*

According to the S2 LULC map [Fig. 4(c)], coffee covers 64% and 60% of the total landscape in UM1 and UM2, respectively, whereas the annual crops occupy 43% in UM3 (PLAND in Table 6). Other perennials and annual crops occupy 29% and 22% of UM4, respectively. All the dominant patch types that formed the matrix in each AEsZ had the LPI as shown in Table 6. Coffee patches in UM2 were more fragmented than in UM1. The most dominant patch (given by LPI) in UM1 is almost twice the size of the dominant patch in UM2 (UM1 = 63 and UM2 = 33). Similarly, the MPA for coffee in UM1 is more than three times the size in UM2 to UM4 (UM1 = 2.65 ha, UM2 = 0.74 ha, UM2 = 0.16 ha, and UM4 = 0.14 ha). Agroforest patches occupy 14% of UM1 and 17% of UM2 and UM3, with the least cover in UM4, occupying ~5% of the landscape. Coffee intercropped with banana is more prevalent in UM2 and UM3,

Table 5 PA, UA, and *F*1 score for all the classification models used in the study to map the LULC classes of the study transect.

Model	(%)	Annuals	Banana	Bareland	Coffee	Agroforest	Grassland	Perennials	Settlement	Tea	Waterbody
PS bands	UA	76	69	91	80	85	97	74	90	95	100
	PA	89	41	91	81	85	96	53	92	91	100
	F1	82	52	91	80	85	97	62	91	93	100
PS bands, VI, and texture	UA	79	76	90	84	88	97	75	95	95	100
	PA	89	46	90	87	91	97	47	90	91	100
	F1	84	57	90	86	89	97	57	93	93	100
S2 bands	UA	91	92	92	91	99	98	97	97	99	100
	PA	96	83	95	95	94	100	92	92	94	100
	F1	93	87	94	93	96	99	94	94	96	100
S2 bands, VI, and texture	UA	95	94	95	93	97	99	97	97	100	100
	PA	99	74	95	96	96	99	95	96	92	100
	F1	97	83	95	94	97	99	96	96	96	100
L8 bands	UA	91	88	92	87	92	92	90	91	93	100
	PA	96	75	89	91	93	84	68	100	95	38
	F1	94	81	91	89	93	88	78	95	94	55
L8 bands, VI, and texture	UA	88	89	73	86	90	91	93	71	95	90
	PA	98	56	81	94	78	87	98	74	92	94
	F1	93	69	77	90	83	89	95	72	94	92

Note: PS, PlanetScope; S2, Sentinel 2; and L8, Landsat 8.

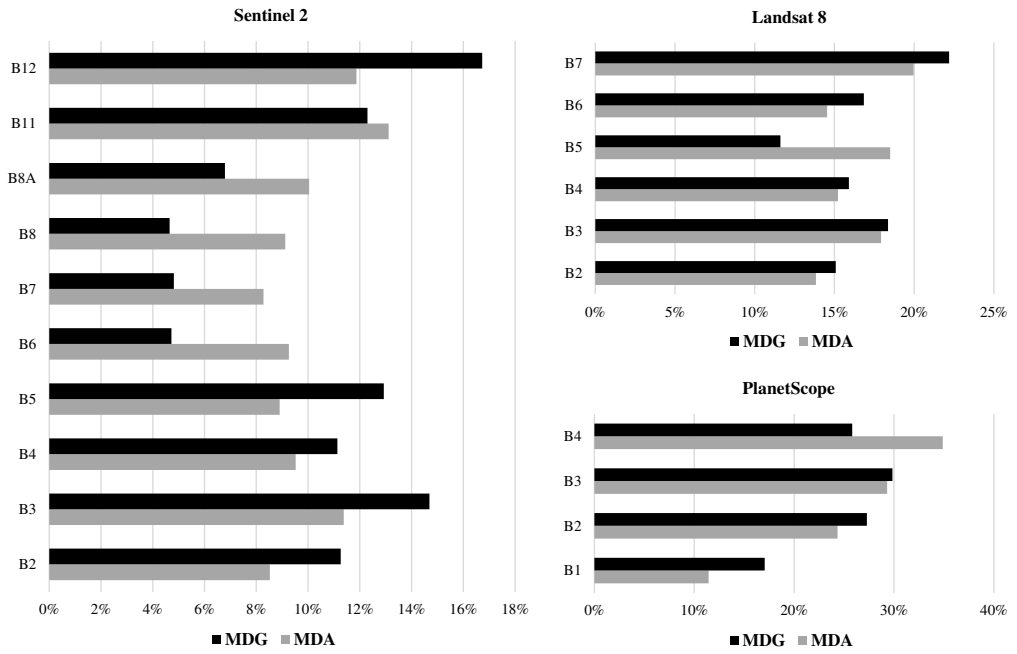

Fig. 3 Variable importance of wavelength bands for Sentinel 2, Landsat 8, and PS datasets converted into percentage.

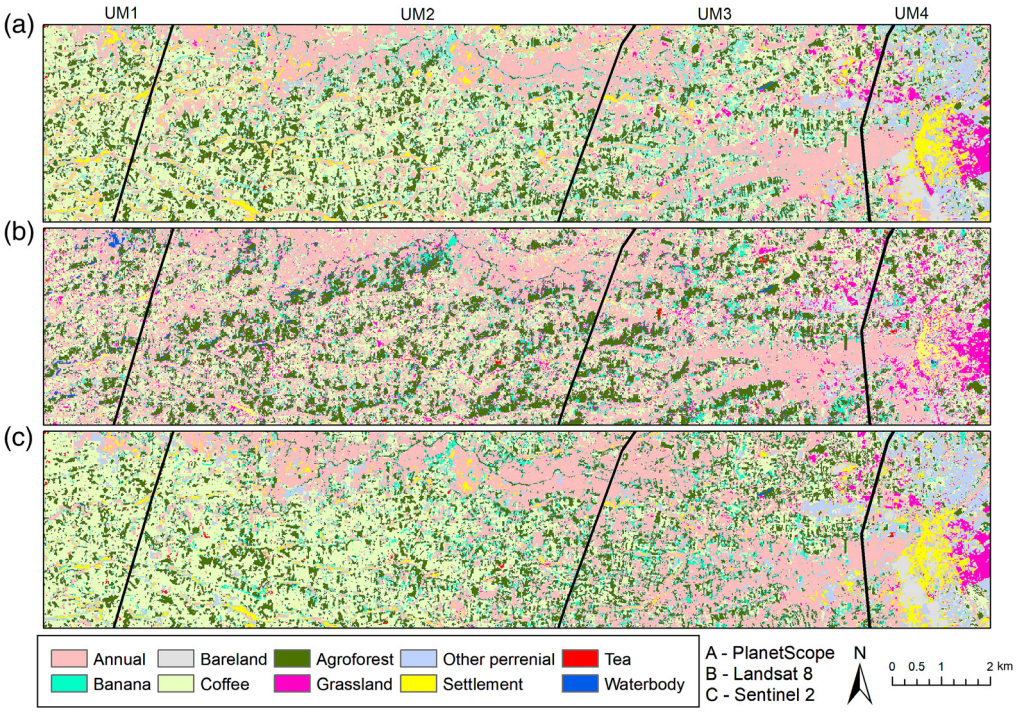

Fig. 4 LULC maps of coffee growing transect in Murang'a county produced using (a) PS, (b) Landsat 8, and (c) Sentinel 2 datasets, and the RF classifier. The black lines show the boundaries of the UM zones.

where banana patches have <0.06% of LPI. UM4 is the marginal coffee growing area characterized by more settlements, grassland, and bareland than all the other subzones. Coffee is highly fragmented in UM3 and UM4 as compared to UM1, with PDs of 157 and 120, respectively (Table 6). Additionally, forest cover and bananas have a high PD in UM2 and UM3, while for annuals, shrubs, grassland, and settlements, the PD is high in UM4.

Table 6 PLAND (%), LPI (%), PD (number of patches/100 ha), and MPA (ha) for each LULC class across the four AEsZs of the study transect in Murang'a County, Kenya

	Metric	Annuals	Banana	Bareland	Coffee	Agroforest	Grassland	Perennials	Settlement	Tea	Waterbody
UM1	MPA (ha)	0.25	0.05	0.37	0.14	0.08	0.24	0.4	0.13	0.03	0.16
	PLAND (%)	3.77	2.40	0.03	64.25	14.31	0.10	4.62	2.98	7.54	0.01
	LPI (%)	0.87	0.02	0.01	62.97	0.20	0.01	0.22	0.23	0.33	0.00
	PD	34.24	81.84	0.46	24.27	68.81	2.98	32.71	32.90	56.12	0.19
	MPA (ha)	0.11	0.03	0.07	2.65	0.21	0.03	0.14	0.09	0.13	0.03
UM2	PLAND (%)	18.83	5.65	0.05	50.99	17.35	0.17	3.64	2.95	0.35	0.01
	LPI (%)	5.36	0.04	0.02	33.20	0.17	0.01	0.08	0.11	0.01	0.00
	PD	53.41	145.22	0.36	69.05	114.08	5.51	27.09	36.81	10.10	0.24
	MPA (ha)	0.35	0.04	0.15	0.74	0.15	0.03	0.13	0.08	0.03	0.04
UM3	PLAND (%)	43.02	5.83	0.39	25.30	17.86	1.78	3.36	2.37	0.04	0.04
	LPI (%)	23.32	0.03	0.09	3.06	0.20	0.12	0.51	0.09	0.00	0.02
	PD	63.22	118.81	1.27	157.37	132.84	24.18	24.55	30.28	1.95	0.53
	MPA (ha)	0.68	0.05	0.31	0.16	0.13	0.07	0.14	0.08	0.02	0.08
UM4	PLAND (%)	22.84	1.37	5.17	17.26	5.35	10.75	29.30	7.76	0.18	0.02
	LPI (%)	5.07	0.05	0.89	1.42	0.17	1.56	10.10	2.00	0.02	0.01
	PD	93.06	26.45	13.86	120.62	67.11	44.71	73.42	58.83	6.89	0.10

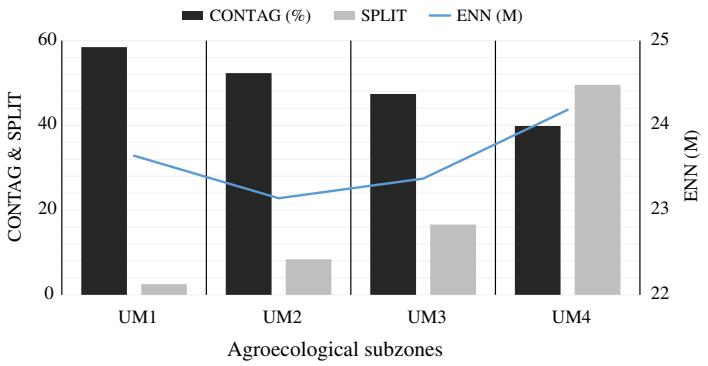

Fig. 5 Landscape metrics that describe landscape connectivity [CONTAG index (%)], landscape fragmentation (SPLIT index), and patch isolation [ENN (m)] in UM1, UM2, UM3, and UM4.

4.3.2 *Landscape level metrics*

Distinct landscape structures exist in each AEsZ. Landscape connectedness, which is measured by the CONTAG (contagion) index is highest in UM1 and gradually declines along the altitudinal gradient, with the least connectedness in UM4 (Fig. 5).

Furthermore, landscape fragmentation described by the SPLIT index shows that UM4 is fragmented five times more than UM1. The high CONTAG index in UM1 also describes a landscape with low diversity in cover types. Interestingly, the distance between patches is similar across all the AEsZ (ENN range 23 to 24 m), which means that UM4 has more cover types that are well interspersed across the landscape, UM1 has a similar interspersion but with the same or fewer patch types than UM4.

5 Discussion

This study explored three multisource satellite images from 3-m PS, 10-m Sentinel 2 (S2), and 30-m Landsat 8 (L8) to identify the optimum dataset for mapping LULC of a coffee-based landscape in the highlands of East Africa that are dominated by smallholding farming. Furthermore, we examined the landscape composition and the level of fragmentation on each AEsZ. The results showed that the spectral resolution of a sensor is a critical factor in delineating vegetation types in a heterogeneous agricultural landscape and that each AEsZ has a unique landscape physiognomy. This study fills a gap on the scarcity of detailed LULC maps, especially in Africa, since the available maps in public databases often generalize agricultural landscapes and are not up-to-date. We delineated detail LULC types that govern ecological processes in each AEsZ and outlined its potential in improving production in agricultural landscapes (e.g., coffee production) while conserving biodiversity and providing ecosystem services.

5.1 *Model Accuracy Assessment*

Our results showed that S2 datasets had the highest OA in all the models, followed by L8, whereas PS had the least OA, despite having the highest spatial resolution. Htitiou et al.[40] and Shoko and Mutanga[50] reported similar findings when S2 dataset outperformed other multispectral datasets in mapping crops and grasslands. The different accuracies in our mapping results could be associated with the differences in the spectral resolution (bandwidth and number of bands) among our sensor images.[51] S2 measures a broader region of the electromagnetic spectrum ranging from the red, green and blue to the SWIR region with additional red-edge bands, in contrast with PS, which covers only the visible and NIR regions with lower waveband data bits. Furthermore, the pixel depth (radiometric resolution) of S2 allows the bands to capture more details per pixel. This could explain the observed high accuracy of the wavelength bands with minor improvement when VI bands were added. Due to advancement of S2 sensor specifications, subtle differences in cover types can now be captured with the added advantage of shorter

revisit times compared to L8. Additionally, the higher resolution in PS captures more features, which increases spectral confusion between classes with the narrow spectral bands.

The commission and omission errors of LULC types described by UA and PA, respectively, were relatively lower when using S2 wavelengths bands. This reinforces our finding that S2 imagery is more suitable for discriminating LULC types in heterogeneous and complex landscapes. In our study area, where coffee and agroforest are important cover types and are often difficult to differentiate, coffee and agroforest were mapped with a higher PA using S2 when compared to L8 and PS. Previous studies, however, attempted to map coffee from other LULC types using Landsat datasets; their results were comparable to what we obtained using L8. For example, Ortega-Huerta et al.[52] differentiated between the open canopy and closed canopy coffee in Southwest El Salvador with an OA of 85.7% using Landsat Thematic Mapper while Cordero-Sancho and Sader[53] attempted to classify shade coffee and sun coffee in Costa Rica with PAs of 91.8% and 86.2%, respectively, using Landsat Enhanced Thematic Mapper. Both S2 and L8 are freely available datasets, but our results showed more accurate mapping results from S2 as opposed to L8. This study, therefore, elucidates the potential of using S2 with limited resources to generate detailed LULC maps, especially for Africa, which is often missed in global LULC datasets. Furthermore, the temporal resolution of the S2 sensor of five days means that researchers and other stakeholders can have access to up-to-date maps to inform their policies.

5.2 *Variable Importance*

Surprisingly, SWIR 1 and SWIR 2 bands were the most important variables in S2 and L8 as opposed to NIR, red edge, red, or green bands, which are known to be the most important bands in LULC classification of vegetation types. In coffee-based landscapes, SWIR bands have been shown to be particularly important due to the soil background that is mixed with the spectral signal of coffee leaves.[54] Often, coffee trees are planted with spacing in between the rows. Though these rows were not visible in our study due to smallholdings with low coffee density and possible defoliation of coffee leaves due to leaf rust disease, they still influenced the pixel purity. Notably also from our study the significance of the red edge band in increasing class separation (MDG = 13%). Red edge bands measure the abrupt rise in the reflectance within the transition zone of the red and NIR regions, this region of the electromagnetic spectrum detects subtle variability in vegetation types, which would otherwise be generalized when using broadband widths such as the case of L8.[32]

The VIs (including biophysical variables from S2) models had a lower OA compared to only wavelength bands, whereas the texture variables had the least OA in all the satellite datasets. The most significant VIs included BI, RI, NDPI, and LAI_CW. On the other hand, contrast, correlation, mean, variance, and to a lesser extent, homogeneity was the most essential texture variables. When VIs and texture variables were combined with the wavelength bands, there was no considerable change in OA for S2, but for PS, there was an increase in OA; we associate this to the unique information that RI contributed to the model. RI and BI are soil-based indices that measure the color properties of soil.[55] These two indices further explain the particular soil background characteristics that were captured by the SWIR band in the wavelength bands model. Given the limited spectrum and data depth of PS, RI captured similar information in the SWIR band; hence, the utility of VIs, especially in sensors with limited spectral bands, is vital. Despite the significance of VIs, the low OA observed in the VIs models is associated with the oversimplification of VIs, especially in heterogeneous landscapes where there is more than one vegetation type and species that could co-occur and occupy the same pixel.[30]

5.3 *Landscape Fragmentation in Each AEsZ*

The results further revealed that landscape composition varied according to AEsZs, which correspond to various elevation zones (Fig. 4). Coffee is the dominant cover in UM1 and UM2 (elevation ranging 1900 to 1600 m), whereas in UM3 and UM4 (elevation ranging 1500 to 1300 m), the dominant cover types are annual crops and other perennials, respectively. The agroforest system in the study area is highly fragmented in UM3 (PD = 157) with the least cover in UM4. Visual interpretation of the LULC map showed that the landscape physiognomy of

agroforest cover in UM3 and UM4 is mostly hedgerows; however, in UM1 and UM2, it is an intersperse cluster in a matrix of full-sun coffee. Notably, bananas occupied a significant percentage of the landscape in UM2 and UM3 (5% of the landscape). Intercropping coffee with banana is commonly practiced by many smallholder farmers to complement their food crop and income generation in many coffee-based landscapes in East Africa.[56]

We showed that landscape connectivity is higher in UM1 than in UM4 (CONTAG index in Fig. 5). LULC of UM1 can facilitate the flow of species from one patch to the other, which increases their survival capacity, unlike in UM4, which has more fragmented patches, as shown by the SPLIT index.[57] For instance, the contiguous patches of coffee farms that form the matrix of UM1 and UM2 can facilitate the flow of coffee pests that solely depend on coffee trees as their primary hosts.[32] For example, movements of the coffee berry borer, *Hypothenemus hampei* Ferrari, may be limited where coffee farms are pockets of fragmented patches, such as in UM3 and UM4.[8] Among other factors, such as higher temperatures in UM3 and UM4,[58] fragmentation could potentially result in overutilization of the available patches by such coffee pests, leading to increased severity in infestation levels. With similar consequences, fragmentation may strengthen pest life history traits involved in adaptation to changing environments, leading to a greater chance of survival.[59] More pockets of agroforest cover, as observed in UM1 and UM2 when compared to UM3 and UM4 can benefit from biodiversity conservation and ecosystem services, such as providing habitats to birds, other pest natural enemies, and pollinators. Ecosystem services benefit coffee production by improving yields through microclimate and soil quality improvement and pest and disease natural regulation.[60]

5.4 *Study Implications and Limitations*

In this study, we showed the robustness of RF and S2 in capturing subtle changes within such landscapes. This methodology can be adopted in other coffee growing regions in East Africa, Asia, and South America. The generated LULC maps can be used as baseline data to guide the restoration of degraded landscapes, development of land use policies such as the agri-environment scheme adopted in Europe, and model ecosystem services, especially from shade coffee generating integrated land management systems. The LULC maps developed in this study could also be integrated with crop phenological and climatic variables to understand the occurrence, abundance, and spread of coffee pests and diseases.

A potential limitation to this study was the use of reference datasets obtained from very high-resolution Google Earth Pro (GE) images in lieu of field reference data. This limitation is also a growing opportunity for using GE alongside crowdsourced data from mobile apps,[61] Global Biodiversity Information Facility[62] and Open Street Map[63] to provide reference data for classification, especially in the era of big data.[44] Landmann et al.[64] mapped rain-fed and irrigated lands in Zimbabwe using reference data obtained from GE, which showed the effectiveness of these new data sources for validation of LULC classification in data-scarce environments. Furthermore, our study area (i.e., the transect) was somewhat small, but as previously mentioned, the transect was chosen as part of a bigger project to essentially cover the four coffee AEsZs. In future, the methods employed in this study should be applied in larger geographical areas to test its up-scalability. Since our study transect covers a gradient of 1300 to 2000 m a.s.l, this could have influenced our LULC mapping results. Further studies should include topographic variables such as elevation and slope to reduce their expected confound effect in coffee mapping experiments in areas of varying topography.[44]

6 Conclusions

In this study, we have shown that S2 is a reliable satellite data to map LULC types with a high level of accuracy in heterogeneous landscapes, such as coffee growing areas dominated by smallholder farms. This is due to a high number of spectral bands that delineate vegetation-based and other LULC types better than other satellites. The SWIR bands were the most important in the LULC classification. Since S2 is freely available, the approach used in this study can be replicated in a resource-constrained context. We studied a coffee growing area to highlight the

complex landscape dynamics in agrosystems that varied within an AEZ. Land-use policies on agricultural landscape management should recommend landscape-specific practices instead of blanket recommendations to improve landscape resilience and connectivity. For future studies, we recommend that detailed studies be conducted to quantify the ecological significance of unique landscape structures in each AEsZ.

7 Appendix

Tables 7–9 provide scene description for satellite images used in the study and variable importance for vegetation and texture variables when used independently and when combined with wavelength bands, respectively.

Table 7 Scene description for each satellite dataset

Satellite imagery	Scene identity (ID)	Date of acquisition	Cloud cover (%)	Source
PS	Analytic Ortho Tile no.807181	03/10/2017	0	https://www.planet.com/
Sentinel 2 (S2)	S2A_MSIL1C_20170827T075211_N0205_R092_T37MBV	27/08/2017	2	scihub.esa.int/dhus/
Landsat 8 (L8)	LC08_L1TP_168061_20171228_20180103_01_T1	28/12/2017	0	https://earthexplorer.usgs.gov/

Table 8 Variable importance (%) for vegetation indices and texture variables (described in Table 2) from Sentinel 2, Landsat 8, and PS datasets.

Vegetation indices							Texture variables						
	Sentinel 2		Landsat 8		PS			Sentinel 2		Landsat 8		PS	
Variable	MDA	MDG	MDA	MDG	MDA	MDG	Variable	MDA	MDG	MDA	MDG	MDA	MDG
BI	25	42	16	22	26	27	MAX	9	8	9	7	9	7
BI2	18	14	11	8	15	14	Homogeneity	11	7	11	8	8	8
GEMI	12	10	9	8	10	12	GLCM variance	11	16	12	18	13	16
GNDVI	10	15	14	18	11	11	GLCM mean	8	13	10	15	11	16
MSAVI	10	9	8	12	11	14	GLCM correlation	11	12	11	12	11	13
NDPI	29	30	28	20	—	—	Entropy	8	8	9	7	8	7
RI	19	20	16	12	28	22	Energy	8	8	7	7	8	8
MCARI	13	10	—	—	—	—	Dissimilarity	10	8	10	9	10	7
MTCI	14	6	—	—	—	—	Contrast	15	11	14	11	14	9
REIP	11	4	—	—	—	—	ASM	9	8	7	7	7	8
S2REP	11	4	—	—	—	—							
LAI	11	19	—	—	—	—							
LAI_CW	15	19	—	—	—	—							
LAI_CAB	10	16	—	—	—	—							
FCOVER	7	10	—	—	—	—							
FAPAR	7	11	—	—	—	—							

Table 9 Variable importance (%) for wavelength bands (described in Table 1) combined with vegetation indices and texture variables (described in Table 2) for PS, Landsat 8, and Sentinel 2 datasets

Sentinel 2			PS			Landsat 8		
Variable	MDA	MDG	Variable	MDA	MDG	Variable	MDA	MDG
B12	4	9	RI	12	12	B7	5	8
B3	4	7	B2	6	10	B4	4	6
B5	4	7	BI	5	9	NDPI	7	6
B11	4	5	B3	5	8	GLCM mean	4	6
NDPI	5	5	B1	5	8	GLCM variance	4	5
BI	3	5	GLCM mean	5	6	B2	3	5
B2	3	5	BI2	6	6	GLCM correlation	4	5
GLCM variance	3	4	GLCM variance	5	6	GNDVI	3	5
RI	4	4	B4	6	5	B3	4	5
GLCM mean	2	3	MSAVI	4	5	Contrast	6	4
GLCM correlation	3	3	GEMI	4	4	Dissimilarity	5	4
B4	2	3	GLCM correlation	6	4	BI	3	4
LAI_CW	3	3	GNDVI	5	3	B6	4	4
Contrast	4	3	Contrast	5	2	RI	5	4
Energy	2	2	Energy	4	2	B5	5	4
ASM	2	2	ASM	4	2	Entropy	4	3
LAI	2	2	Dissimilarity	4	2	MSAVI	3	3
MAX	3	2	Entropy	4	2	GEMI	4	3
GNDVI	2	2	Homogeneity	4	2	BI2	4	3
LAI_CAI	2	2	MAX	4	2	Homogeneity	5	3

Acknowledgments

We gratefully acknowledge the financial support for this research by the following organizations and agencies: French Development Agency (AFD); UK's Foreign, Commonwealth & Development Office (FCDO); Swedish International Development Cooperation Agency (SIDA); the Swiss Agency for Development and Cooperation (SDC); Federal Democratic Republic of Ethiopia; and the Kenyan Government. Gladys Mosomtai was supported by In-Region Postgraduate Scholarship from German Academic Exchange Service (DAAD) and L'Oréal-UNESCO For Women in Science Fellowship. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. T. Tscharntke et al., "Landscape moderation of biodiversity patterns and processes—eight hypotheses," *Biol. Rev.* **87**(3), 661–685 (2012).

2. G. Martel et al., "Impact of farming systems on agricultural landscapes and biodiversity: from plot to farm and landscape scales," *Eur. J. Agron.* **107**, 53–62 (2019).
3. D. Saah et al., "Land cover mapping in data scarce environments: challenges and opportunities," *Front. Environ. Sci.* **7**, 150 (2019).
4. D. Garrity, J. Dixon, and B. Jean-Marc, "Understanding African Farming Systems, Science and Policy Implications," 2012, https://aifsc.aciar.gov.au/sites/default/files/understanding_african_farming_systems_report_for_aifsc_conference.pdf (accessed 30 May 2019).
5. R. Duflot et al., "Landscape level processes driving carabid crop assemblage in dynamic farmlands," *Popul. Ecol.* **58**(2), 265–275 (2016).
6. T. Diekötter and T. O. Crist, "Quantifying habitat-specific contributions to insect diversity in agricultural mosaic landscapes," *Insect. Conserv. Divers.* **6**(5), 607–618 (2013).
7. A. C. Smith, L. Fahrig, and C. M. Francis, "Landscape size affects the relative importance of habitat amount, habitat fragmentation, and matrix quality on forest birds," *Ecography* **34**(1), 103–113 (2011).
8. J. Avelino et al., "Landscape context and scale differentially impact coffee leaf rust, coffee berry borer, and coffee root-knot nematodes," *Ecol. Appl.* **22**(2), 584–596 (2012).
9. C. Vancutsem et al., "Remote sensing harmonizing and combining existing land cover/land use datasets for cropland area monitoring at the African Continental Scale," *Remote Sens.* **5**, 19–41 (2013).
10. M. S. Reis et al., "Towards a reproducible LULC hierarchical class legend for use in the Southwest of Pará State, Brazil: a comparison with remote sensing data-driven hierarchies," *Land* **7**(2), 65 (2018).
11. M. C. Hansen et al., "Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data," *Remote Sens. Environ.* **185**, 221–232 (2016).
12. J. Li and D. P. Roy, "A global analysis of Sentinel-2A, Sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring," *Remote Sens.* **9**(9), 903 (2017).
13. ICO, "Country coffee profile," (2019).
14. F. Anthony et al., "The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers," *Theor. Appl. Genet.* **104**(5), 894–900 (2002).
15. C. Bunn et al., "A bitter cup: climate change profile of global production of Arabica and Robusta coffee," *Clim. Change* **129**(1-2), 89–101 (2015).
16. H. Ketema et al., "Quantifying smallholder farmers' managed land use/land cover dynamics and its drivers in contrasting agro-ecological zones of the East African Rift," *Glob. Ecol. Conserv.* **21**, e00898 (2020).
17. L. Breiman, "Random forests," *Mach. Learn.* **45**(1), 5–32 (2001).
18. P. Ochungo et al., "Multi-sensor mapping of honey bee habitats and fragmentation in agro-ecological landscapes in Eastern Kenya," *Geocarto Int.* **0**(0), 1–22 (2019).
19. R. Jaetzold et al., *Farm Management Handbook of Kenya*, Subpart B2, Vol. II (2007).
20. F. M. DaMatta et al., "Ecophysiology of coffee growth and production," *Braz. J. Plant. Physiol.* **19**(4), 485–510 (2007).
21. GoK, "Murang'a County Integrated Development Plan 2018–2022," (2018).
22. Planet Labs, "Planet imagery product specification," 2016, https://www.planet.com/products/satellite-imagery/files/Planet_Imagery_Product_Specs.pdf.
23. H. K. Zhang et al., "Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences," *Remote Sens. Environ.* **215**, 482–494 (2018).
24. B. Tawona et al., "Is it possible to discern Striga weed (*Striga hermonthica*) infestation levels in maize agro-ecological systems using in-situ spectroscopy?" *Int. J. Appl. Earth Obs. Geoinf.* **85**, 102008 (2020).
25. M. Ovuka and S. Lindqvist, "Rainfall variability in Murang'a district, Kenya: meteorological data and farmers' perception," *Geogr. Ann. Ser. A* **82**(1), 107–119 (2000).
26. G. Vivone et al., "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.* **53**(5), 2565–2586 (2015).
27. R. Mathieu et al., "Relationships between satellite-based radiometric indices simulated using laboratory reflectance data and typic soil color of an arid environment," *Remote Sens. Environ.* **66**(1), 17–28 (1998).

28. J. P. Lacaux et al., "Classification of ponds from high-spatial resolution remote sensing: application to Rift Valley Fever epidemics in Senegal," *Remote Sens. Environ.* **106**(1), 66–74 (2007).
29. J. Qi et al., "A modified soil adjusted vegetation index," *Remote Sens. Environ.* **48**(2), 119–126 (1994).
30. J. Xue and B. Su, "Significant remote sensing vegetation indices: a review of developments and applications," *J. Sens.* **2017**, 1–17 (2017).
31. R. Escadafal, "Remote sensing of soil color: principles and applications," *Remote Sens. Rev.* **7**(3-4), 261–279 (1993).
32. J. Delegido et al., "A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems," *Eur. J. Agron.* **46**, 42–52 (2013).
33. A. Chemura et al., "Mapping spatial variability of foliar nitrogen in coffee (*Coffea arabica* L.) plantations with multispectral Sentinel-2 MSI data," *ISPRS J. Photogramm. Remote Sens.* **138**, 1–11 (2018).
34. T. B. Raper and J. J. Varco, "Canopy-scale wavelength and vegetative index sensitivities to cotton growth parameters and nitrogen status," *Precis. Agric.* **16**(1), 62–76 (2015).
35. M. Weiss and F. Baret, "S2ToolBox Level 2 products: LAI, FAPAR, FCOVER – Version 1.1. Sentin ToolBox Level2 Prod," 53, 2016, http://step.esa.int/docs/extra/ATBD_S2ToolBox_L2B_V1.1.pdf.
36. R. M. Haralick, K. Shanmugam, and D. I. Haralick, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **SMC-3**(6), 610–621 (1973).
37. G. Redzwan and M. F. Ramli, "Geo-referencing the satellite image from Google Earth by relative and absolute positioning," *Malays. J Sci.* **26**(2), 135–141 (2007).
38. M. Belgiu and L. Drăgu, "Random forest in remote sensing: a review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.* **114**, 24–31 (2016).
39. A.-L. Boulesteix et al., "Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2**(6), 493–507 (2012).
40. A. Htitiou, A. Boudhar, and Y. Lebrini, "The performance of random forest classification based on phenological metrics derived from Sentinel-2 and Landsat 8 to map crop cover in an irrigated semi-arid region," *Remote Sens. Earth Syst. Sci.* **2**, 208–224 (2019).
41. H. Han, X. Guo, and H. Yu, "Variable selection using mean decrease accuracy and mean decrease gini based on random forest," in *Proc. IEEE Int. Conf Softw. Eng. Serv. Sci.*, pp. 219–224 (2016).
42. A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News* **2**(3), 18–22 (2002).
43. R Core Team, "R: the R project for statistical computing," r-project.org/.
44. P. Hurskainen et al., "Auxiliary datasets improve accuracy of object-based land use/land cover classification in heterogeneous savanna landscapes," *Remote Sens. Environ.* **233**, 111354 (2019).
45. P. Verma et al., "Appraisal of kappa-based metrics and disagreement indices of accuracy assessment for parametric and nonparametric techniques used in LULC classification and change detection," *Model. Earth Syst. Environ.* **6**, 1045–1059 (2020).
46. K. McGarigal et al., "FRAGSTATS: spatial pattern analysis program for categorical maps," Computer Software Program (2002).
47. L. Kumsa et al., "Patch area and current coffee management determine woody plant diversity in patches of semi-forest coffee embedded in an agricultural matrix," *Glob. Ecol. Conserv.* **8**, 230–240 (2016).
48. G. E. Crozier and G. J. Niemi, "Using patch and landscape variables to model bird abundance in a naturally heterogeneous landscape," *Can. J. Zool.* **81**(3), 441–452 (2003).
49. S. Krishnan et al., "Status of pollinators and their efficiency in coffee fruit set in a fragmented landscape mosaic in South India," *Basic Appl. Ecol.* **13**(3), 277–285 (2012).
50. C. Shoko and O. Mutanga, "Examining the strength of the newly-launched Sentinel 2 MSI sensor in detecting and discriminating subtle differences between C3 and C4 grass species," *ISPRS J. Photogramm. Remote Sens.* **129**, 32–40 (2017).
51. N. Levin, "Fundamentals of remote sensing," *Resour. Policy* **2**(1), 65 (1976).

52. M. A. Ortega-Huerta et al., "Mapping coffee plantations with land sat imagery: an example from El Salvador," *Int. J. Remote Sens.* **33**(1), 220–242 (2012).
53. S. Cordero-Sancho and S. A. Sader, "Spectral analysis and classification accuracy of coffee crops using Landsat and a topographic-environmental model," *Int. J. Remote Sens.* **28**(7), 1577–1593 (2007).
54. M. S. de Oliveira Pires, M. de Carvalho Alves, and E. A. Pozza, "Multispectral radiometric characterization of coffee rust epidemic in different irrigation management systems," *Int. J. Appl. Earth Obs. Geoinf.* **86**, 102016 (2020).
55. R. Escadafal and A. R. Huete, "Soil optical properties and environmental applications of remote sensing," *Int. Arch. Photogramm. Remote Sens.* **29**(B7), 709–715 (1993).
56. T. Liebig et al., "Local and regional drivers of the African coffee white stem borer (*Monochamus leuconotus*) in Uganda," *Agric. For. Entomol.* **20**(4), 514–522 (2018).
57. P. Kindlmann and F. Burel, "Connectivity measures: a review," *Landsc. Ecol.* **23**(8), 879–890 (2008).
58. J. Jaramillo et al., "Thermal tolerance of the coffee berry borer Hypothenemus hampei: predictions of climate change impact on a tropical insect pest," *PLoS One* **4**(8), e6487 (2009).
59. Y. Ziv and G. Davidowitz, "When landscape ecology meets physiology: effects of habitat fragmentation on resource allocation trade-offs," *Front. Ecol. Evol.* **7**, 1–8 (2019).
60. F. Lescourret et al., "A social–ecological approach to managing multiple agro-ecosystem services," *Curr. Opin. Environ. Sustain.* **14**, 68–75 (2015).
61. T. Landmann et al., "Application of hyperspectral remote sensing for flower mapping in African savannas," *Remote Sens. Environ.* **166**, 50–60 (2015).
62. GBIF.org, "GBIF home page," https://www.gbif.org/ (2020).
63. OpenStreetMap contributors, "Planet OSM," https://www.openstreetmap.org/ (2020).
64. T. Landmann et al., "Optimizing harmonics from Landsat time series data: the case of mapping rainfed and irrigated agriculture in Zimbabwe," *Remote Sens Lett.* **10**(11), 1038–1046 (2019).

Biographies of the authors are not available.