 doi:10.20944/preprints202010.0517.v1

Article

Sophisticated cropping patterns mapping in Kenya using vegetation indices and phenology from Sentinel-2 and Sentinel-1 radar backscatter data

Grace Rebecca Aduvukha[1, 2*], Elfatih M. Abdel-Rahman[1, 4], Arthur W. Sichangi[2], Godfrey Ouma Makokha[2, 6], Tobias Landmann[1, 3], Bester Tawona Mudereri[1, 5], Henri E. Z. Tonnang[1] and Thomas Dubois[1]

[1] International Centre of Insect Physiology and Ecology (*icipe*), P.O. Box 30772, 00100 Nairobi, Kenya gaduvukha@icipe.org (G.R.A); eabdel-Rahman@icipe.org (E.M.A-R); bmudereri@icipe.org (B.T.M); htonnang@icipe.org (H.E.Z.T); tdubois@icipe.org (T.D)

[2] Institute of Geomatics, GIS & Remote Sensing, Dedan Kimathi University of Technology, Private Bag, 10143 Nyeri, Kenya; arthursichangi@yahoo.com (A.W.S); makokha.godfrey@gmail.com (G.O.M)

[3] RSS-Remote Sensing Solutions Gmbh, Dingolfinger Str. 9, 81673 Munich, Germany; landmann@rssgmbh.de (T.L)

[4] Department of Agronomy, Faculty of Agriculture, University of Khartoum, Khartoum North 13314, Sudan

[5] Department of Animal and Wildlife Science, Midlands State University, P. Bag 9055, Gweru, Zimbabwe

[6] School of Science and Informatics, Taita Taveta University, P.O. Box 635, 80300 Voi, Kenya

* Correspondence: gaduvukha@icipe.org

Abstract: The proportion of area under various crops at a given point in time, known as a cropping pattern, plays an essential role in determining the level of agricultural production. In this study, cropping patterns of three sub-counties in Murang'a County, a typical African smallholder farming area in Kenya, were mapped. Specifically, we compared the performance of eight classification scenarios for mapping cropping patterns; namely using (i) only Sentinel-2 reflectance bands (S2), (ii) S2 and S2 derived vegetation indices (VIs); (iii) S2 and S2 vegetation phenology (VP); (iv) S2 and Sentinel-1 radar backscatter data (S1); (v) S2, VIs, and S1; (vi) S2, VP, and S1; (vii) S2, VIs and VP, and (viii) S2, VIs, VP and S1. Reference data of the dominant cropping patterns and non-croplands were collected. The guided regularized random forest (GRRF) algorithm was used to select the optimum variables and to perform the respective classification for each scenario. The most accurate result of the overall accuracy of 93.16% was attained from the scenario (viii) S2, VIs, VP, and S1. The McNemar's test of significance did not show significant differences ($p \leq 0.05$) among the tested scenarios. Our study demonstrated the strength of GRRF and the synergetic advantage of S2 and S1 derivatives to map cropping patterns in a heterogeneous landscape where high resolution imagery are inaccessible. Our cropping pattern mapping approach can be used in other sites of relatively similar agro-ecological conditions.

Keywords: agricultural productivity; cropping patterns; Kenya; multi-data analysis

1. Introduction

Cropping patterns and food security in Africa are highly affected by a myriad of factors including climate change and variability, and high pests and disease infestation levels, thus causing low production rates [1]. The proportion of area under various crops at a given point in time, known as a cropping pattern [2], plays an essential role in determining the level of agricultural production, particularly for small-scale farmers in Africa [3]. Despite the importance of coherent cropland information, studies have indicated that the small-scale structures (< 0.1 ha) and fragmentation of these cropping patterns in Africa, triggered by their high intra- and inter-seasonal dynamics, prohibit their accurate detection and characterization [4]. This complexity is further aggravated by the individual farmers' crop selection for planting in a particular season. The choice of crop for planting

by farmers varies spatially and among farmers across the different agro-ecological systems. These cropping patterns are also influenced by the availability of optimum environmental conditions for the crop, its traditional use as food and feed, its profitability as well as its tolerance and resistance to insect pests and diseases [5].

Therefore, accurate detection of these complex cropping patterns, which render information feeds beyond 'just' crops and cropland, is key for providing verifiable, realistic, and site-specific advice to farmers and policymakers. The types of cropping patterns can include monocropping, crop rotation, and intercropping [6]. These cropping patterns possess several benefits and drawbacks. For instance, monocropping patterns ensure specialized crop production and expected higher earnings due to mass crop production. However, monocropping practice is characterized by, for example, high risk of pests and low soil microbe diversity [7]. On the other hand, crop rotation reduces soil erosion and increases soil fertility, which results in improved crop yields [8]. Notwithstanding, the mandatory and regular crop diversification required in crop rotation may be a strain for the farmers if there are no readily available resources for its implementation. On the other hand, intercropping promotes soil fertility, reduces pest risks, and maximizes land profit. Nevertheless, the different crops in an intercropping practice may require different uptake of resources such as water and fertilizer, which may not be adequately utilized by the different crops [6]. Intercropping comprises variants such as rows, relay, and mixed cropping, but the present study refers to the different groups of intercropping patterns as mixed cropping in general. The present study focuses on two types of cropping patterns, i.e. monocropping patterns and mixed cropping patterns that are commonly practiced in the study's test site.

To apply on-farm technologies, advice, and policymaking require precise spatiotemporal information on cropping patterns [4]. Usually, ground-based surveys and inspection methods are used to determine the commonly grown crops in an area, and hence zonal advice or plans are provided for all farmers in that area. This approach is often inadequate, expensive, has a long-time lag, is laborious, and provides incomplete information necessary for precision agriculture and efficient utilization of resources and effective pest management [4]. In contrast, the recent advancement in remote sensing technology provides efficient, timely, synoptic, and inexpensive data that could effectively capture the cropping variability at different spatiotemporal scales [9], in our case at 10 m pixel size using Sentinel-2 and-1 (S2 and S1) data.

The use of remotely sensed data for mapping and modeling cropping patterns and other agronomic practices are well documented in the literature. This has been evidenced by studies that were carried out for mapping of various croplands [10-12]; crop types [13-15] and cropping patterns [16-18] with a view of investigating the relevance of different remote sensing systems and image classification methods for improving classification accuracies, reliability, and reproducibility of the results. Diverse imaging systems, analytical techniques, and spectral variables have also been explored for cropping pattern classification. These systems range from optical multispectral [19-21] and hyperspectral imaging [22,23] to radar sensors [24,25] with high, medium, and coarse spatial resolutions. Also, studies have used different remotely sensed variables like single date [26] and multi-date (time-series) vegetation indices [16] and phenometrics [27,28], to map cropping patterns.

However, challenges in mapping cropping patterns are evident in Africa [29]. These challenges include inter- and intra-season changes in crop phenological cycles due to variability in farming practices or weather conditions, and highly fragmented landscapes on which the crops are grown [14,16,30]. Previous studies have thus proposed the use of optical and radar datasets of better spatial resolutions [11], the use of crop phenological variables extracted from the relatively new satellite sensors such as S2 [31], the use of multisource remotely sensed, and ancillary data to improve the quality and timeliness of in-season cropping patterns mapping [32]. Therefore, this study evaluates the strengths of multi-date medium resolution S2 reflectance imagery, S2 derived vegetation indices (VIs) and vegetation phenological variables (VP), and S1 radar backscatter data (S1) for mapping cropping patterns in an agro-natural production system in Murang'a County, Kenya. Specifically, we assessed the performance of eight multi-sensor classification scenarios for delineating cropping patterns using the guided regularized random forest (GRRF) machine learning classifier [33]. GRRF

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0517.v1

has shown to be a successful method in reducing the dimensionality in large datasets and simultaneously handles very well the multicollinearity in the data [34,35]. Compared to the standard random forest (RF) classifier, GRRF is superior in its ability to identify optimum uncorrelated variables for analysis [33].

2. Study area

The study was conducted using the case of Murang'a County of Kenya (Figure 1). Murang'a lies between latitudes of 0° 34′ 00′′ S and 1° 07′ 00′′ S and longitudes of 36° 00′ 00′′ E and 37° 27′ 00′′E. The elevation ranges between 900 m and 3,355 m above sea level with an area coverage of 2,326 km^2. Murang'a County is characterized by six agro-ecological zones and three climatic regions i.e. equatorial, subtropical, and semiarid climatic regions [36]. The region has a bimodal rainfall pattern, with long rains in March-May and short rains in October-November. On average, annual rainfall ranges from >800 mm in the southeast to >2,600 mm in the northwest, while the annual temperature range is 12-20°C [37]. Our study area has a complex heterogeneous landscape, translating into heterogeneous cultivation of crops like avocado, maize, common bean, sweet potato, arrowroot, macadamia, pineapple, banana, tea, and coffee, planted in different cropping patterns at different points in time across the cropping seasons. Monocropping is majorly practiced for commercial purposes and mixed cropping mostly for subsistence consumption. However, some crops such as avocado, grown in mixed cropping patterns by smallholder farmers in Murang'a, are also used for commercial purposes [38]. Murang'a County contributes to 40% of total avocado production in Kenya and is a horticultural crop whose export value has been on the rise in the last decade [39]. We selected three sub-counties in Murang'a county (Kandara, Maragwa, and Gatanga), which are dominant for avocado production [39]. Mapping of the cropping patterns in this region can be utilized to improve the understanding of the contribution of each cropping pattern in the abundance of avocado pests and pollinators across different vegetation intensity for sustainable management [40].

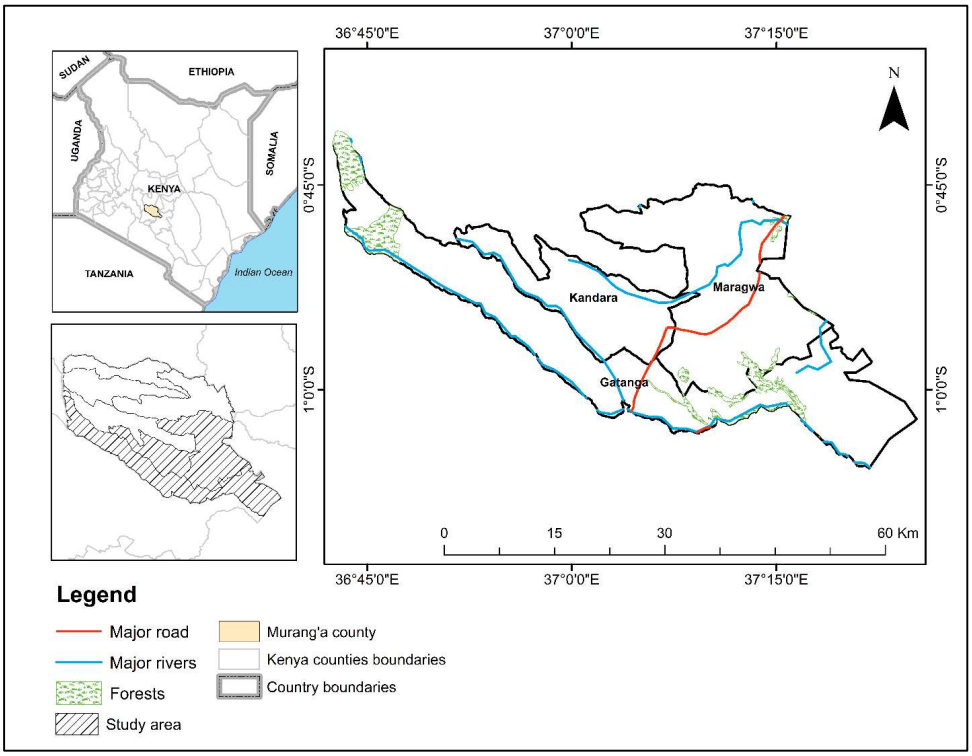

Figure 1. Location of the study area comprising of the three sub-counties, i.e. Kandara, Maragwa, and Gatanga, in Murang'a County, Kenya.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0517.v1

3. Methodology

Figure 2 illustrates the methodological approach used to map the cropping patterns in the study area using different remotely sensed variables extracted from S2 and S1 imageries.

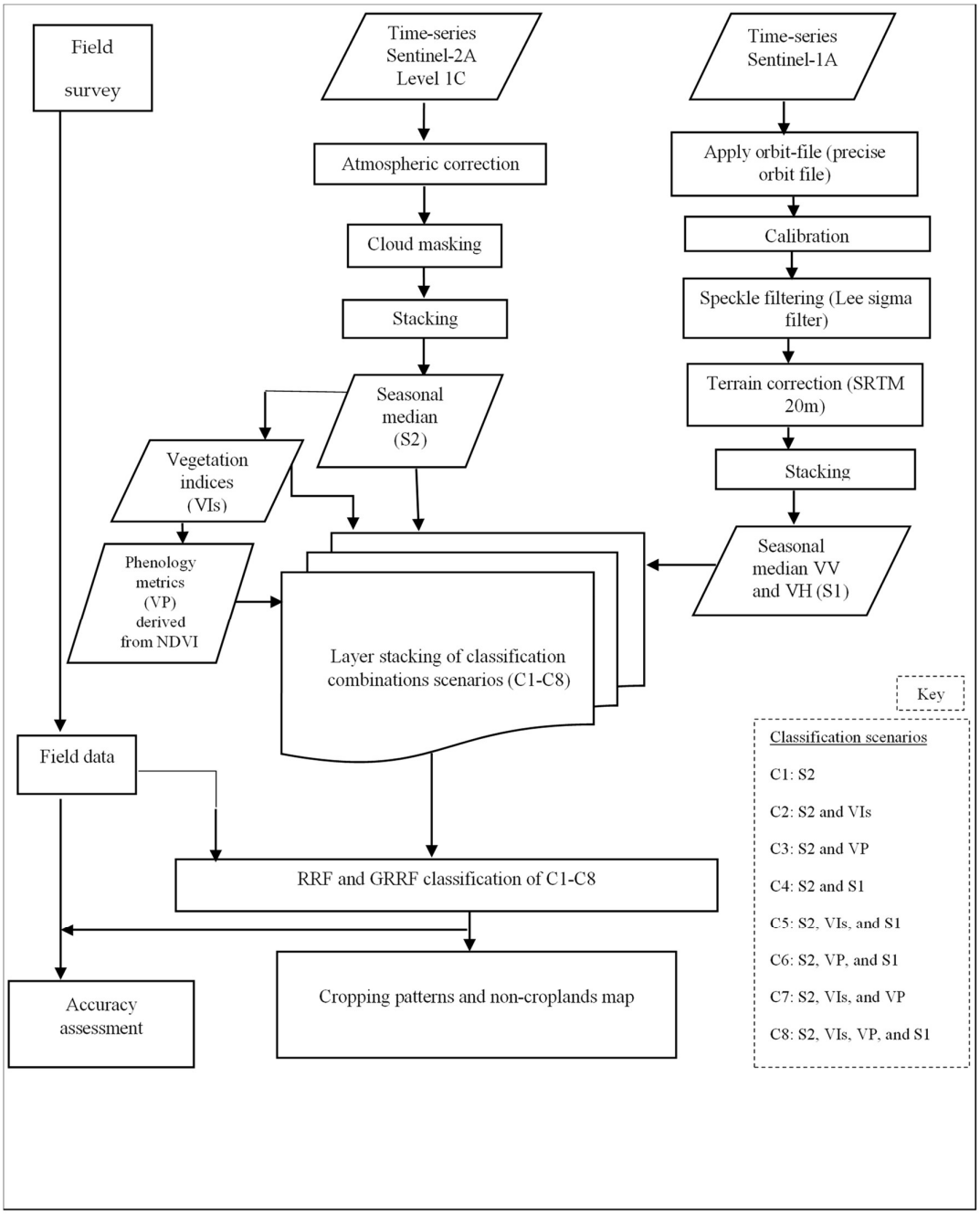

Figure 2. Flow diagram of the methodology adopted for cropping patterns classification. NDVI = normalized difference vegetation index, RRF = regularized random forest, GRRF = guided regularized random forest, SRTM = shuttle radar topography mission, VV = vertical transmit and vertical receive and VH = vertical transmit and horizontal receive.

3.1. Remotely sensed data

The remotely sensed data utilized in this study included multi-date reflectance bands, vegetation indices, phenological metrics from S2, and backscatter from S1. These freely available

datasets of S2 and S1 were selected to emphasize the various parts of the electromagnetic spectrum for accurate cropping pattern mapping.

3.1.1. Sentinel-2

Freely available multi-date S2A level 1C imagery (n=126 scenes) captured during four seasons from 10 December 2017 to 15 December 2018 (Table 1**)** with cloud cover less than 20% were acquired from the European space agency (ESA) Copernicus open access hub [41]. S2 has a swath width of 290 km² providing 13 spectral bands that range in pixel size from 10, 20, and 60 m across the visible, near-infrared, and the shortwave infrared spectrum [41], as shown in Table 2. The S2 images were atmospherically corrected using the Sen2cor module in the sentinel application platform (SNAP) toolbox [41]. In SNAP, we also performed cloud masking, resampling of the S2 20 m spectral bands to 10 m spatial resolution, layer stacking, and computation of the median pixel image for each season. We then used the Murang'a sub-county boundaries to subset the respective median images to our study area. Coastal and aerosol (band 1), water vapor (band 9) and cirrus (band 10) bands (Table 2) were excluded from this analysis as they mostly contribute to atmospheric and geophysical parameters [41], which were not the focus of our current study.

Table 1. Sentinel-2 multispectral data acquisition dates and their specific seasons of the year 2017-2018.

Acquisition date	Season
10 December 2017-25 February 2018	1. Short dry
25 February 2018-20 June 2018	2. Long wet
20 June 2018-20 September 2018	3. Long dry
20 September 2018-15 December 2018	4. Short wet

Table 2. Sentinel-2 multispectral sensor wavebands, description, resolution, and their respective central wavelengths. Bands 1, 9, and 10 were excluded from this analysis.

Band	Resolution (m)	Central wavelength (nm)	Description
B1	60	443	Ultra-blue (Coastal and aerosol)
B2	10	490	Blue
B3	10	560	Green
B4	10	665	Red
B5	20	705	Vegetation red edge (RE1)
B6	20	740	Vegetation red edge (RE2)
B7	20	783	Vegetation red edge (RE3)
B8	10	842	Near-infrared (NIR)
B8a	20	865	Narrow NIR (NNIR)
B9	60	940	Short wave infrared (SWIR1) water vapor
B10	60	1375	Short wave infrared (SWIR2)-cirrus
B11	20	1610	Short wave infrared (SWIR3)
B12	20	2190	Short wave infrared (SWIR4)

3.1.2. Vegetation indices

VIs are a combination of spectral characteristics of two or more wavelength bands that indicate the relative abundance of vegetation components such as chlorophyll and water contents[42]. The VIs described in Table 3 were selected and derived for this study from the median images of our multi-date S2 imagery. Composite images of each of the VIs (n=8) were then created. These VIs were selected for cropping patterns mapping as they are the most robust indices that mimic vegetation seasonality and are known to reduce residual contamination due to atmospheric noise and soil background [26]. The multi-date and multi-season aspects in the VIs can also cater for NDVI anisotropic effects [43].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0517.v1

Table 3. Vegetation indices that were used in the present study.

No.	Index	Abbreviation	Formula	Reference
1	Normalized difference vegetation index	NDVI	(NIR-Red)/(NIR+Red)	[44]
2	Enhanced vegetation index	EVI	2.5 *((NIR-Red)/(NIR+6*Red-7.5* Blue+1))	[45]
3	Two-band enhanced vegetation index	EVI2	2.5 *((NIR-Red)/(NIR+2.4*Red+1))	[46]
4	Normalized difference water index	NDWI	(NIR-SWIR3)/(NIR+SWIR3)	[47]
5	Modified soil adjusted vegetation index	MSAVI	(NIR-Red)(1+L)/(NIR+Red+L)	[48]
6	Soil adjusted vegetation index	SAVI	1.5 *((NIR-Red)/(NIR+Red+0.5))	[49]
7	Green normalized difference vegetation index	GNDVI	(NIR-Green)/(NIR+Green)	[50]
8	Atmospherically resistant vegetation index-2	ARVI2	-0.18+1.17((NIR-Red)/(NIR+Red))	[51]

Note: NIR = near infrared, SWIR = short wave infrared; Blue, Green, Red, NIR, and SWIR3 are surface reflectance at bands 2, 3, 4, 8, and 11 of Sentinel-2 respectively; and L= 1-(2*s*(NIR-Red)*(NIR-s*Red))/(NIR+Red) where s is the slope of the soil line from a plot of red versus near-infrared brightness values.

3.1.3. Vegetation phenological variables

VP variables mimic the growth life cycle of a plant [52]. The VP variables as described by Araya [53] were derived in the present study and are summarized in Table 4. The VP variables were simulated from the multi-date NDVI curve of the S2 images of each of the four seasons (Table 1). A composite image for the VP variables (n=15) was created and used in the cropping pattern classification experiment.

Table 4. Description of vegetation phenological variables derived in the present study.

No.	Phenological variable	Definition of the normalized difference vegetation index (NDVI) curve and physiological description
1	Onset_value	The NDVI value at the start of the growth (seedling growth stage)
2	Onset_time	The time when the growth onset is achieved
3	Max_value	The maximum NDVI value in the season
4	Max_time	The time when the max_value is attained (anthesis growth stage)
5	Offset_value	The NDVI value at the end of the season
6	Offset_time	The time when growth offset is attained (senescence growth stage)
7	LengthGS	The length of the growing season
8	BeforeMaxT	The length of time between onset and max_value
9	AfterMaxT	The length of time between max_value and offset
10	GreenUpSlope	The rate of increase in NDVI value between onset and offset
11	BrownDownSlope	The rate of decrease in NDVI value between max_value and offset
12	TINDVI	The area under the NDVI curve between onset and offset
13	TINDVIBeforeMax	The area under the NDVI curve between onset and max_value
14	TINDVIAfterMax	The area under the NDVI curve between max_value and offset
15	TINDVIAsymmetry	The difference between BeforeMaxTINDVI and AfterMaxTINDVI

 doi:10.20944/preprints202010.0517.v1

3.1.4. Sentinel-1 radar backscatter data

Multi-date S1 imagery (53 in total) of four seasons were acquired from the ESA Copernicus open access hub [41] and used in this analysis (Table 5). The acquisition dates of S1 imagery were slightly different from those of S2 due to image availability in the ESA archive. S1 sensor provides C-band synthetic aperture radar (SAR) images in both singular and dual-polarization with a repeat cycle of 12 days [41]. Acquisition of these images are in four modes i.e., stripmap (SM), interferometric wide swath (IW), extra-wide swath (EW), and wave (WV) with different processing levels i.e. Level-0, Level-1 (Single Look Complex-(SLC), ground range detected-(GRD)) and Level-2 [41]. In this study, Level-1 S1 products of GRD and IW were used. The S1 images were dual-polarized in vertical transmit and vertical receive (VV) and vertical transmit and horizontal receive (VH) mode. The pre-processing procedures of S1 images were performed in the SNAP toolbox [41]. Based on Filipponi [54], these processes included applying the precise orbit file (provides accurate satellite position and velocity information), radiometric calibration using calibration beta, and speckle filtering using the lee sigma filter. Similarly, we computed the terrain correction using 20 m shuttle radar topography mission (SRTM) data and resampled the data to 10 m pixel size. These processed images were then stacked to produce per season median pixel value of the VV and VH image bands that were then subset to the extent of the study area.

Table 5. Sentinel-1 radar backscatter data acquisition dates and their specific seasons in the years 2017 and 2018.

Acquisition date	Season
10 December 2017-1 March 2018	1. Short dry
1 March 2018-15 June 2018	2. Long wet
15 June 2018-15 October 2018	3. Long dry
15 October 2018-31 December 2018	4. Short wet

3.2. Field data collection

A stratified random sampling approach was used for the field data collection, using three vegetation intensity classes (i.e. low, medium, and high) as strata at a landscape scale using the *K*-means clustering method [40]. The *K*-means clustering method groups similar data points together and uncovers underlying patterns [55]. The stratified random sampling was done to provide a comprehensive representation of the cropping patterns within different vegetation intensity in the study area at a landscape scale. Field reference data for the cropping patterns and non-croplands were collected during the period ranging from 13 December 2018 to 19 December 2018, corresponding to the short wet season. The cropping patterns classes included monocrop avocado, monocrop coffee, monocrop maize, monocrop tea, monocrop pineapple, mixed crop avocado, and mixed crop maize. The non-cropland classes included water bodies, built-up area, grassland, shrubland, and natural vegetation. The non-cropland classes were included due to the heterogeneity of the landscape in the study area. Reference points were field sampled randomly within each of the three vegetation intensity strata, ensuring that all classes had more than thirty data entries well spread across the strata. The on-field data of each class were collected as points (i.e. pixels), which were then converted to homogenous units (i.e. polygons) using on-screen digitization on high-resolution Google Earth imagery [56]. The pixel values within the homogenous polygons were then extracted to be utilized for the cropping pattern predictions as summarized in Table 6.

Table 6. The number of polygons and pixels used per class.

Class	No. of Polygons	No. of Pixels
Built-up area	86	375
Grassland	76	936
Mixed crop avocado	85	191
Mixed crop maize	54	299
Monocrop avocado	74	242
Monocrop coffee	49	1,462

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0517.v1

Monocrop maize	56	479
Monocrop pineapple	76	1,451
Monocrop tea	78	406
Natural vegetation	85	3,005
Shrubland	59	743
Water	38	1,397
Total	**872**	**10,986**

3.3. Predictor variables selection and classification

The GRRF algorithm was used to select the most important variables in eight individual classification scenarios (Table 7). GRRF algorithm works with using a concept similar to that of the random forest (RF) method. However, RF uses an adaptable number of predictor variables and does not adhere to the assumption of a normal distribution of the predictor variables [57], whereas GRRF additionally uses the importance scores from a preliminary RF to guide the feature selection process of regularized random forest (RRF). In RRF, a feature is selected by building only one ensemble where the feature importance is evaluated on a part of the training data at each node [33]. However, the limitation of RRF feature selection is that many features can share the same information gain at a node with a small number of instances and a large number of features, causing a likelihood of RRF to select a feature that is not strongly relevant. To counter for the limitations of RRF, GRRF assigns a penalty coefficient to each feature by changing the importance coefficient of gamma value ($\gamma \in (0, 1)$), which controls the weight of normalized importance. This ensures that the most relevant variables are retained while still ensuring accurate classification [58]. Thus, in this study γ=0.7 was used in the "CoefReg" function in the "caret" package in R software to select the most important variables [59]. The mean decrease in accuracy score value depicts the level of importance of the variable.

We set the number of decision trees grown (*ntree*) to 500 default value and the number of variables used to split the trees (*mtry*) to the square root of the number of variables, which is also a default setting [57]. Furthermore, we used the flagReg=1, which initiates the penalization of the variables at each node and to determine the most important variables with a tune length of 3. The tune length parameter controls the algorithm to attempt different values for the main parameter in the prediction. The eight individual classification scenarios (C1-C8) combined different multi-date S2 and S1 variables (Table 7) and their counterparts that combined only the selected S2 and S1 variables i.e. $C1_{selected}$-$C8_{selected}$ were then carried out using the GRRF algorithm.

Table 7. The classification scenarios and the remotely sensed variables that were used in each scenario.

Classification Scenario	No. of total variables combined	Variables
C1	40	Sentinel-2 reflectance bands (S2)
C2	48	S2 and S2-based vegetation indices (VIs) i.e. (S2 and VIs)
C3	55	S2 and S2-based vegetation phenology (VP) i.e. (S2 and VP)
C4	48	S2 and Sentinel-1 radar backscatter bands (S1) i.e. (S2 and S1)
C5	56	S2, VIs, and S1
C6	63	S2, VP, and S1
C7	63	S2, VIs, and VP
C8	71	S2, VIs, VP, and S1

3.4. Classification accuracy assessment

We used a 2-fold cross-validation method to train and validate the performance of the classification scenarios [60]. In this approach, each fold had a chance to be a training and testing set. The performance of each classification scenario was then evaluated using the area under class method [61]. The area

 doi:10.20944/preprints202010.0517.v1

under class method establishes the unbiased total area of each class since it includes the area of the map omission error of each class, leaving out the commission error. The producer's accuracy (PA), user's accuracy (UA), and overall accuracy (OA) were then computed from the classification error matrix of estimated area proportions, taking into account the proportion of each class in the study area. Moreover, a class-wise accuracy assessment was performed for each class using the F1-score criterion [62]. F1-score is a measurement that balances the difference between PA and UA for each class i through the formulation of the harmonic mean of PA and UA as shown in Equation 1.

$$(F1)_i = (2 \times PA_i \times UA_i) \div (PA_i + UA_i) \quad (1)$$

The unbiased estimated area of each class and a confidence interval of 95% [61] were also calculated from the most accurate scenario. Additionally, a McNemar's chi-square test [63] was carried out to test for any statistically significant differences ($p \leq 0.05$) among the cropping patterns mapping results from the classification scenarios.

4. Results

4.1. Variable selection using the guided regularized random forest (GRRF) algorithm

S2 reflectance bands across the four seasons in all the GRRF scenarios (Table 8) dominated in the most relevant bands selected. Specifically, the S2 bands of NNIR, SWIR3, RE1, Red, Green, and Blue were selected. Additionally, NDWI and GNDVI were the most relevant VIs selected from the eight indices investigated. The most relevant VP variables selected were the Offset_value, TINDVIBeforeMax, and TINDVIAfterMax, whereas the S1 bands of VV and VH were selected. The specific variables selected in each of the eight scenarios are shown in Figure 3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0517.v1

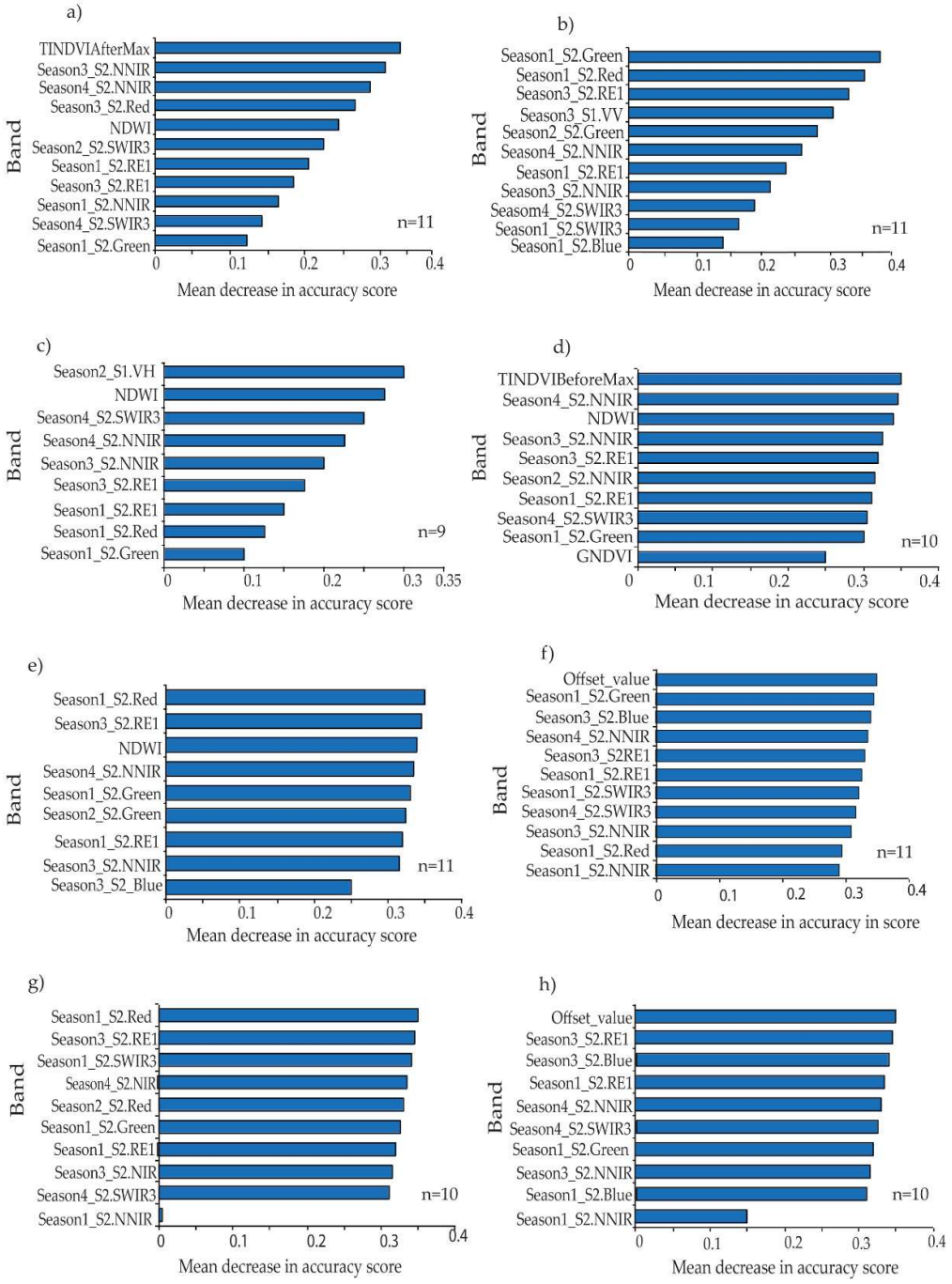

Figure 3. Guided regularized random forest (GRRF) selected variables in each of the eight scenarios i.e.,**(a)** Sentinel-2 (S2), vegetation indices (VIs), vegetation phenology (VP) and Sentinel-1 radar backscatter data (S1); **(b)** S2 and S1; **(c)** S2, VIs, and S1; **(d)** S2, VP, and VIs; **(e)** S2 and VIs; **(f)** S2 and VP; **(g)** S2, and **(h)** S2, VP, and S1.

4.2. Cropping patterns mapping

The cropping pattern maps are shown in Figure 4. The southeastern side of the study area is mainly characterized by monocrop pineapple, the southwestern side is characterized by monocrop coffee, the western side is characterized by monocrop tea while the eastern side is dominated by mixed crop maize, monocrop maize, mixed crop avocado, and monocrop avocado. The non-croplands like the shrublands and grasslands are shown to be dominant on the eastern side whereas natural vegetation is most evident on the northwestern side.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0517.v1

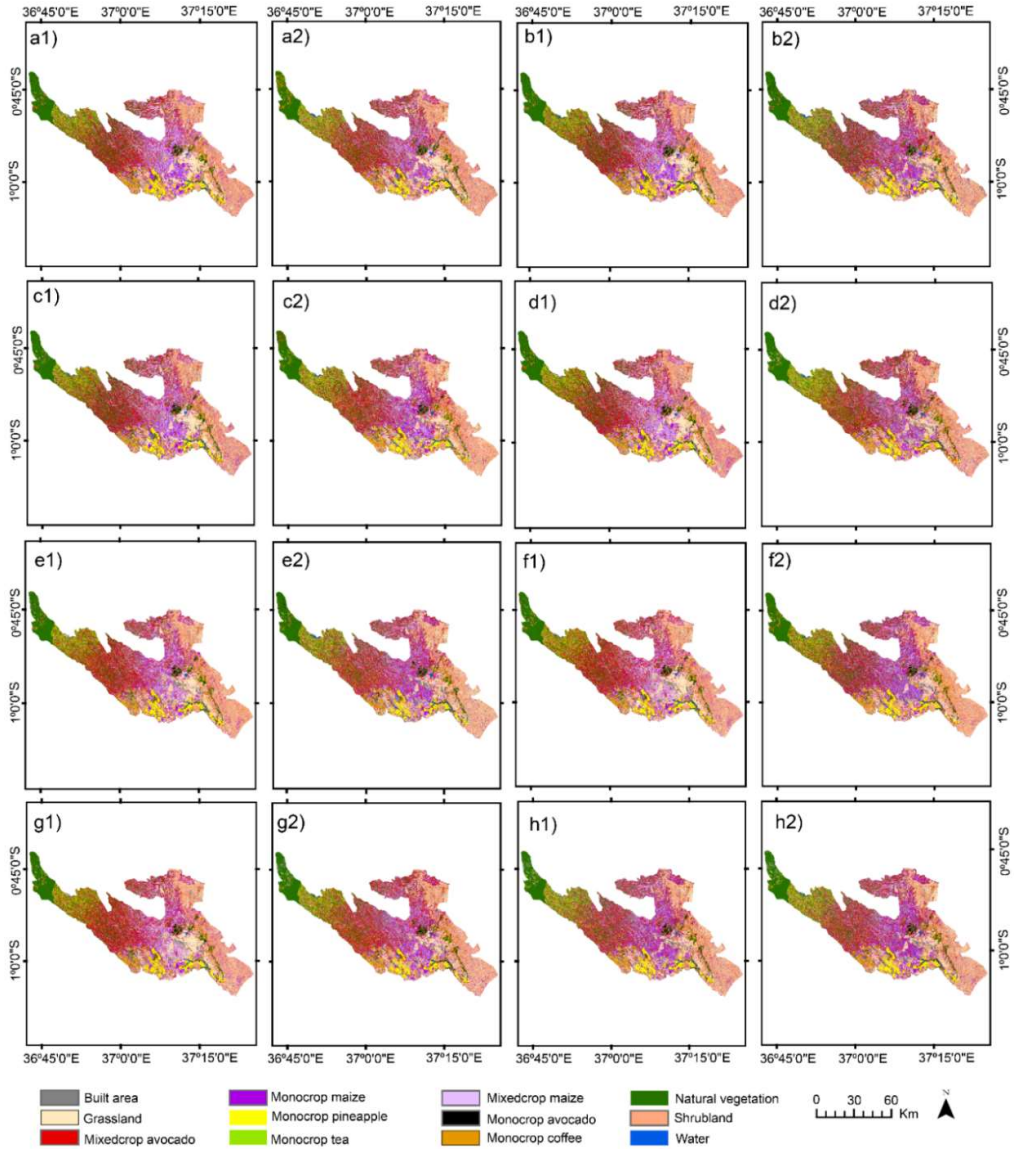

Figure 4. Cropping pattern maps produced from classifications using combined scenarios of variables of **(a1)** Sentinel-2 (S2), vegetation indices (VIs), vegetation phenology (VP) and Sentinel-1 radar backscatter (S1); **(b1)** S2 and S1; **(c1)** S2, VIs and S1; **(d1)** S2, VP and VIs; **(e1)** S2 and VIs; **(f1)** S2 and VP; **(g1)** S2; **(h1)** S2, VP and S1, and guided regularized random forest (GRRF) selected variables of **(a2)** S2, VIs, VP, and S1; **(b2)** S2 and S1; **(c2)** S2, VIs and S1; **(d2)** S2, VP and VIs; **(e2)** S2 and VIs; **(f2)** S2 and VP; **(g2)** S2 and **(h2)** S2, VP and S1.

4.3. Mapping accuracy assessment

The OA of all the classification scenarios was above 90.00% as shown in Figure 5. Among the cropping patterns, the combined scenarios C1-C8 (Table 7) showed the highest PA (100.00%) and F1-score (99.97%) for monocrop pineapple while the highest UA (100%) was recorded for monocrop tea. On the other hand, the lowest PA (86.60%) was shown for monocrop coffee from scenario S2 and VP (C3), the lowest UA (69.63%) for mixed crop avocado from scenario S2, VP and V1s (C7), and the lowest F1-score (80.86%) for mixed crop avocado from scenario S2, VP, and VIs (C7).

Regarding the GRRF selected scenarios ($C1_{selected}$-$C8_{selected}$), the highest PA (100.00%) was shown for monocrop pineapple from the GRRF selected variables of S2 ($C1_{selected}$), S2 and VIs ($C2_{selected}$), S2 and VP ($C3_{selected}$), and S2 and S1 ($C4_{selected}$). Monocrop tea had the highest UA (100.00%) from the GRRF variables of S2 and VIs ($C2_{selected}$), and S2 and VP ($C3_{selected}$), while the highest F1-score (99.97%)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0517.v1

was shown for monocrop pineapple from the GRRF selected variables of S2 and VIs ($C2_{selected}$). In addition, the lowest PA (82.29%) was predicted for mixed crop maize in the GRRF selected variables of S2 ($C1_{selected}$). The lowest UA (67.02%) and F1-score (78.30%) were recorded for mixed crop avocado from the GRRF selected variables of scenario S2 and V1s ($C2_{selected}$). In as much as the non-croplands were not a focus of this present study, water was mapped with class-wise accuracy of 99.00% across all scenarios. A summary of the class-wise accuracies of all scenarios is shown in Figures 6 and 7. However, the McNemar's test for significance showed no statistical significance ($p \leq 0.05$) difference between the combined scenarios and GRRF selected scenarios (Table 8).

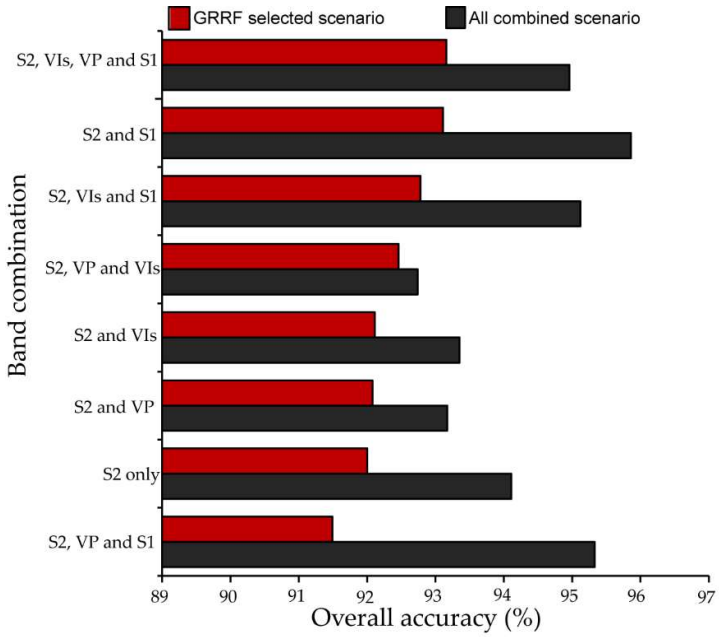

Figure 5. Summary of the overall accuracy of cropping patterns and non-croplands classification using the combined and the guided regularized random forest (GRRF) selected bands of each scenario of Sentinel-2 (S2), vegetation indices (VIs), vegetation phenology (VP), and Sentinel-1 radar backscatter data (S1).

4.4. Area estimation

The unbiased area estimate calculated from the best GRRF scenario i.e. $C8_{selected}$ among the cropping patterns revealed that mixed crop avocado had the greatest acreage (13,940.18 ha) while the lowest acreage in the cropping patterns was the monocrop avocado (1,929.92 ha) class. In non-croplands, shrubland had the greatest acreage (21,594.77 ha) while water had the lowest acreage (1,167.00 ha). The acreages of other cropping patterns and non-croplands are summarized in Figure 8.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0517.v1

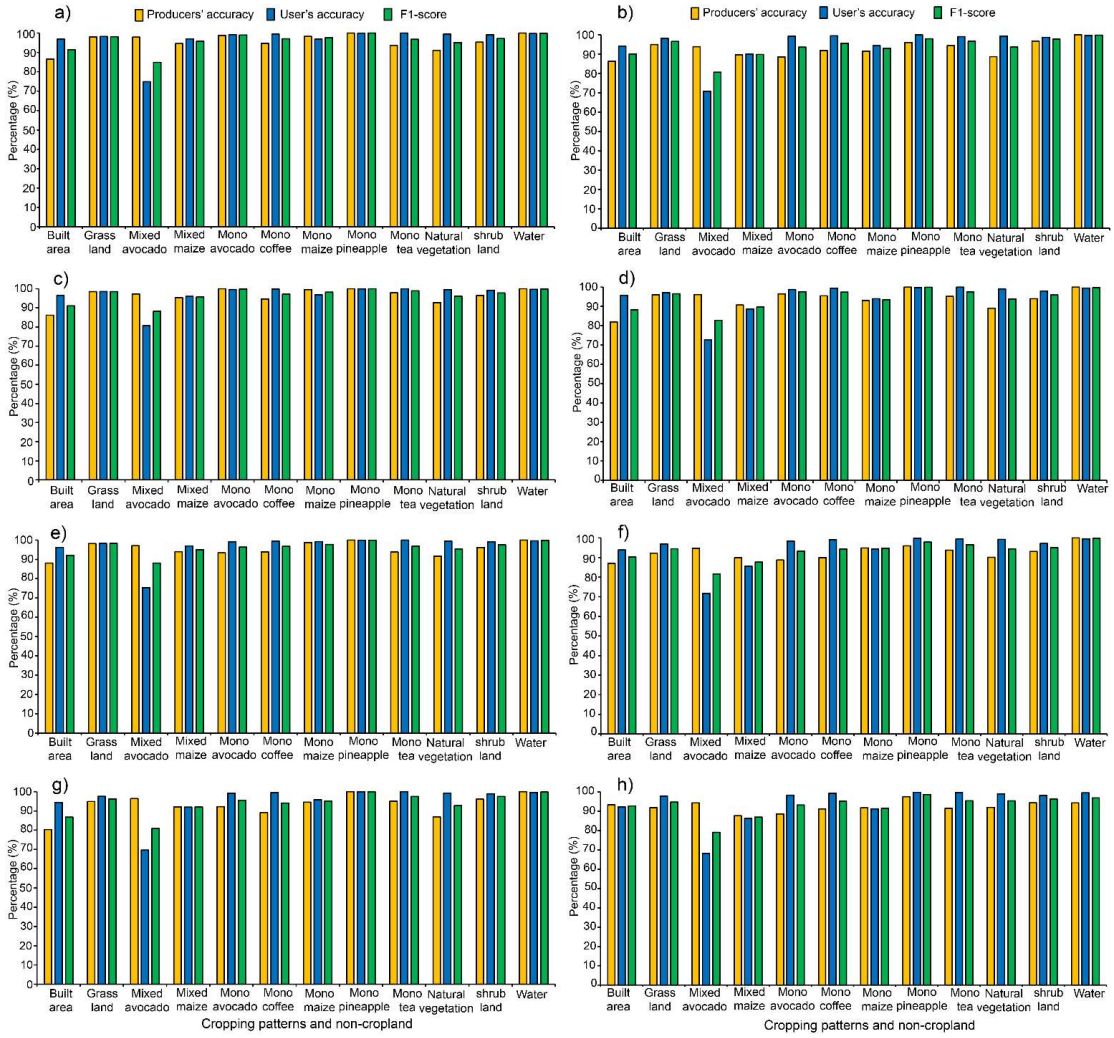

Figure 6. Summarized class-wise accuracies i.e. producer's accuracy (PA), user's accuracy (UA), and F1-score of all combined variables and their corresponding guided regularized random forest (GRRF) selected scenarios. **(a)** Combined Sentinel-2 (S2), vegetation indices (VIs), vegetation phenology (VP) and Sentinel-1 radar backscatter data (S1); **(b)** GRRF selected S2, VIs, VP and S1; **(c)** combined S2 and S1; **(d)** GRRF selected S2 and S1; **(e)** combined S2, VIs, and S1; **(f)** GRRF selected S2, VIs, and S1; **(g)** combined S2, VP and VIs, and **(h)** GRRF selected S2, VP, and VIs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0517.v1

Figure 7. Summarized class-wise accuracies i.e. producer's accuracy (PA), user's accuracy (UA), and F1-score of all combined variables and their corresponding guided regularized random forest (GRRF) selected scenarios i.e. **(a)** combined Sentinel-2 (S2), vegetation indices (VIs); **(b)** GRRF selected S2 and VIs; **(c)** combined S2 and vegetation phenology (VP); **(d)** GRRF selected S2 and VP; **(e)** combined S2; **(f)** GRRF selected S2; **(g)** combined S2, VP, and Sentinel-1 radar backscatter data (S1) and **(h)** GRRF selected S2, VP, and S1.

Table 8. McNemar's test for comparing the performance of all the classification scenarios (C1-C8) for mapping the studied cropping patterns. The level of significance (*p*-value) is set at ≤ 0.05.

Comparison	*p*-value	Chi-square
C1 vs $C1_{selected}$	0.969	0.015
C2 vs $C2_{selected}$	0.967	0.017
C3 vs $C3_{selected}$	0.969	0.001
C4 vs $C4_{selected}$	0.974	0.001
C5 vs $C5_{selected}$	0.967	0.001
C6 vs $C6_{selected}$	0.970	0.001
C7 vs $C7_{selected}$	0.972	0.001
C8 vs $C8_{selected}$	0.974	0.001

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2020 doi:10.20944/preprints202010.0517.v1

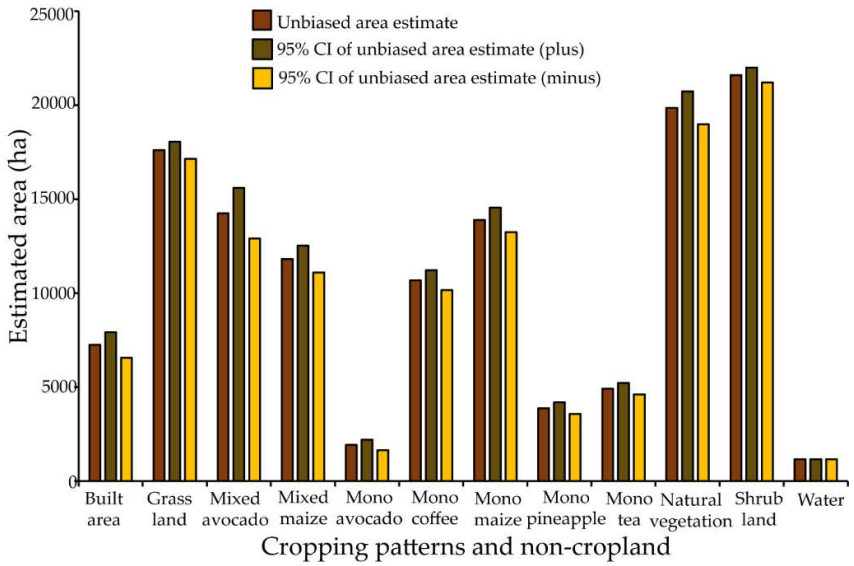

Figure 8. The unbiased area estimates and approximate 95% confidence interval (CI) of the cropping patterns and non-croplands obtained using the area under class method of the most accurate variables combination result from scenario $C8_{selected}$ i.e. guided regularized random forest (GRRF) selected variables of Sentinel-2 (S2), vegetation indices (VIs), vegetation phenology (VP) and Sentinel-1 radar backscatter data (S1).

5. Discussion

This study illustrated the utility of multi-date and multi-sensor variables for classifying cropping patterns for a heterogeneous agro-natural landscape in Kenya. We tested the performance of eight classification scenarios that resulted from combining different vegetation indices and vegetation phenological variables from S2 spectral and S1 SAR backscatter imagery. Seven cropping patterns were classified. The specific combination of the selected variables in each scenario using the GRRF algorithm was also analyzed. We utilized satellite-based data acquired over four seasons (i.e. short and long dry as well as short and long wet season) that were assumed to have captured all the changes in inter- and intra-vegetation dynamics. The use of optical satellite time-series data increased the chances of acquiring cloud-free imagery [64, 65]. Persistent cloud coverage can considerably affect the quality of optical S2 imagery [66]. S2 and S1 have varied revisit periods at 5 and 12 days, respectively, and these differences were accounted using seasonal median composites.

In general, the overall accuracies achieved were above 90% for all scenarios (combined and selected variables using GRRF). However, the higher accuracies of all combined scenarios (C1-C8) have been treated with caution because of the effects of multicollinearity that can lead to overfitting [58]. We thus placed more emphasis on the performance of the GRRF selected scenarios ($C1_{selected}$-$C8_{selected}$) which have the capabilities of reducing multidimensionality and the expected multicollinearity by maintaining the most relevant variables for the analysis. Therefore, the most important variables selected across the eight combined scenarios were S2 bands of NNIR, SWIR3, RE1, Red, Green and Blue; VI bands of NDWI and GNDVI; VP bands of Offset_value, TINDVIBeforeMax and TINDVIAfterMax; and S1 bands of VV and VH. The selection of the S2 bands could be explained by the narrowness of the NIR waveband region at 865 nm. The narrow NIR spectral region is known to be less contaminated by water vapor and represents the NIR plateau for vegetation while also being sensitive to some soil chemical properties [41]. Moreover, the reflectance in the SWIR3 band is sensitive to the leaf internal structure [68] whilst the red edge band (RE1), in addition to visible bands of red, green, and blue, are also sensitive to vegetation chlorophyll and other biochemical contents [69-71].

Furthermore, the VIs were also useful in specific crop type identification [72] by measuring the photosynthetic size of specific plant canopies that could improve the individual cropping patterns

classification [73]. The selection of the VIs i.e., NDWI and GNDVI can be attributed to the sensitivity of the NIR band in the NDWI to leaf internal structure and leaf dry matter content while the SWIR3 band in the index is sensitive to the vegetation water content and the spongy mesophyll structure in vegetation canopies reflecting biochemical metrics of vegetation [47]. On the other hand, the GNDVI is more sensitive to the chlorophyll content of the plant since it is composed of the green channel instead of the red band [50]. In addition, time-series phenology variables are found to be the best in differentiating temporal and spectral variability of crop growth [74]. The selection of the VP variables of Offset_value, TINDVIBeforeMax, and TINDVIAfterMax could be explained by the sensitivity of Offset_value to land use and land cover differences, while TINDVIBeforeMax and TINDVIAfterMax describe the pre- and post-anthesis stages, respectively which can differ among different plants [53].

Regarding S1 radar backscatter data, the inter-channel phase information enabled by dual-polarization, i.e. VV and VH, allowed the enhanced analysis of the backscattering properties of the target areas of our study hence improved the classification [75]. Nonetheless, S1 data may be sensitive to soil moisture in the background of sparsely vegetated areas [76]. The synergetic importance of combining the S2 bands, S2 derived VIs, VP and S1 radar backscatter data was also demonstrated in this study in that the GRRF selected bands of scenario S2, VIs, VP, and S1 had the highest overall accuracy (OA=93.16%) compared to the rest of the GRRF selected band scenarios. The GRRF selected bands in S2, VIs, VP and S1 scenario were TINDVIAfterMax, NDWI, SWIR3, NNIR, RE1, Red, and Green. S2 bands' dominance confirmed the strength and importance of the raw spectral bands of S2 in discriminating vegetation [77].

We mapped cropping patterns that included monocrop avocado, mixed crop avocado, monocrop maize, mixed crop maize, monocrop pineapple, and monocrop coffee, grown in a very complex, heterogeneous, and dynamic agro-natural setup in Murang'a County, Kenya. In the field, we observed that most of the mixed fields were small-scale avocado and small-scale maize fields that are mixed with other crops like the common bean, banana, and macadamia. We speculate that the farmers intend to maximize the profit of their land by mixing the crops in the same piece of the land, particularly with the uncertainty in rainfall trend, and invasion of insect pests like the fall armyworm that could damage their maize crop [78]. High class-wise UA, PA, and F1-scores were observed in the mapping of monocropping patterns of avocado, coffee, tea, and pineapple. This could be due to the high spectral uniformity of the monocropping patterns, which also resulted in less intra class variability. However, mixed crop avocado and mixed crop maize classes had F1-scores below 90.00% compared to other cropping pattern classes in our study. This can be explained by high inter-class variability associated with the different vegetation composition including non-croplands such as natural vegetation, grasslands, and shrublands, hence a high rate of misclassification within the farms [23]. Nonetheless, the mixed crop avocado class had the greatest acreage (13,940.18 ha) compared to other crops, which could be attributed to the popularity of avocado farming by small-scale farmers in Murang'a County due to its increasing export value in Kenya. Furthermore, we assessed the performance of our classification experiments using an error matrix that takes into account the estimated area proportion of each class [61]. This is a robust and reliable classification accuracy assessment method compared to the traditional classification confusion matrix that considers the number of mapped instances of each class. Understanding the variability of cropping patterns and non-croplands can also contribute to the sustainable management of crop insect pests and pollinators that can be influenced by landscape structure [40].

Although non-croplands such as grassland, shrubland, natural vegetation, and water were not the focus of this present study, it is worth noting that relatively high UA, PA, and F1-scores were reported for mapping these classes across all (combined and GRRF) classification scenarios. The insignificant statistical differences from the McNemar's test between the combined scenarios and the GRRF selected variables could be explained by the use of related samples (i.e. the same training data used in the classification of all the scenarios) [63]. **T**he use of the same training data in all the classification scenarios was necessary in this study to enable the unbiased comparison of the performance of the classification scenarios. Also, the insignificant differences among the performance of our classification scenarios indicate the novelty and robustness of our variable selection

 doi:10.20944/preprints202010.0517.v1

experiment using GRRF, in such a way that a few selected numbers of variables performance were not statistically different to all combined variables.

The mapping of cropping patterns is of profound importance in understanding the sustainability of food systems and how they are affected by climate and also for modeling the abundance and spread of crop insect pests and disease and pollinators [19, 79]. Furthermore, cropping patterns are valuable parameters in many crop modeling frameworks that estimate crop production as a function of many biotic and abiotic factors [80].

6. Conclusions

This study investigated the contribution of time-series S2 reflectance bands, its derivatives of vegetation indices and phenology metrics, and S1 radar backscatter data in the classification of cropping patterns in three sub-counties of Murang'a County, Kenya using eight scenarios and GRRF machine learning algorithm. The GRRF selected scenarios that consisted of S2 bands, their derivatives of vegetation indices and phenology metrics, and S1 radar backscatter data accurately (OA > 93.15%) mapped cropping pattern in a complex dynamic heterogeneous landscape. Our study also provided information on cropping pattern acreage that is necessary for crop management, and different other crop produce harvesting and marketing plans. Consequently, our approach can be used to map cropping patterns in other sites of similar agroecological conditions when high resolution imagery are not readily available.

Author Contributions: Conceptualization, E.M.A.-R., T.L., and T.D.; methodology, G.R.A., E.M.A.-R., A.W.S., G.O.M., B.T.M., T.L., H.E.Z.T., and T.D.; software, G.R.A. and B.T.M.; validation, G.R.A., E.M.A.-R., A.W.S., G.O.M., B.T.M., T.L, H.E.Z.T., and T.D.; formal analysis, G.R.A., and B.T.M.; investigation, E.M.A.-R., G.R.A., B.T.M., A.W.S., and G.O.M.; resources, T.D.; data curation, T.L., and G.R.A.; writing—original draft preparation, G.R.A.; writing—review and editing, G.R.A., E.M.A.-R., A.W.S., G.O.M., B.T.M., T.L., H.E.Z.T., and T.D.; visualization, G.R.A., E.M.A.-R., B.T.M., A.W.S., and G.O.M.; supervision, E.M.A.-R., A.W.S., and G.O.M.; project administration, T.D.; funding acquisition, T.D. All authors have read and agreed to the content of the manuscript.

Funding: This research was funded by the Federal Ministry for Economic Cooperation and Development (BMZ), through the project "Integrated pest and pollinator management (IPPM) in avocado-cucurbits production systems in Kenya and Tanzania", grant number 17.7860.4-001.00; UK's Foreign, Commonwealth & Development Office (FCDO); the Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); the Federal Democratic Republic of Ethiopia; and the Government of the Republic of Kenya. The views expressed herein do not necessarily reflect the official opinion of the donors. The views expressed herein do not necessarily reflect the official opinion of the donors.

Acknowledgments: We thank the avocado farmers in Murang'a for their cooperation and for providing us with the information needed for this project. Much appreciation also goes to Stella Muthoni Gachoki, Eunice King'ori, Marian Salim Adam, and Ephantus Kimani for their help in the field data collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. FAO. FAO and Traditional Knowledge: The Linkages with Sustainability, Food Security and Climate Change Impacts Rome, Italy 2009.
2. Husain, M. Systematic Agricultural Geography. Reprinted 2004, Rawat Publication, Jaipur and New Delhi, 1996, pp. 217, 218.
3. Jayne, T.S.; Yamano, T.; Weber, M.T.; Tschirley, D.; Benfica, R.; Chapoto, A.; Zulu, B. Smallholder income and land distribution in Africa: Implications for poverty reduction strategies. *Food Policy*, 2003, *28*(3), 253–275. https://doi.org/10.1016/S0306-9192(03)00046-0
4. Bégué, A.; Arvor, D.; Bellón, B.; Betbeder, J.; Abelleyra, D. de.; Ferraz, R.P.D.; Lebourgeois, V.; Lelong, C.; Simões, M.; Verón, S.R. Remote Sensing and Cropping Practices: A Review. *Remote Sens*, 2018, *10*(1), 99. https://doi.org/10.3390/rs10010099
5. Ansari, N.A. Dynamic cropping pattern within the last two decades: A case study of Gautam Buddh Nagar District, National Capital Region, India. 2015 Apr 3.

6. Lithourgidis, A.S.; Dordas, C.; Damalas, C.; Vlachostergios, D.N. Annual intercrops: An alternative pathway for sustainable agriculture. *Austral. J. Crop Sci.* 2011, *15*, 396-410.
7. Ekroth, A.K.E.; Rafaluk-Mohr, C.; King, K.C. Diversity and disease: Evidence for the monoculture effect beyond agricultural systems. *BioRxiv*, 668228. https://doi.org/10.1101/668228
8. Jalilian, J.; Najafabadi, A.; Zardashti, M.R. Intercropping patterns and different farming systems affect the yield and yield components of safflower and bitter vetch. *J. Plant Interact.* 2017, *12*(1), 92–99. https://doi.org/10.1080/17429145.2017.1294712
9. Nellis, M.D.; Price, K.P.; Rundquist. D. Remote sensing of cropland agriculture. In T. A. Warner, M. D. Nellis, & G. M. Foody, The SAGE Handbook of Remote Sensing (pp. 368–380). SAGE Publications, Inc. https://doi.org/10.4135/9780857021052.n26.
10. Xiong. J.; Thenkabail, P.S.; Gumma, M.K.; Teluguntla, P.; Poehnelt, J.; Congalton, R.G.; Yadav, K.; Thau, D. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS J. Photogramm. Remote Sens*, *126*, 225–244. https://doi.org/10.1016/j.isprsjprs.2017.01.019
11. Hao, P.; Löw, F.; Biradar, C. Annual Cropland Mapping Using Reference Landsat Time Series—A Case Study in Central Asia. *Remote Sens.* 2018, *10*(12), 2057. https://doi.org/10.3390/rs10122057
12. Tomppo, E; Antropov, O; Praks, J. Cropland classification using sentinel-1 time series: Methodological performance and prediction uncertainty assessment. *Remote Sens.* 2019, *11*(21), 2480. https://doi.org/10.3390/rs11212480
13. Forkuor, G.; Conrad,C.; Thiel, M.; Ullmann, T.; Zoungrana, E. Integration of Optical and Synthetic Aperture Radar Imagery for Improving Crop Mapping in Northwestern Benin, West Africa. *Remote Sens*. 2014, 6, 6472-6499. https://doi:10.3390/rs6076472
14. Heupel, K.; Spengler, D.; Itzerott, S. A Progressive Crop-Type Classification Using Multitemporal Remote Sensing Data and Phenological Information. *PFG – J. Photogramm. Remote Sens. Geoinformation Sci.* 2018, *86*(2), 53–69. https://doi.org/10.1007/s41064-018-0050-7
15. Brinkhoff, J.; Vardanega, J.; Robson, A.J. Land Cover Classification of Nine Perennial Crops Using Sentinel-1 and -2 Data. *Remote Sens.* 2020, *12*(1), 96. https://doi.org/10.3390/rs12010096
16. Mondal,S.; Jeganathan, C.; Sinha, N.K.; Rajan, H.; Roy, T.; Kumar, P. Extracting seasonal cropping patterns using multi-temporal vegetation indices from IRS LISS-III data in Muzaffarpur District of Bihar, India. *Egypt J. Remote Sens. Space Sci.* 2014, *17*(2), 123–134. https://doi.org/10.1016/j.ejrs.2014.09.002
17. Waldhoff, G.; Lussem, U.; Bareth, G. Multi-data approach for remote sensing-based regional crop rotation mapping: A case study for the Rur catchment, Germany. *Int. J Appl. Earth Obs. Geoinf.* 2017, *61*, 55–69. https://doi.org/10.1016/j.jag.2017.04.009
18. Chen, Y.; Lu, D.; Moran, E.; Batistella, M.; Dutra, L.V.; Sanches, I.D.; Silva, B.F.R.; Huang, J.; Luiz, J.B.A.; Oliveira, A.F.M. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *Int. J. Appl. Earth Obs. Geoinf.* 2018. https://doi.org/10.1016/j.jag.2018.03.005
19. Kyalo, R.; Abdel-Rahman, E.M.; Subramanian, S.; Nyasani, J.O; Thiel, M.; Jozani, H.; Borgemeister, C.; Landmann, T. Maize Cropping Systems Mapping Using RapidEye Observations in Agro-Ecological Landscapes in Kenya. *Sensors* 2017. https://doi.org/10.3390/s17112537
20. Singha, M.; Wu, B.; Zhang, M. An Object-Based Paddy Rice Classification Using Multi-Spectral Data and Crop Phenology in Assam, Northeast India. *Remote Sens.* 2016, *8*(6), 479. https://doi.org/10.3390/rs8060479.
21. Belgiu, M.; Csillik, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* 2018, *204*, 509–523. https://doi.org/10.1016/j.rse.2017.10.005
22. Nidamanuri, R.R.; Zbell, B. Use of field reflectance data for crop mapping using airborne hyperspectral image. *ISPRS J. Photogramm. Remote Sens.* 2011, *66*(5), 683–691. https://doi.org/10.1016/j.isprsjprs.2011.05.001
23. Boitt, M.; Ndegwa, C.; Pellikka, P. Using Hyperspectral data to identify crops in a cultivated agricultural landscape - A Case Study of Taita Hills, Kenya. *J. Earth Sci. Clim. Change* 2014, 05(09). https://doi.org/10.4172/2157-7617.1000232
24. Kenduiywo, B.K.; Bargiel, D.; Soergel, U. Crop-type mapping from a sequence of Sentinel 1 images. *Int. J. Remote Sens.* 2018, *39*(19), 6383–6404. https://doi.org/10.1080/01431161.2018.1460503
25. Useya, J.; Chen, S. Exploring the Potential of Mapping Cropping Patterns on Smallholder Scale Croplands Using Sentinel-1 SAR Data. *Chin. Geogr. Sci.* 2019, *29*(4), 626–639. https://doi.org/10.1007/s11769-019-1060-0
26. Kobayashi, N.; Tani, H.; Wang, X.; Sonobe, R. Crop classification using spectral indices derived from Sentinel-2A imagery. *J Inf. Telecommun*. 2019, *0*(0), 1–24. https://doi.org/10.1080/24751839.2019.1694765

27. Zhong, L.; Hawkins, T.; Biging, G.; Gong, P. A phenology-based approach to map crop types in the San Joaquin Valley, California. *Int. J. Remote Sens.* 2011, *32*(22), 7777–7804. https://doi.org/10.1080/01431161.2010.527397
28. Liu, J.; Zhu, W.; Atzberger, C.; Zhao, A.; Pan, Y.; Huang, X. A Phenology-Based Method to Map Cropping Patterns under a Wheat-Maize Rotation Using Remotely Sensed Time-Series Data. *Remote Sens*. 2018, *10*(8), 1203. https://doi.org/10.3390/rs10081203
29. Fritz, S.; See, L.; Rembold, F. Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. *Int. J. Remote Sens.* 2010, *31*(9), 2237–2256. https://doi.org/10.1080/01431160902946598
30. Petitjean, F.; Inglada, J.; Gancarski, P. Satellite Image Time Series Analysis Under Time Warping. *IEEE Trans. Geosci. Remote Sens.* 2012, *50*(8), 3081–3095. https://doi.org/10.1109/TGRS.2011.2179050
31. Duncan, J.M.A.; Dash, J.; Atkinson, P.M. The potential of satellite-observed crop phenology to enhance yield gap assessments in smallholder landscapes. *Front Environ. Sci.* 2015, *3*. https://doi.org/10.3389/fenvs.2015.00056
32. Song, Q.; Hu, Q.; Zhou, Q.; Hovis, C.; Xiang, M.; Tang, H.; Wu, W. In-Season Crop Mapping with GF-1/WFV Data by Combining Object-Based Image Analysis and Random Forest. *Remote Sens.* 2017, *9*(11), 1184. https://doi.org/10.3390/rs9111184
33. Deng, H.; Runger, G. Gene selection with guided regularized random forest. *Pattern Recognit.* 2013, *46*(12), 3483–3489. https://doi.org/10.1016/j.patcog.2013.05.018
34. Mureriwa, N.; Adam, E.; Sahu, A.; Tesfamichael, S. Examining the spectral separability of prosopis glandulosa from co-existent species using field spectral measurement and guided regularized random forest. *Remote Sens.* 2016, *8*(2). https://doi.org/10.3390/rs8020144
35. Mudereri, B.T.; Dube, T.; Niassy, S.; Kimathi. E.; Landmann, T.; Khan, Z.; Abdel-Rahman, E.M. 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.* 2020, *85*, 102008. https://doi.org/10.1016/j.jag.2019.102008
36. Murang'a County. Available online: https://muranga.go.ke/?page_id=3130 (accessed 20 May 2019).
37. Ovuka, M.; Lindqvist, S. Rainfall Variability in Murang'a District, Kenya: Meteorological Data and Farmers' Perception. *Geogr. Ann. Ser. A, Phys. Geogr.* 2000, *82*(1), 107–119.
38. Amare, M.; Mariara, J.; Oostendorp, R.; Pradhan, M. The impact of smallholder farmers' participation in avocado export markets on the labor market, farm yields, sales prices, and incomes in Kenya. *Land Use Policy* 2019, *88*, 104168. https://doi.org/10.1016/j.landusepol.2019.104168
39. Johnny, E.G.; Kabubo-Mariari, J.; Mulwa, R.; Ruigu, G.M. Smallholder Avocado Contract Farming in Kenya: Determinants and Differentials in Outcomes. *Afr. J. Econ. Rev.* 2019, *07*(2). https://doi.org/10.22004/ag.econ.292365
40. Toukem, N.K.; Yusuf, A.A.; Dubois, T.; Abdel-Rahman, E.M.; Adan, M.S.; Mohamed, S.A. Landscape Vegetation Productivity Influences Population Dynamics of Key Pests in Small Avocado Farms in Kenya. *Insects* 2020, *11*(7), 424 https://doi.org/10.3390/insects11070424
41. European Space Agency. Available online: https://www.esa.int/ESA (accessed 20 June 2019).
42. Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *J. Sens.* 2017. https://doi.org/10.1155/2017/1353691
43. Coburn, C.A.; Gaalen, E.V.; Peddle, D.R.; Flanagan, L.B. Anisotropic reflectance effects on spectral indices for estimating ecophysiological parameters using a portable goniometer system. *Can. J. Remote Sens.* 2010, *36*(sup2), S355–S364. https://doi.org/10.5589/m10-066
44. Tucker, C.J.; Elgin, J.H.; McMurtrey, J.E.; Fan, C.J. Monitoring corn and soybean crop development with hand-held radiometer spectral data. *Remote Sens. Environ*. 1979 *8*(3), 237–248. https://doi.org/10.1016/0034-4257(79)90004-X
45. Ahamed, T.; Tian, L.; Zhang, Y.; Ting, K.C. A review of remote sensing methods for biomass feedstock production. *Biomass Bioenergy*. 2011, *35*(7), 2455–2469.
46. Jiang, Z.; Huete, A.R.; Kim, Y.; Didan, K. 2-band enhanced vegetation index without a blue band and its application to AVHRR data. (W. Gao & S. L. Ustin, Eds.; p. 667905). https://doi.org/10.1117/12.734933
47. Gao, B. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ*. 1996, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3
48. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ*. 1994, *48*(2), 119–126.

49. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 1988, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X
50. Gitelson, A.A.; Merzlyak, M.N. Remote sensing of chlorophyll concentration in higher plant leaves. *Adv. Space. Res.* 1998, *22*(5), 689–692. https://doi.org/10.1016/S0273-1177(97)01133-2
51. Kaufman, Y.J.; Tanre, D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Trans. Geosci. Remote Sens.* 1992, *30*(2), 261–270. https://doi.org/10.1109/36.134076
52. Kimball, J. Vegetation Phenology. In: Njoku EG, editor. Encyclopedia of Remote Sensing. New York, NY: *Springer* 2014. https://doi.org/10.1007/978-0-387-36699-9_188
53. Araya, S. Multi-temporal Remote sensing for the Estimation of Plant Available Water-holding Capacity of soil. Phd Thesis, University of Adelaide, Australia, 2017.
54. Filipponi, F. Sentinel-1 GRD Preprocessing Workflow. In Proceedings of the 3rd International Electronic Conference on Remote Sensing, 22 May–5 June 2019.
55. Likas, A.; Vlassis, N.; Verbeek, J.J. The global k-means clustering algorithm. *Pattern Recognit.* 2003, *36*(2), 451–461. https://doi.org/10.1016/S0031-3203(02)00060-2
56. Google Earth. Available online: https://earth.google.com/web (accessed on 26 June 2020).
57. Breiman, L. Random Forests. *Mach. Learn.* 2001, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
58. Izquierdo-Verdiguier, E.; Zurita-Milla, R. An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* 2020, *88*, 102051. https://doi.org/10.1016/j.jag.2020.102051
59. R Core Team. R: *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing, 2019.
60. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1995. p. 1137–1143.
61. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* 2014, *148*, 42–57. https://doi.org/10.1016/j.rse.2014.02.015
62. Chinchor, N. MUC-4 Evaluation Metrics. In Proceedings of the Fourth Message Understanding Conference (MUC-4), McLean, Virginia, June 16-18, 1992.
63. Foody, G. Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogramm. Eng. Remote Sens.* 2004, *70*, 627–633. https://doi.org/10.14358/PERS.70.5.627
64. Mtibaa, S.; Irie, M. Land cover mapping in cropland dominated area using information on vegetation phenology and multi-seasonal Landsat 8 images. *Euro-Mediterr. J. Environ. Integr.* 2016, 1(1), 6. https://doi.org/10.1007/s41207-016-0006-5
65. Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W-T. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* 2018. https://doi.org/10.1016/j.jag.2018.06.007
66. Eberhardt, I.D.R.; Schultz, B.; Rizzi, R.; Sanches, I.D.; Formaggio, A.R.; Atzberger, C.; Mello, M.P.; Immitzer, M.; Trabaquini, K.; Foschiera, W.; José Barreto Luiz, A. Cloud Cover Assessment for Operational Crop Monitoring Systems in Tropical Areas. *Remote Sens.* 2016 *8*(3), 219. https://doi.org/10.3390/rs8030219
67. Rui, M.; Ma, C.; Hao, Y.; Guo, J.; Rui, Y.; Tang, X.; Zhao, Q.; Fan, X.; Zhang, Z.; Hou, T.; Zhu, S. Iron Oxide Nanoparticles as a Potential Iron Fertilizer for Peanut (Arachis hypogaea). *Front. Plant Sci.* 2016, *7*. https://doi.org/10.3389/fpls.2016.00815
68. Deepak, M.; Keski-Saari, S.; Fauch, L.; Granlund, L.; Oksanen, E.; Keinänen, M. Leaf Canopy Layers Affect Spectral Reflectance in Silver Birch. *Remote Sens.* 2019, *11*(24), 2884. https://doi.org/10.3390/rs11242884
69. Immitzer, M.; Vuolo, F.; Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.* 2016 *8*(3), 166. https://doi.org/10.3390/rs8030166
70. Zabala Ramos, S. Comparison of multi-temporal and multispectral Sentinel-2 and Unmanned Aerial Vehicle imagery for crop type mapping. Lund University GEM Thesis Ser. 2017.
71. Chuanliang, S.; Bian, Y.; Zhou, T.; Pan, J. Using of Multi-Source and Multi-Temporal Remote Sensing Data Improves Crop-Type Mapping in the Subtropical Agriculture Region. *Sensors* 2019, *19*, 2401. https://doi.org/10.3390/s19102401

72. Sonobe, R.; Yamaya, Y.; Tani, H.; Wang, X.; Kobayashi, N.; Mochizuki, K. Crop classification from Sentinel-2-derived vegetation indices using ensemble learning. *J. Appl. Remote Sens*. 2018, *12*(2), 026019. https://doi.org/10.1117/1.JRS.12.026019
73. Wiegand, C.L.; Richardson, A.J.; Escobar, D.E.; Gerbermann, A.H. Vegetation indices in crop assessments. *Remote Sens. Environ.* 1991 *35*(2), 105–119. https://doi.org/10.1016/0034-4257(91)90004-P
74. Ghazaryan, G.; Dubovyk, O.; Löw, F.; Lavreniuk, M.; Kolotii, A.; Schellberg, J; Kussul, N. A rule-based approach for crop identification using multi-temporal and multi-sensor phenological metrics. *Eur. J. Remote Sens.* 2018 *51*, 511–524. https://doi.org/10.1080/22797254.2018.1455540
75. Van Tricht, K.; Gobin, A.; Gilliams, S.; Piccard, I. Synergistic use of radar Sentinel-1 and optical Sentinel-2 imagery for crop mapping: a case study for Belgium. *Remote Sens.* 2018. https://doi.org/10.20944/preprints201808.0066.v1
76. Gao, Q.; Zribi, M.; Escorihuela, M. J.; Baghdadi, N.; Synergetic Use of Sentinel-1 and Sentinel-2 Data for Soil Moisture Mapping at 100 m Resolution. *Sensors*2017), 17(9). https://doi.org/10.3390/s17091966
77. Mudereri, B.T.; Dube, T.; Abdel-Rahman, E.M.; Niassy, S.; Kimathi, E.; Khan, Z.; Landmann, T. A comparative analysis of Planetscope and Sentinel-2 space-borne sensors in mapping striga weed using guided regularised random forest classification ensemble. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2019 *XLII-2/W13*, 701–708. https://doi.org/10.5194/isprs-archives-XLII-2-W13-701-2019
78. De Groote, H.; Kimenju, S.C.; Munyua, B.; Palmas, S.; Kassie, M.; Bruce, A. Spread and impact of fall armyworm (Spodoptera frugiperda J.E. Smith) in maize production areas of Kenya. *Agric. Ecosyst. Environ.* 2020, *292*, 106804. https://doi.org/10.1016/j.agee.2019.106804
79. Panigrahy, S.; Ray, S.S.; Manjunath, K.R.; Pandey, P.S.; Sharma, S.K.; Sood, A.; Yadav, M.; Gupta, P. C.; Kundu, N.; Parihar, J.S. A Spatial Database of Cropping System and its Characteristics to Aid Climate Change Impact Assessment Studies. *J. Indian Soc. Remote Sens.* 2011, *39*(3), 355–364. https://doi.org/10.1007/s12524-011-0093-3
80. Dury, J.; Schaller, N.; Garcia, F.; Reynaud, A.; Bergez, J.E. Models to support cropping plan and crop rotation decisions. A review. *Agron. Sustain. Dev.* 2012, *32*(2), 567–580. https://doi.org/10.1007/s13593-011-0037-x