Article

Delineation of Cocoa Agroforests Using Multi-Season Sentinel-1 SAR Images: Low Grey Level Range Reduces Uncertainties in GLCM Texture-Based Mapping

Frederick N. Numbisi [1,2,*], Frieke M. B. Van Coillie [1] and Robert De Wulf [1]

[1] Ghent University, Laboratory of Forest Management and Spatial Information Techniques, Coupure Links 653, 9000 Gent, Belgium; (Frederick.NkeumoeNumbisi, Frieke.Vancoillie, Robert.DeWulf) @UGent.be
[2] World Agroforestry Centre (ICRAF), West and Central Africa Regional Programme, P. O. Box 16317, Yaoundé, Cameroon; f.nkeumoe@cgiar.org
* Correspondence: frednumbisi@gmail.com; Tel.: +49-178-310-7190

Abstract: Delineating the cropping area of cocoa agroforests is a major challenge for quantifying the contribution of the land use expansion to tropical deforestation. Discriminating cocoa agroforests from tropical transition forests using multi-spectral optical images is difficult due to a similarity in the spectral characteristics of their canopy; moreover, optical sensors are largely impeded by the frequent cloud cover in the tropics. This study explores multi-season Sentinel-1 C-band SAR image to discriminate cocoa agroforests from transition forests for a heterogeneous landscape in central Cameroon. We use an ensemble classifier, random forest, to average SAR image texture features of GLCM (Grey Level Co-occurrence Matrix) across seasons; next, we compare classification performance with results from RapidEye optical data. Moreover, we assess the performance of GLCM texture feature extraction at four different grey level quantization: 32bits, 8bits, 6bits, and 4bits. The classification overall accuracy (OA) of texture-based maps outperformed that from an optical image; the highest OA of 88.8% was recorded at 6bits grey level. This quantization level, in comparison to the initial 32bits in SAR images, reduced the class prediction error by 2.9%. Although this prediction gain may be large for the landscape area, the resultant thematic map reveals the decrease and fragmentation of forest cover by cocoa agroforests. According to our classification validation, the Shannon entropy (H) or uncertainty provides a reliable validation for class predictions and reveals detail inference for discriminating inherently heterogeneous vegetation categories. The texture-based classification achieved a reliable accuracy considering the heterogeneity of the landscape and vegetation classes.

Keywords: mapping cocoa agroforests; Congo Basin rainforest; sentinel-1; SAR; GLCM textures; grey level quantization; random forest algorithm; machine learning; classification uncertainty

Conference paper:
Numbisi, F. N.; Van Coillie, F.; De Wulf, R. Multi-Date Sentinel1 Sar Image Textures Discriminate Perennial Agroforests in a Tropical Forest-Savannah Transition Landscape. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-1*, 339–346, doi:10.5194/isprs-archives-XLII-1-339-2018.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

1. Introduction

The mapping of cocoa commodity cropland is essential to quantify its ecosystem services, and as well the disservices related tropical forest cover loss. Agricultural land expansions, predominantly for oil palm, rubber, and cocoa plantations, contribute significantly to tropical deforestation [1–3]. Moreover, these commodity cropping lands, amongst others, provide different ecological services in terms of carbon sequestration, habitat provision, conservation of biodiversity [4,5]. Thus, a reliable and recurrent mapping of such cropping area is crucial for customizing forest landscape management to the respective land use expansion.

Agroforestry has been suggested as an agricultural option for sustainable cocoa production; Cocoa Agroforestry refers to the system of growing cocoa tree crop in the understorey of multi-strata canopy trees [6], which comprise a diversity of timber, fruits, and NTFP (Non-Timber Forest Product) producing tree species [5,7,8]. Cocoa is a tree crop of high economic importance, and particularly in tropical sub-Saharan Africa [2,9] that contributes about 70% of global cocoa dry beans export [10]. Regrettably, the expansion of cocoa production lands contributes significantly to the loss of forest cover [11,12]. Such expansions are somewhat specific to countries and production landscapes [13–16]; therefore, some are more destructive to forests than others. On a global scale, cocoa production was responsible for 57% of the global agricultural land expansion rate of 132,000ha y^{-1} in the period 2000–2013 [17]. However, such figures need validation, at national levels, through mapping of actual cropping lands.

From an ecological standpoint, compared to the intensive mono-stratum cocoa plantations and other high canopy commodity crops such as oil palm, rubber, etc., cocoa agroforests sustain ecosystem services at a scale that is considered second to transition forests [18–20]. Regarding management in most cocoa-producing nations, available literature does rarely address spatial mapping of cocoa production area. Management projections on production area are based on the FAO's (Food and Agricultural Organization) database on crops, FAOSTAT. This database depends on sporadic Country reports of annual harvest area, and projection from these reports, which may not represent the actual ground reality [21] In the case of Cameroon, cocoa is predominantly grown in small-scale agroforests of 1-3ha [5,7,8]. Thus, the cocoa area of 123,120 ha in Cameroon's Centre Region production hotspot for example [22], is in effect the harvested area, which is based on seasonal records by local farmers' and Cocoa cooperatives. Thence, the national statistics is monitored and published by the National Cocoa and Coffee Board (NCCB). Depending on the variety and propagation technique, established cocoa farms require about 3 or more years of crop tending, before the harvest of first produce [23]. Consequently, the FAO records of the harvested area may be, at the minimum, 3 years short of possible expansions in cocoa farms. On the assumption of continuous expansion of cocoa production land, therefore, the associated impact on forest cover is far greater than management extrapolations made solely from published data on harvested areas. For a sustainable management of cocoa production landscapes, national government programs that stimulate export of dry cocoa beans [2,9,22] need support for reliable and updated estimations of both harvest and expansion areas for cocoa agroforests.

The application of earth observation data provides large-scale mapping of commodity cropping area. Unfortunately, the discrimination of cocoa agroforest areas with multi-strata canopy, using optical reflectance and vegetation indices is yet not successful [17,24]. Cocoa agroforests have similar canopy structure as transition forests [16]. In moist tropical zones, a high frequency of clouds and atmospheric aerosols hampers application of optical satellite data. SAR images on the contrary, provide cloud and season independent information about land surface features. Based on texture information extraction, the analysis of SAR images has been used for discrimination of cropland [25,26] and forest biomass estimation [27]. Unlike optical imagery that captures reflectance of trees and forest canopies, SAR data capture the water content (a dielectric property) and structure (geometric property) of target features. The later information is only provided if the targets' size is lower or close to the wavelength of the SAR sensor. Thus, use of SAR imagery, e.g. for vegetation mapping, is predetermined by sensor wavelength, and necessitates image processing procedures that vary with vegetation type and scale of assessment.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

Long wavelength SAR such as L-band (λ≈25 cm) provide details on volume scattering from branches and stems, which are essential for aboveground biomass estimation [28]. A SAR-based index, Radar Vegetation Index (RVI), was developed for biomass monitoring using L-band data. However, the application of such an index requires removal of the contribution from soil surface backscatter [29]. Although such bands may improve discrimination of vegetation with high tree canopy, they are less reliable for mapping features of low or no vegetation. In the latter scenario, therefore, other SAR wavelengths may be equally reliable.

The C-band sensor is the wavelength (λ≈5 cm) with the largest temporal series in SAR remote sensing. Its utility has been less explored for mapping tropical land cover, and especially so in commodity croplands in the predominantly heterogeneous farming conditions in sub-Saharan Africa. The C-Band SAR penetrates the vegetation canopy only to a limited extent. However, as in the case of settlement and grassland land cover classes, forests have a high temporal stability of SAR backscatter signals. To Thiel et al. [30], the contrast between these land cover classes and agricultural land is high in the cross-polarized (HV or VH) SAR image bands. Besides, Stimulus et al. [31] reported that texture measures are needed to discriminate settlement areas from forests. Thus, considering the seasonal changes in structure and water content of vegetation canopy elements, a temporal metric of texture from C-band SAR images may be able to discriminate perennial agroforestry land cover.

Texture measures from Grey Level Co-occurrence Matrix (GLCM) provide reliable information on the spatial relationship of images pixels [32]. The GLCM provides a joint probability distribution or co-occurrence frequency of grey levels (or intensity tones) in an image based on three parameters: the pixel(s) distance, angular displacement and image sub-region - analysis window size. Several second order ,between two pixels, statistics from the GLCM were proposed [33] to describe the texture in an image. The use of GLCM texture measures depends on the geometry of target features and their characteristic spatial structure in the landscape[32] . For land cover classification in a heterogeneous landscape, Mishra et al. [34] observed that texture information was more valuable to improve classification accuracy in SAR image than for optical image. Yet, the authors [34] suggested that an optimum combination of texture features is needed for the specific type of landscape heterogeneity. Land cover classification using GLCM texture extraction have focused on scale or window size [32,34–37]. However, the importance of grey level quantization in GLCM texture analysis has been emphasized [38–40]. Moreover, for texture feature extraction, use of image grey levels beyond the depth of pixels (range of values) may increase uncertainty in results [41]. Thus, application of GLCM matrix for land use and land cover (LULC) classification [30,32,37] do not take grey level quantization in account, which may be vital in mapping heterogeneous agricultural landscapes - in particular, the inherently heterogeneous LULC categories.

The purpose of this study was to assess the temporal contribution of SAR volume scattering, essentially by vegetation canopy, in discriminating perennial cocoa agroforests land use. We explored multi-seasonal multi-polarization Sentinel-1 C-band SAR images for the following objectives: (1) Evaluate the performance of GLCM texture-based discrimination of cocoa agroforest land use from transition forests cover in comparison to typical classification from multi-spectral optical image – using a RapidEye image; (2) Assess the contribution of grey level quantization in improving texture-based classification performance. We compared four different grey level quantization or dynamic pixel range: 32bits, 8bits, 6bits and 4bits; (3) Assess the information gain from Shannon Entropy (H) or uncertainty as a classifier performance estimator.

We applied the GLCM to estimate four selected texture statistics based on [32]: contrast, entropy, variance, and correlation, which provided texture information on structure and water content of vegetation – volume scattering. Then, conducted an averaging of SAR volume scattering across seasons by using a machine (random forest) learning classification algorithm; we include other land cover classes in classification analysis to derive a thematic land cover map of the heterogeneous landscape

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 179; doi:10.3390/ijgi8040179

2. Study site

This study was conducted in the landscape of Bakoa (32N 734280m E 510975m N and 747435m N 501480m E, 123.28 km2), which is located in the Bokito District of the Mbam and Inoubou Department, in the Centre Region of Cameroon (Figure 1). This area is classified as a savannah-forest transition zone. The topography features a rolling terrain and the altitude ranges between 500 – 900m a.s.l. The vegetation is a mosaic of bush-savannah, subsistence farming, and perennial cocoa agroforests. These perennial agroforests are established mainly within or along patches of transition and gallery forests. The study area is situated in the humid forest bimodal agro-ecological zone, which is characterized by two dry and wet seasons. The total annual rainfall ranges between 1300 – 1500mm with a long rainy season from August to November. The main dry season lasts about 5 months from November – April. The mean annual temperature is 25°C.

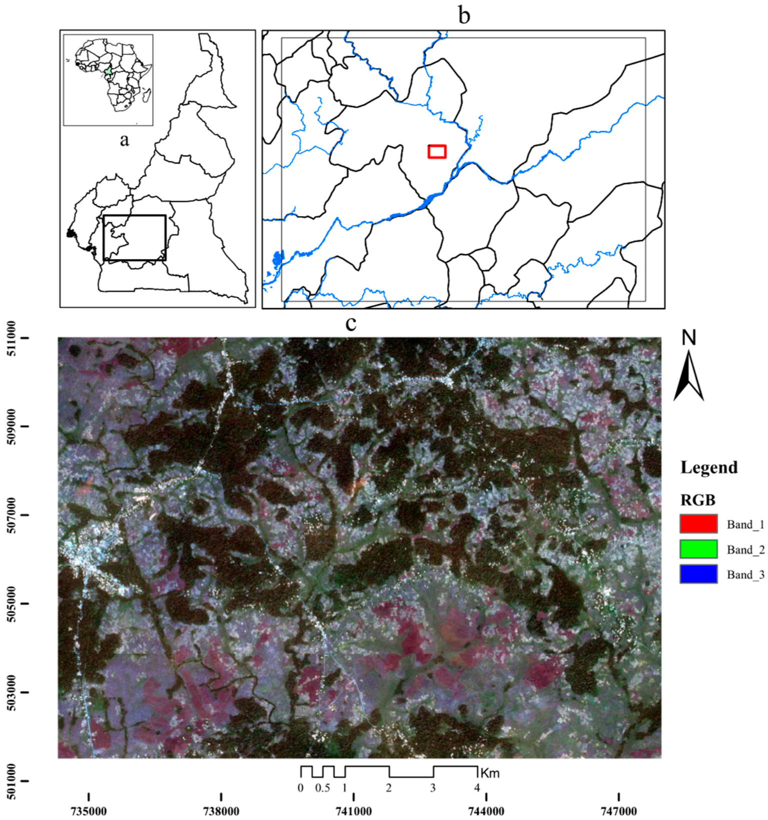


Figure 1. (**a**) Study area located in the centre region of Cameroon; (**b**) Study landscape, in red footprint, within the Bam and Inoubou administrative department; (**c**) RapidEye false colour image (RGB: Blue, Green, and Red spectral bands) revealing a mosaic of forest and savannah vegetation in the landscape.

3. Materials and methods

3.1 Satellite data: optical and radar imagery

We acquired a multispectral optical image of 5m spatial resolution from RapidEye, which was recorded in the dry season of 2015. The image comprised five spectral bands in the Blue (400-510 nm), Green (520-590 nm), Red (630-685 nm), RedEdge (690-730 nm), and Near Infrared (760-850 nm) range of the electromagnetic spectrum. Four different image tiles, acquired on the same date, were needed to cover the study landscape.

We accessed Sentinel-1A C-band (λ=5.5 cm) SAR images for the study area from the Sentinel Scientific Data Hub of the European Space Agency (ESA). The SAR data were acquired in dual (VV and VH) polarization, the Interferometric Wide swath (IW) observation mode, and pre-processed to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

Level-1 or ground range detection level (GRDH) of 10m spatial resolution. We selected a temporal series of 50 images acquired between March 2015 and April 2017 that covered both the dry and wet seasons. Using the image processing tools of Sentinel Application Platform (SNAP) version5.0, we prepared image subsets and pre-processed them sequentially from radar backscatter intensity values to sigma naught (sigma0) backscatter coefficients: thermal noise removal, filtering with orbit file, radiometric calibration, geometric rectification, and terrain correction. The digital elevation model (DEM) of the Shuttle radar topographic mission (SRTM) was applied, in SNAP with Sentinel-1 toolbox (S-1TBX), for terrain correction and geometric rectification of SAR images. We used both the co- (VV) and cross-polarised (VH) bands of all images. We then projected the pre-processed 10m resolution images in WGS 1984 UTM Zone 32 N. A list of the remote sensing data is presented in Table 1.

Table1. Used remote sensing data: mono-date RapidEye (5m) and multi-date Sentinel-1 SAR (10m) data.

Satellite mission	Scene ID(s)	Acquisition date (DD/MM/YYYY/)	Sensing stop time (HH:MM:SS UTC)	Acquisition mode (Polarization)	Data level
RapidEye: RE-3	3241224_ 3241225_ 3241124_ 3241124_	09-JAN-2015	10:35:41.00	MSI , Optical	L3A
Sentinel-1A	_006256_008304_78DE	06-JUN-2015	17:28:11.147769	IW Ascending (Dual: VV,VH)	Level1 GRD
	_007306_00A05D_2111	17-AUG-2015	17:28:14.323283		
	_007831_00AE86_4926	22-SEP-2015	17:28:15.577539		
	_008706_00C641_B612	21-NOV-2015	17:28:15.454239		
	_010456_00F838_64CF	20-MAR-2016	17:28:13.302867		
	_011156_010D64_7E35	07-MAY-2016	17:28:15.219784		
	_012031_012962_8F08	06-JUL-2016	17:28:18.678294		
	_012906_01465C_878C	04-SEP-2016	17:28:21.557129		
	_014831_0182BC_16C4	14-JAN-2017	17:28:19.166952		
	_015706_019D94_BC50	15-MAR-2017	17:28:18.681981		

3.2 Field campaigns

During the field campaigns conducted in 2015, 2016 and 2017, we collected ground information on land use and cover. The field data comprised ground GPS information and inventory of representative areas that characterize the different land cover and uses in the landscape (see Figure 2 and Table 2).

Table 2. Description of the thematic land cover types used for classification of land cover (Figure 2).

Class acronym	Class Name	Description
Bu	Built up	Residents, commercial/markets, industrial, administrative settings
Es	Earth road/bare soil	Land areas of exposed soil, bare rocks
Sv	Shrub/grassland Savannah	*Imperata sp.* savannah land: Shrubby and grassland areas which have not been converted to farm
W	Water	River, ponds, seasonal and permanent swamps

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

Af	Perennial cocoa agroforests	Land areas used for cocoa production with various degrees of canopy stratification: canopy/shade trees are mainly deciduous
Fa	Subsistence farming	Savannah and forest land areas that have been converted essentially for permanent or seasonal subsistence crop production; including farm fallows
Sf	Transition/Secondary forest	Disturbed and gallery forest patches, secret/cultural forest, hunting forest: have a more permanent and less stratified canopy structure

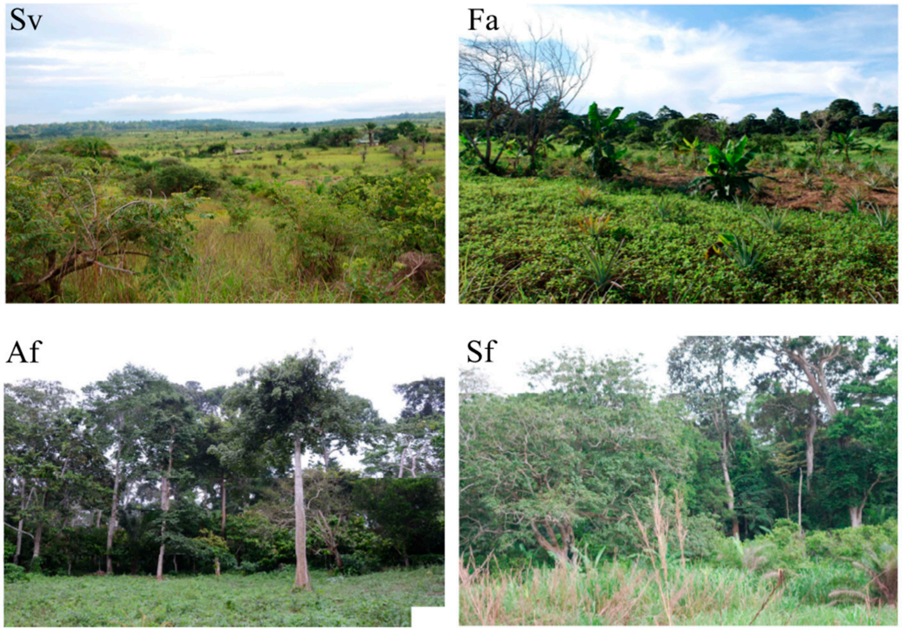


Figure 2. The range of vegetation land cover differ mainly in the density of woody biomass, which changes with season or phenological period. Class acronyms are described in Table 2.

3.3 Image processing workflow

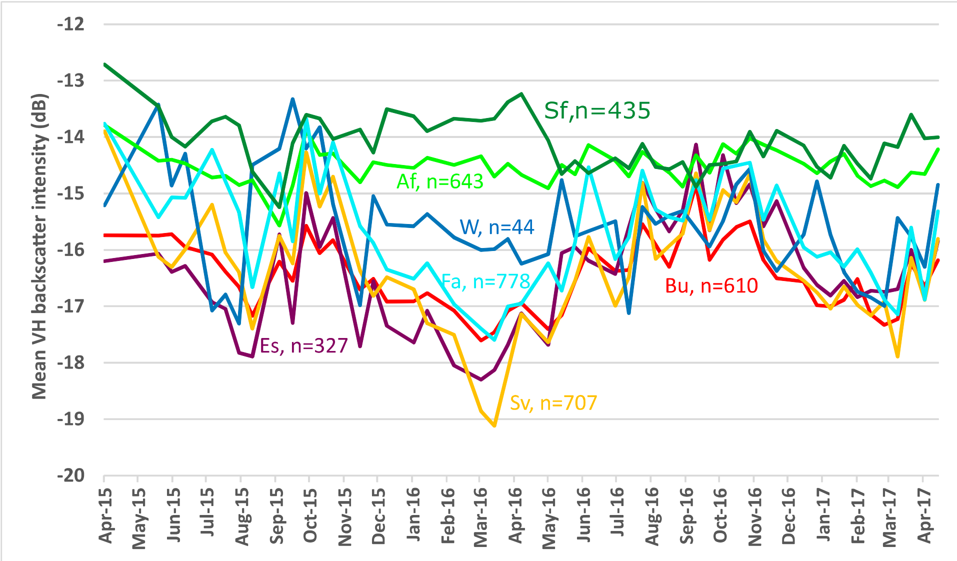
For RapidEye images, the following pre-processing protocol was conducted: atmospheric rectification by dark object subtraction (DOS), radiometric calibration to reflectance values, geometric correction, and finally computing different vegetation indices from a mosaicked image. After subsetting the SAR images, we used the batch processing mode of SNAP for the following pre-processing steps: radiometric calibration to Sigma0 (decibels), and geocoding with SRTM 3sec DEM using RangeDoppler Terrain Correction. We used intensity backscatter profile (Figure 3) and the Random Forest (RF) important variables criterion to remove noisy SAR images, and selected a subsample of 10 (of the 50) important images that represent six wet and four dry seasons between 2015 and 2017 (see summary description in Table 1 and Figure 4).

The image processing steps, detailed in following subsections, comprised of three major categories: (a) Feature extraction: this consisted of computing images of vegetation indices and GLCM texture images; (b) Image classification: we co-registered the vegetation index and texture images into the separate stacks or models described in Table 3. Then we ran eight RF ensemble (machine learning) classification algorithms, using the image stacks as input (Table 3), and (c) Post-processing: estimation of uncertainties in classified maps, in addition to accuracy metrics, as the basis for validating the classifier models. Finally, for the texture-based model with the highest overall

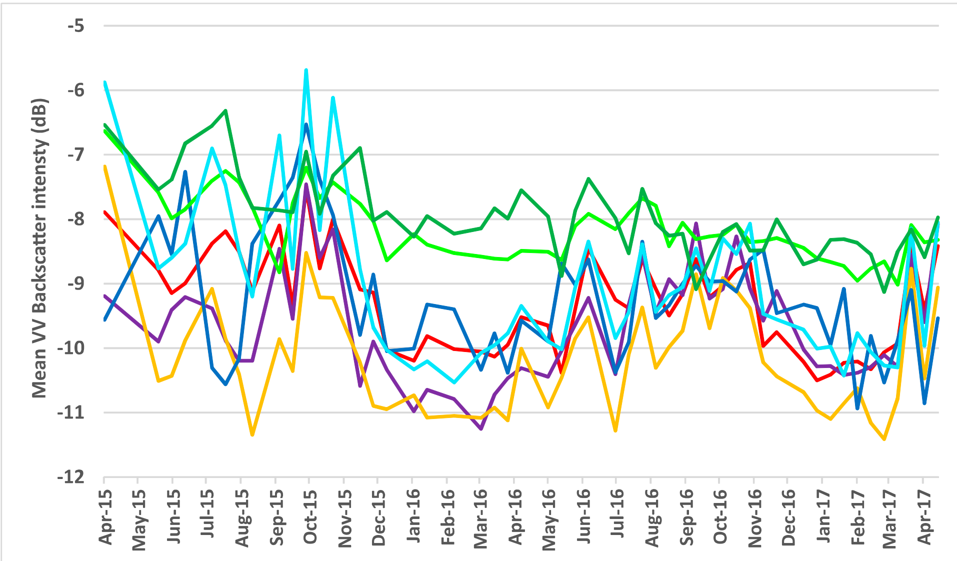
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

accuracy, we evaluated GLCM texture images at four different grey level quantization, to improve classification uncertainties.



(a)



(b)

Figure 3. Radar backscatter intensity temporal profiles for the different land use/cover types using the SAR images. (**a**) Vertical transmitted, horizontal receive backscatter; (**b**) Vertical transmitted, vertical received backscatter. The number, n, for each label pertains to the amount of the sample pixels (10×10m²)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

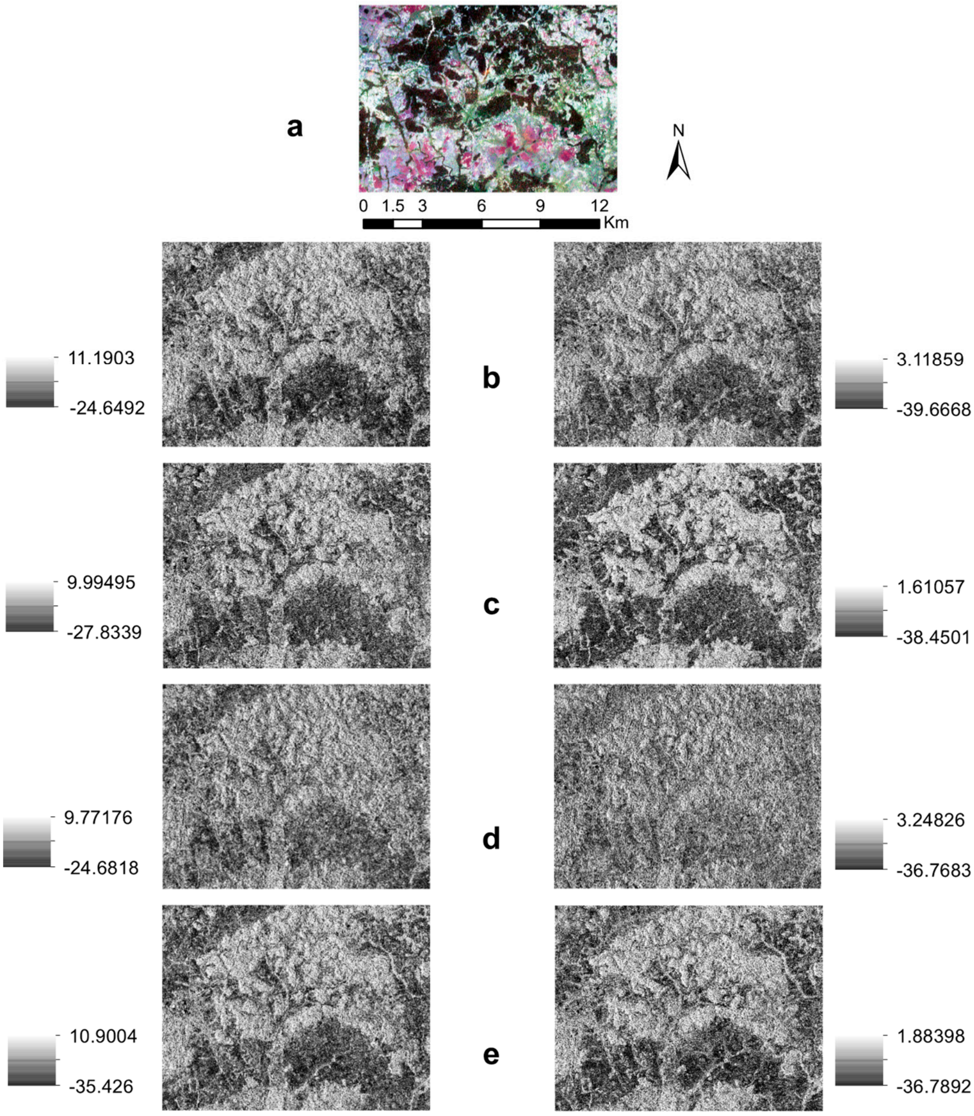


Figure 4. Radar backscatter intensity (dB) images of study landscape for key seasons from 2015 to 2017. (**a**) RapidEye false colour (RGB: bands 1, 2, and 3) composite image; (**b**) start of wet season: 17 August, 2015; (**c**) peak dry season: 20 March, 2016; (**d**) peak wet season: 4 September, 2016; (**e**) mid of dry season: 14 January, 2017. The left and right columns are, respectively, the VV and VH backscatter for each image. The North and scale bar is applicable to all images.

3.3.1. Feature extraction: Vegetation Indices (VIs) and GLCM texture features

The monitoring of vegetation status and extent is often based on normalisation ratios of spectral bands, in the Visible and Near Infrared (NIR) spectrum [42] in spaceborne imagery. These ratios are based on the contrasting spectral response of vegetation to the Red and NIR wavelengths.

The application of indices such as NDVI (Normalised Difference Vegetation Index) for vegetation monitoring have faced several challenges [43], notably issue of saturation for biomass above certain thresholds, and which is common in moist tropical vegetation. And, although saturation may not be an issue over agricultural landscapes, reflectance from soil background often perturb discrimination of sparse vegetation or cropland from bare soil [44]. In this study, we used VIs whose values indicate the status and abundance of vegetation and biomass, and that minimise the effect of soil background on vegetation reflectance values [45]: NDVI, gNDVI (green NDVI), EVI2 (Enhanced Vegetation Index), SAVI (Soil Adjusted Vegetation Index), and MSAVI (Modified SAVI) [46–50]. However, to provide additional information on vegetation characteristics and vitality, recent

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

optical sensors include an additional spectral band - the red edge band featured in RapidEye and Sentinel2 [51]. This band is located between red absorption (by chlorophylls) zone, and the NIR waveband.Since radar backscatter signals from a ground resolution cell are pseudorandom, the interaction of microwaves with terrain objects may be difficult to predict. Moreover, SAR images have speckle effect because the response signal of a resolution cell is a coherent interference from multiple scattering elements within the cell. Based on texture information extraction, the analysis of SAR images has been used for discrimination of cropland [26] and forest biomass estimation [27]. Often, the GLCM statistical approach is used in analysing SAR textures. The GLCM is a sparse matrix that stores co-occurrence probabilities of inter-pixel grey levels in an image [33]. These probabilities provide a second-order measure for texture features in an image: they represent conditional joint probabilities of all pairwise combination of grey levels (G) in the spatial window of analysis, and depend on both the spatial orientation (θ) and displacement distance (δ). Computation of GLCM is faster for images with fewer grey levels, because the matrix is dimensioned to G. The conditional probabilities are estimated as follows:

$$\Pr(x) = \{C_{ij} | (\theta, \delta)\} \tag{1}$$

Where, C_{ij} = co-occurrence probability between grey level i and j; and is defined by

$$C_{ij} = P_{ij} / \textstyle\sum_{i,j=1}^{G} P_{ij} \tag{2}$$

Where, P_{ij} = number of occurrence of grey levels i and j within the given window, for a certain (θ, δ) pair; G = the quantized number of grey levels. The denominator sums up to the total number of grey level pairs (i, j) within the analysis window.

Although different second-order statistics are commonly used to classify single images [52], some GLCM texture measures are auto-correlated [33]: a selection of a few texture measures may be reliable in achieving specific image analysis objective(s) [32]. We assess the accuracy of SAR images, covering several seasons, in discriminating perennial agroforestry land cover, using four less correlated GLCM texture measures: *Contrast*, *Entropy*, *Correlation*, and *Variance*. We estimated the textures measures from GLCM using a 5×5 moving window, an aggregate orientation of four directions (0°, 45°, 90°, and 135°), and one-pixel displacement (inter-pixel distance).

$$Contrast = \textstyle\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P_{i,j}\, (i-j)^2 \tag{4}$$

$$Entropy = \textstyle\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P_{i,j} \log_2 P_{i,j} \tag{5}$$

$$Variance = {\sigma_i}^2 = \textstyle\sum_{i,j=0}^{G-1} P_{i,j} (i - \mu_i)^2 \tag{6}$$

$$Correlation = \textstyle\sum_{i,j=0}^{G-1} P_{i,j} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j}\right] \tag{7}$$

where $P_{i,j}$ is the joint probability distribution of the grey levels *i* and *j* at two ends of a displacement vector in the assessment window, and G is the number of rows or columns. Since we considered a symmetrical GLCM, $\mu_i \equiv \mu_j$ and ${\sigma_i}^2 \equiv {\sigma_j}^2$. And for Entropy, 0 × ln(0) = 0 , since ln(0) is undefined.

3.3.2. Classification: Random Forest ensemble algorithm

In this study, we used a non-parametric machine learning algorithm, the Random Forest (RF) ensemble as an image classifier. This algorithm builds multiple decision trees for the same dataset based on random bootstrapping of sample training data [53] . The random forest classifier is less influenced by the common issue of over-fitting and is able to handle a large number of variables [54]. Firstly, each tree is built from a random subset (n) of two thirds of the original samples (N) – the 'in-bag' data; and secondly, from a subset (m) randomly selected from the total (M) variables in the

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

dataset – *mtry*, in each decision tree nodes are split using a best split variable – the one that yields the highest decrease in impurity [54]. The algorithm is a soft classifier on the basis of the probability voting of pixels belonging into the respective classes considered (Table 2). Compared to other non-parametric classification algorithms, it is less constrained by the need of extensive training and test data samples; this is due to an integrated out-of-bag (OOB) error estimation and accuracy test following a bootstrap sub-sampling on input data. Several sources provide additional details on the random forest algorithm [54,55].

We ran eight RF models for the different images stacks as classifier input (Table 3). For each model, we evaluated the OOB error curve and *mtry* to prune decision trees to an optimal number. For a spatially explicit and unbiased representation of each land cover class in the RF models, we divided the extracted pixel information for each class into a stratified random sample of 70% and 30% pixels respectively for training and testing the models. Image classification was conducted using the random forest package [56] of R programming software 3.4.3.

Table 3. The respective image stacks, used to compare the Random Forest (RF) classification accuracy.

Data Categories	Model	Image stack
RE: Dry season Multi-spectral RapidEye Image	RE1	TOA Reflectance of B, G, R, Red Edge, and NIR: 5 Bands
	RE2	TOA Reflectance and Vegetation Indices (VIs): 10 Bands
GL: Multi-date and season SAR GLCM Textures	GL1	Multi-date VV GLCM Textures: 40 Bands
	GL2	Multi-date VH GLCM Textures: 40 Bands
	GL3	Multi-date VV and VH GLCM Textures: 80 Bands
GLI: Multi-date and season SAR intensity and GLCM Textures	GLI1	Multi-date SAR VV Simga0 intensity and VV GLCM Textures: 50 bands
	GLI2	Multi-date SAR VH Sigma0 intensity and VH GLCM Textures: 50 bands
	GLI3	Multi-date SAR VV plus VH Sigma0 intensity and, VV plus VH GLCM Textures: 100 bands

Several studies have used RF for classification of forest cover [57] and cropland [26]. Although data over-fitting and poor prediction with RF approach has been reported [58] contrarily, it has been shown to provide relatively better classification of croplands [26] and mangrove vegetation [52]. Thus the performance of the RF ensemble classification algorithm may vary for different landscapes and cropping systems. For example, Loosvelt et al. [59], observed a high classification uncertainties for mixed pixels, at the heterogeneous boundaries of internally homogeneous cropping fields. Likewise, Van Tricht et al. [60], reported low classification confidence at such crop boundaries. Mixed cropping systems are common and the norm in moist tropical landscapes. However, reported research on the processing and use of SAR images for mapping of tropical heterogeneous cropping land, such as perennial agroforestry, is scarce.

3.3.3. Post-processing and classification uncertainty assessment

In remote sensing mapping, the validity and reliability of classified maps are often decided on basis of estimated overall accuracy and kappa coefficient [26]. Values such as user's and producer's accuracy are prone to errors and uncertainties [61]. As a soft classifier, however, the RF algorithm provides the possibility for assessing data- and computation-related uncertainties [59]. In our analysis, we used user's accuracy (omission error), producer's accuracy (commission error), overall accuracy and Kappa statistics – which compares results of a chance classification versus our RF model

 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 179; doi:10.3390/ijgi8040179

accuracy. However, the pixel-based classification methods are prone to uncertainties coming from either the use of unreliable data [61]. Thus, RF algorithm, as a soft classifier, provides a vector (P_u) of classification probability or votes for each image pixel - $P_u = P_1, P_2, P_3,......, P_n$ for a classification with n categories, and P_i denotes the probability of belonging to class i (Table 2).

In this study, in addition to model OOB error estimation, we evaluated classification uncertainties of RF models using the maximum classifier probability (U), and a weighted uncertainty measure entropy: the Shannon entropy (H) (Shannon, 1948; Vajapeyam, 2014). These uncertainties were calculated as:

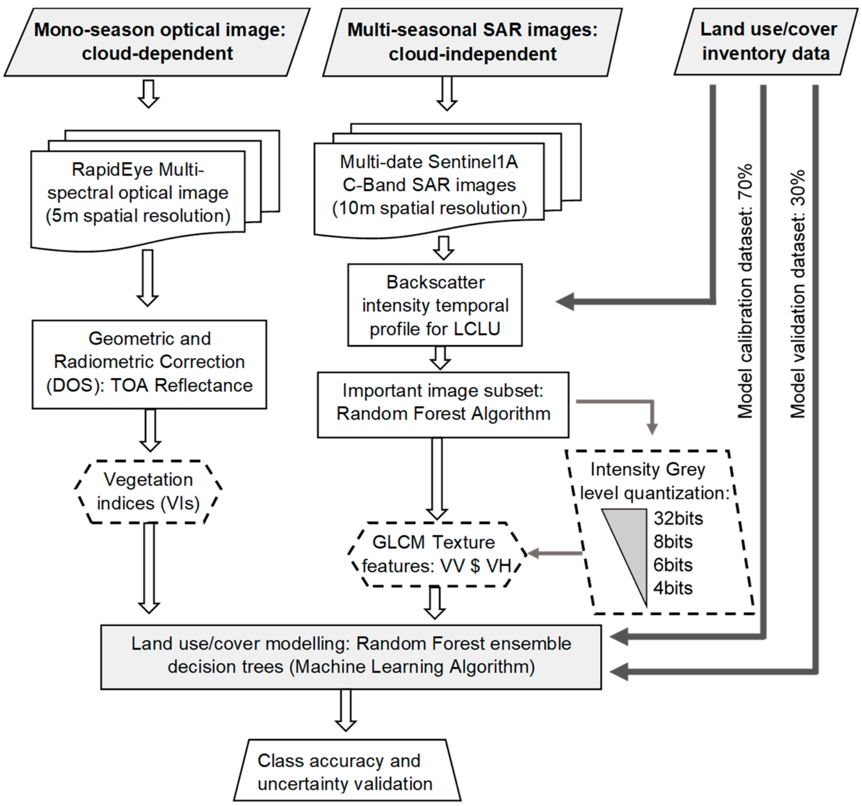
$$U = 1 - P_{max} \tag{8}$$

$$H = -\sum_{i=1}^{N} P_i \, . \log P_i \tag{9}$$

Where, P_i = probability of belonging to class i and P_{max} = maximum probability vote for a pixel's class. N = the total number of classes considered for analysis.

The maximum probability class assignment, by the soft classifier, for a pixel does not always result in assigning the true class label to the concerned pixel. Thus, by considering the entire range of values in a pixel's probability vector, H, compared to U that only makes use of P_{max}, provides a more robust measure of uncertainty; it has a maximum value at highest entropy – equal probability votes for all classes considered.

Loosvelt et al. [59], showed that H is reliable for evaluating uncertainties in mapping cropland from SAR images. However, our study area is characterised by heterogeneous cropping systems and is located in a tropical landscape (Figure 2). For the best-performing RF models, based on kappa accuracy, we computed and analysed U and H uncertainties for the classified maps, and the land cover classes considered in the study area (Table 2). The uncertainty estimations and analysis were conducted in Spyder IDE (Integrated Development Environment) of Anaconda distribution for Python programming software version 3.0 (Anaconda 3)



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

Figure 5. Schematic outline of the image processing and texture feature extraction, and land use/cover (LCLU) delineation by a machine learning algorithm. Input and outputs are shaded, and feature extraction in broken border lines. DOS: Dark object subtraction, TOA: Top of Atmosphere

4. Results

All the RF models had classification accuracies above 70%. Classification error and the sensitivity in discriminating land cover classes was different for each model. The models with the highest classification reliability, in increasing order of importance, were RE1, GLI3, and GL3.

4.1 Classification accuracy

Table 4 summarizes the classification results for all eight RF models. All models had a reliable overall accuracy (OA) above 70%. However, compared to using VV or VH bands separately, the use of both co- and cross-polarization bands (GL3) resulted in the highest classification accuracy. The GL3 model had the highest overall accuracy of 88.1% and kappa of 0.85; and, compared to other models, the OOB error estimate was the least with 12.8%. Also, classification from the multi-spectral optical image (RE1 model) had a reliable overall accuracy of 81.1%, but it had a lower kappa 76.9%. Compared to the GL3 model, the OOB error difference of +7% was observed for the RE1. The GLCM textures are reliable for discriminating the land cover/uses. And considering the heterogeneous and dynamic vegetation in the landscape, an improvement feature selection using the GLCM approach is necessary to reduce class uncertainties.

Table 4. Classification accuracies of different feature models based on the Random Forest (RF) classifier algorithm.

Model	Overall Accuracy-OA % (95% CI)	Kappa	OOB Error %
RE1	81.04 (79.68, 82.35)	0.769	19.18
RE2	80.15 (78.76, 81.48)	0.757	19.46
GL1	82.74 (80.02, 85.23)	0.787	17.12
GL2	81.65 (78.74, 84.32)	0.773	18.47
GL3	88.07 (85.52, 90.31)	0.853	12.85
GLI1	78.80 (75.85, 81.53)	0.738	19.66
GLI2	82.97 (80.21, 85.48)	0.789	18.71
GLI3	85.07 (82.42, 87.47)	0.817	13.69

The thematic land cover map from RE1 and GL3 models are shown in Figure 6. Separately, both VV and VH GLCM derived texture measures were poor in the prediction of non-vegetated land covers, and more so when both bands were included in the same model (Figure 6b). When included as input layers, the SAR backscatter intensity did not improve classification accuracy. Likewise, the inclusion of vegetation indices from the multispectral optical image, taken during a dry season, did not improve classification accuracy (Figure 6a). The texture measures from both VH and VH backscatter provide comparable, and may be complementary, LULC mapping accuracy to the commonly used vegetation indices of optical image.

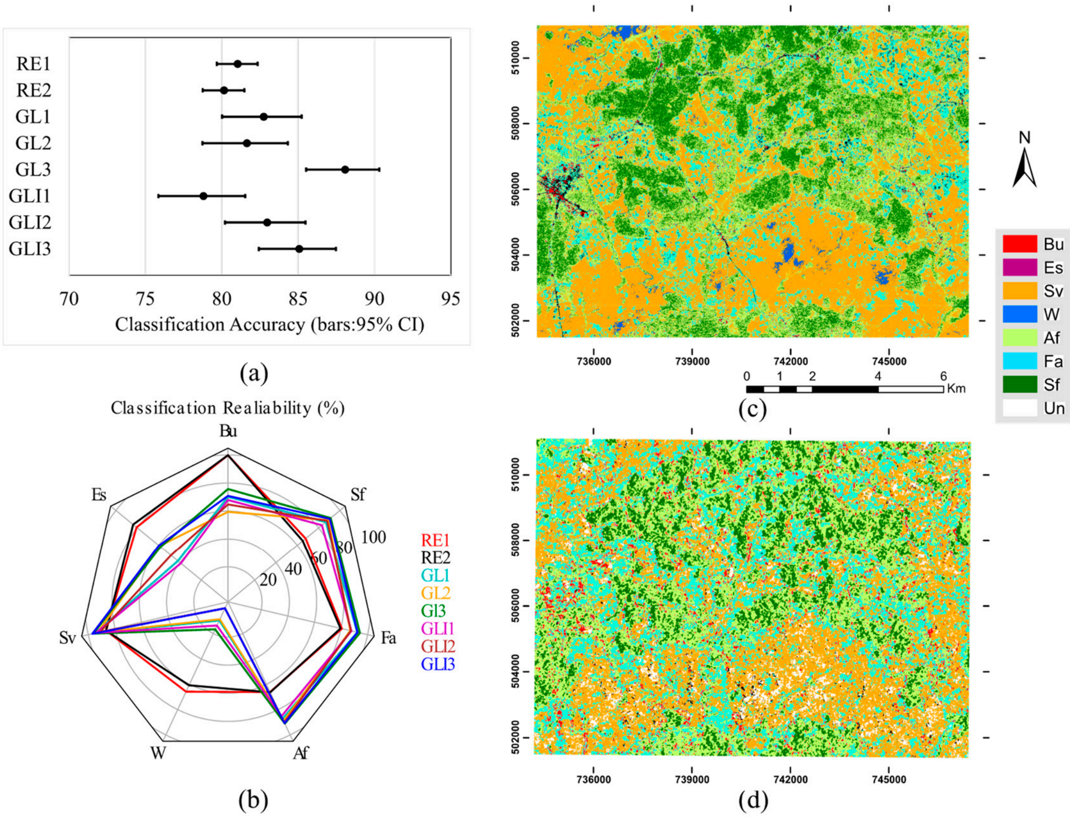


Figure 6. Pixel-based classification result for the eight models evaluated by random forest ensemble algorithm. (**a**) Classification accuracy; (**b**) Class reliability estimates; (**c**) Thematic land cover/use map from model RE1; (**d**) Thematic map model GL3. The scale, legend and north arrow apply to both c and d.

Visually, RE1 map shows a relatively intact and continuous expanse of transition forest patches (Figure 6c). Contrarily, the classified map from GL3 revealed that transition forest cover is highly fragmented by cocoa agroforests into smaller patches (Figure 6d). Also, from classification reliability estimates (100 - commission error) in Figure 6b, the RE1 model was more reliable in delineating non-vegetation land features. SAR-based texture images had a high reliability in delineating vegetation landscape features (Sv, Af, Fa, and Sf). Thus, although multi-spectral optical image had a better classification prediction of land cover classes in general, it was less reliable in discriminating perennial agroforests from transition forest land cover.

4.2 Uncertainty in discriminating vegetation land cover

The classification results from dry season RapidEye multi-spectral optical image (RE1) had a low overall and class uncertainties. From the cumulative estimates of class probabilities in Figure 7, classification uncertainty from the RE1 converges at a probability of around 0.6 for both U and H, whereas the uncertainty from GL3 map converges at higher probabilities – 0.7 and 0.9 respectively for U and H (Figure 7a). About 90% of pixels classified by RE1 had H uncertainties below 0.4, compared to about 50% of pixels for GL3. This difference is less obvious in the cumulative plot of U. Thus, uncertainty difference between RE1 and Gl3 was better revealed by Shannon entropy or H uncertainty.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

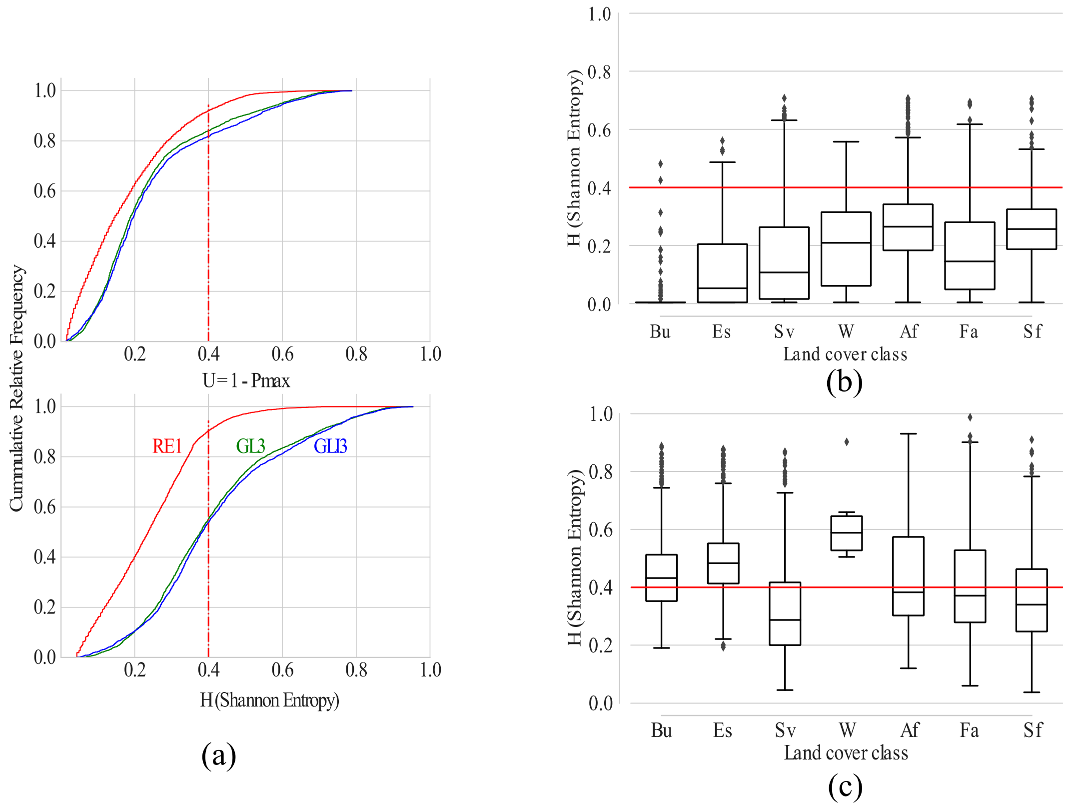


Figure 7. Classification uncertainties as validation of models with highest accuracy. (**a**) The Shannon entropy (H) clearly reveals uncertainty in classification accuracy validation. For the thematic maps from RE1 and GL3 R models: As example, the proportion of pixels with uncertainty below 0.4; (**b**) Individual class uncertainty, (H), for model RE1; (**c**) Class uncertainty, (H), for model GL3.

The individual class uncertainties are compared in Figure 7. Although the classified map from the multi-spectral image (RE1 model) had a lower accuracy, the class uncertainty was, compared to other land cover types, high for perennial cocoa agroforests and transition forest cover (Figure 7b). In comparison to RE1, the multi-seasonal SAR image textures, from the GL3 model, had a high overall uncertainty of a pixel's class prediction. However, perennial agroforests and transition forests were discriminated with relatively lower individual class uncertainty (Figure 7c): The median of class uncertainties were in a range between 0.2 and 0.4, which is comparable to those obtained from the single date multi-spectral image (RE1). The uncertainties in land cover/use discrimination by RE1 may reflect vegetation status/phenology - canopy greenness; while, in the GL3 model, volume scattering of radar signal reflects changes in water content and structure in different vegetation canopy.

4.3 Contribution of pixel depth to texture feature extraction

The SAR backscatter intensity images have a pixel quantization of 32 bits. The likelihood of grey level co-occurrences was lower at such high dynamic pixel range, as observable from unclassified pixels, *Un class*, in Figure 6c. In order to improve classification uncertainty from SAR image textures, we computed and compared three (3) different image pixel quantization or grey levels; 32bits – GL3 (as in original SAR intensity image), 8bits – GL3_B8, 6bits – GL3_B6, and 4bits – GL3_B4.

The contribution of grey levels on GLCM feature co-occurrence and classification probability is shown in Figure 8 and Table 5. Pixel co-occurrence was low at high pixel dynamic ranges (Figure 8a), which resulted in a large clustering of features with low and no predictions in Figures 8a and d. The dynamic pixel range of 64 grey level significantly reduced the prediction error (Figure 8e). The class prediction probability is improved to a maximum of 0.99 in GL3_B6, with a difference of 7% from 0.92 recorded in GL3. The prediction did not improve with further reduction of grey levels to 4

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

bits (Figure 8d). Optimizing the image pixel depth resulted in a marginal improvement of classification accuracy to a kappa value of 86.22 (Table 5). Though, the OOB error of prediction is reduced remarkably from the initial 12.8% (GL3) to 9.9% in GL3_B6. These results show that the dynamic pixel range was vital in feature selection (co-occurrence between pixels), and texture feature extraction by the ensemble algorithm.

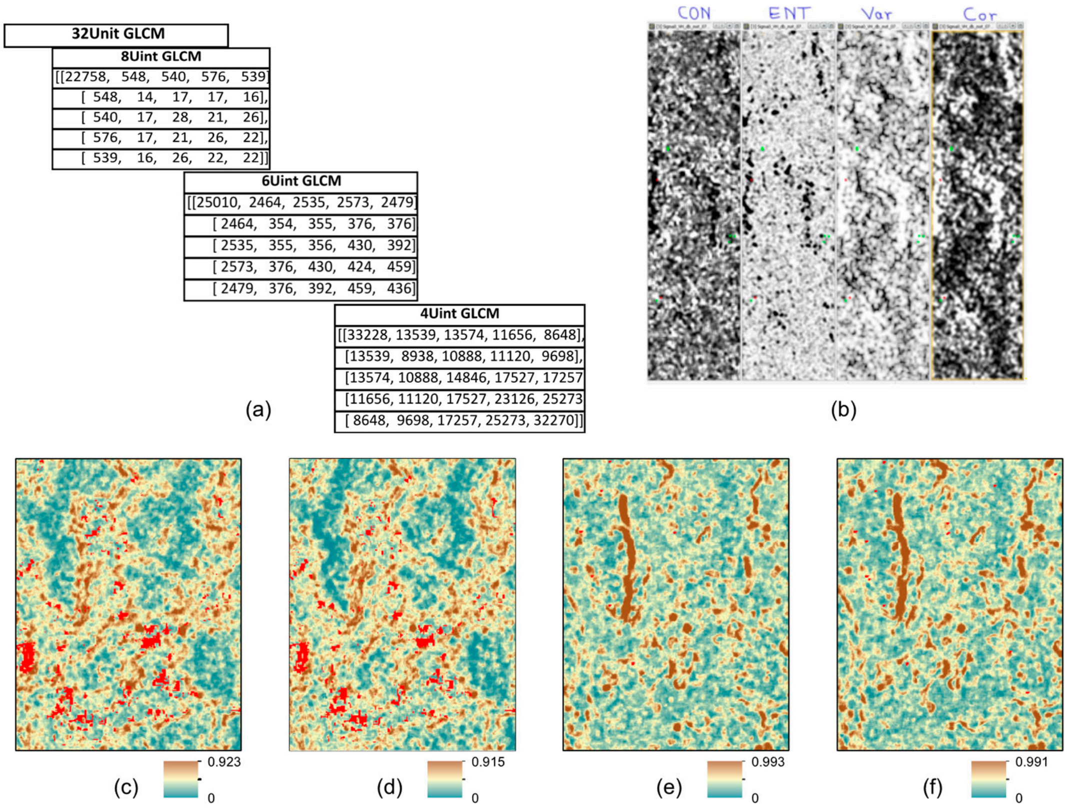


Figure 8. Classification probability maps reveal improved pixel classification with at lower grey level quantization. (**a**) Co-occurrence values in the first 5 × 5 cells of GLCM; (**b**) Sample details of the four GLCM texture measures; (**c**) to (**f**) are snips showing the respective details of the classification probability map for GL3, GL3_B8, GL3_B6, and GL3_B4 models. Unclassified areas are shown as red pixels in (c) and (d)

Table 5. Classification results of different grey level GLCM models and land cover/use surface area estimates. See Appendix A for details of class errors.

Model	Overall Accuracy-OA % (95% CI)	Kappa	OOB Error %
RE1	81.04 (79.68, 82.35)	0.769	19.18
GL3	88.07 (85.52, 90.31)	0.854	12.85
GL3_B8	88.23 (85.74, 90.41)	0.854	11.84
GL3_B6	88.83 (86.48, 90.90)	0.862	9.92
GL3_B4	88.86 (86.50, 90.94)	0.862	10.38

The improvements in the predicted land cover/use maps, after optimizing the dynamic pixel ranges, are illustrated in Figures 9 and 10. Figure 9 shows the predicted maps for the different texture-based models. Compared to the classification from GL3, at 32 bits pixel depth, with an estimated 647.3 ha as unclassified area (Figure 9a), the analysis at 6bits dynamic pixel range show less mosaic in discriminating class areas and reduced the unclassified land area to about 7.4 ha (Figure 9c). See Table A3 in Appendix.

The model validation results, by Shannon Entropy estimates, are shown in Figure 10. The difference in class uncertainty between GL3 and GL3_B6 is most evident for the following classes; Built up (Bu), Cocoa agroforests (Af), subsistence farms (Fa), and transition forests (Sf). The GL3_B6 had a comparable lower class error for these classes: respectively 16.4%, 4.4%, 5.3%, and 6.1% (see Figure A1 in Appendix). Remarkably, the class uncertainty estimates for Sf is low and comparable for GL3_B6 and RE1. Unlike the RE1 model, with no difference in class uncertainty between Af and Sf, the significant distinction between their class uncertainties in the GL3_B6 confirms the reliability of the model to discriminate these two vegetation categories.

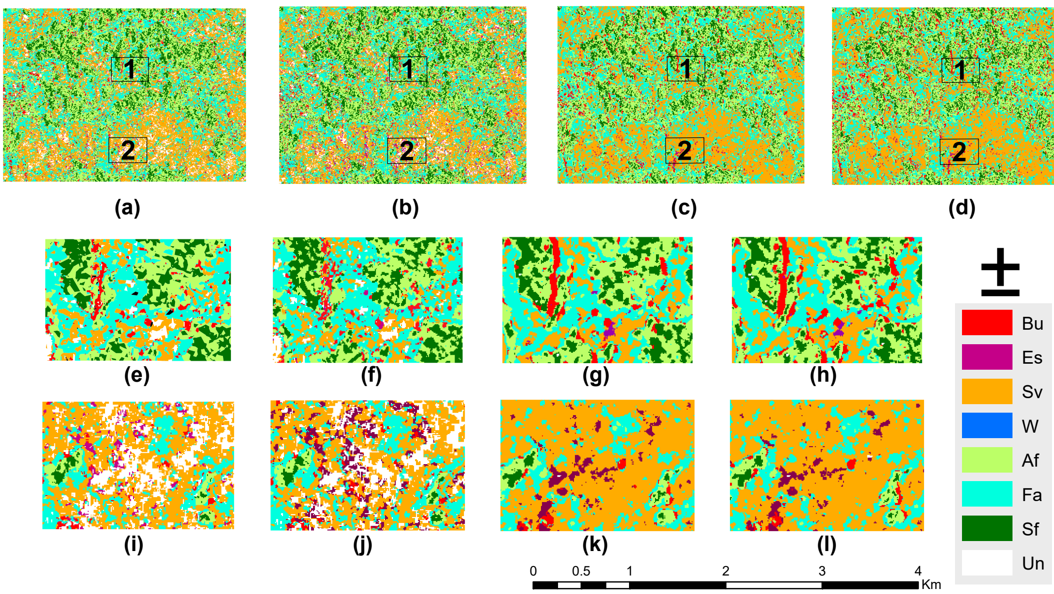


Figure 9. Classification prediction, based on Random forest algorithm, for the GLCM texture-based models. (**a**) GL3; (**b**) GL3_B8; (**c**) GL3_B6; (**d**) GL3_B4; (**e**) to (**h**) are corresponding details of the prediction maps in area 1 ; (**i**) to (**l**) are corresponding details in area 2. The legend and north arrow apply to all images and the scale bar to images (e) to (l).

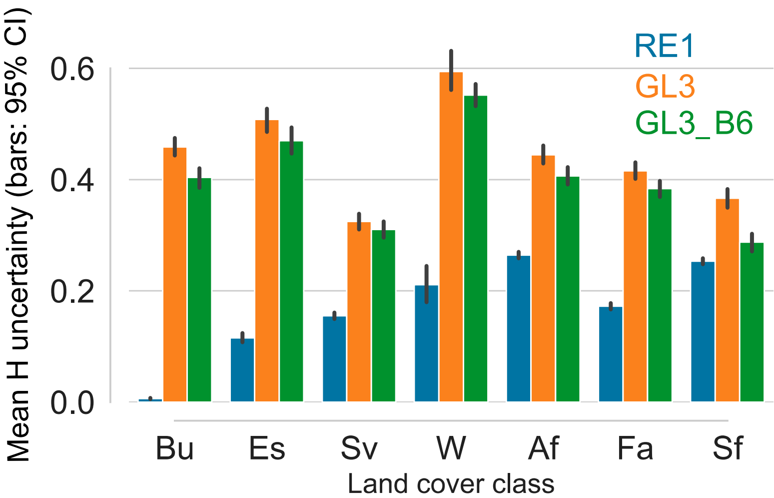


Figure 10. Comparison of validation by Shannon Entropy (H) estimation. Class uncertainty estimate is significantly improve by the GL3_B6, as observed for Built-up areas (Bu), Cocoa agroforests (Af), subsistence farmlands (Fa) and transition forests (Fs).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

The estimated land cover/use against the reference model (RE1) is shown in Figure 11. The total area for each class is summarized in Table A3 (see Appendix). A signification reduction in unclassified area is shown in 11c and d. The area of transition forest is lower for all models than estimated by RE1. Though the class error and uncertainty for GL3_B6 is comparable to that of RE1 (Figure 10). The land area for subsistence farming, cocoa agroforests and built up is remarkably larger in the texture-based predicted maps.

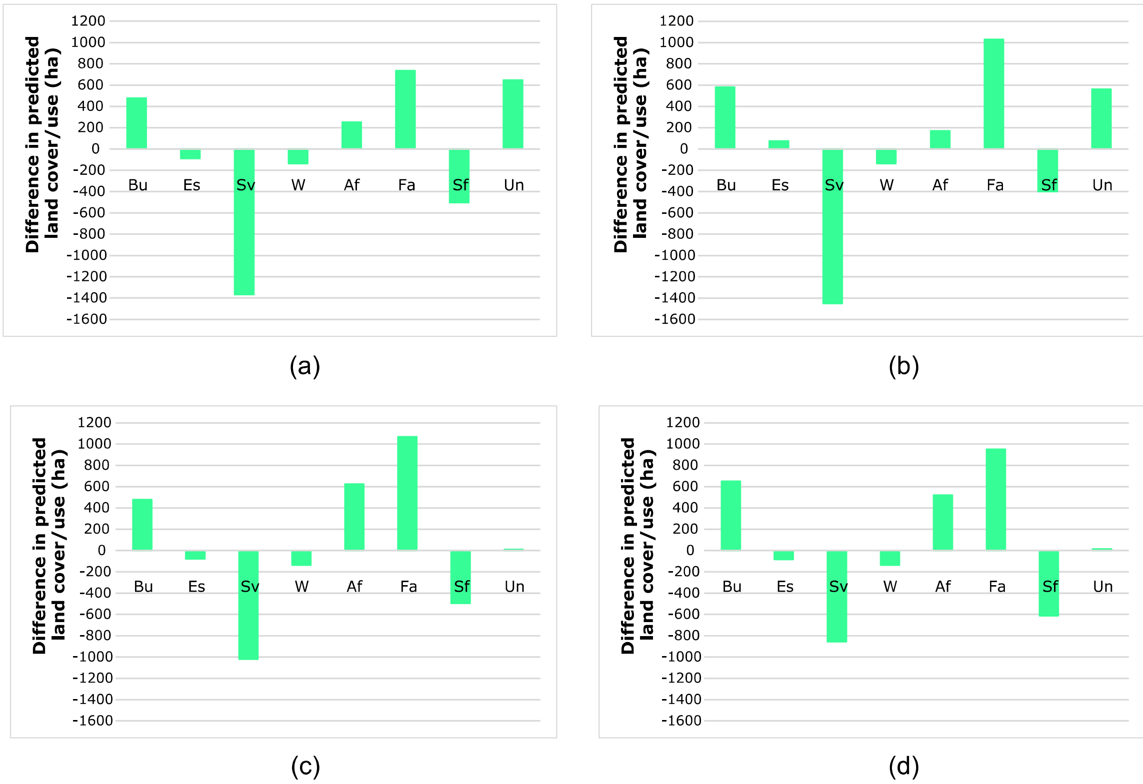


Figure 11. Comparison of the predicted land cover/use from RE1 model (optical image), in the landscape surface area of 11344.61ha, with: (a) GL3; (b) GL3_B8; (c) GLC_B6; and (d) GL3_B4. See Table A2.

5. Discussion

Land use and land cover (LULC) classification of satellite image is commonly applied to mono-date images; we use a rapidEye image that was acquired in the dry season – during minimal or no hindrance from cloud cover. In LULC change detection, a temporal series of images is often used, which require a reliable consistency in the characteristics of the processed images. For SAR image processing, therefore, image filtering or multi-looking is often applied to reduce speckle noise in SAR images [60]. However, such pre-processing also reduces the resolution of images. Thus, considering the landscape structure and the inherently heterogeneous vegetation categories in this study, we did not considered image speckle as noise. The selected SAR images were recorded during the ascending satellite overpass. We conducted a temporal average of measured textures across seasons, which, reduced any potential noise from individual image pixels; meanwhile, the seasonal differences in volume scattering over vegetation cover provided the texture details for discriminating the vegetation types. The temporal SAR data and the texture measures were, however, less sensitive to mapping non-vegetation land cover - notably the water cover. The low classification sensitivity for water areas can be explained by the following: variability in water cover as a result of seasonal swampy areas, the seasonal conversion of some swampy areas into subsistence farms of adapted crops, and the low backscatter intensity over water, which culminates in low or no co-occurring pixels

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

in the GLCM. Consequently, the GLCM had a low likelihood to extract texture for the water class. Nonetheless, this land cover was not of main interest in this study.

Compared to a "business as usual" classification using a mono-season multi-spectral optical image (RE1), a combination of texture from both VV and VH bands and the 6bits grey level quantization prior to GLCM texture classification had the highest OA of 88.1% and kappa of 0.85. Moreover, this accuracy resulted in, respectively, 3% and 9.3% reduction in prediction error over the GLCM texture at the default 32bits grey levels and the optical-image. Albeit the relatively low high prediction error of the RE1 model, the class prediction error was low for land cover with no- and low-vegetation: built up, bare soil, savannah and subsistence farmlands; these have rather distinctive optical spectral signatures. For the vegetation cover with high canopy, Table A1 shows the high confusion between cocoa agroforests and transition forests, both of which have high canopy. Since their canopy structure is similar [16], they may not be reliably discriminated using their spectral signatures [17], and less so during the dry season - of low leaf proliferation. Therefore, the classification performance from optical data is logical, and for the vegetation classes, reflects the vegetation status per season and phonological cycle; consequently, the optical reflectance in the dry season was less distinctive for cocoa agroforests versus secondary forests.

In the texture-based classification, we averaged volume scattering across seasons, which captured differences the vegetation types. As seen in Table A2, the confusion between cocoa agroforests and transition forest is low, compared to other classes. This indicates that optimizing the grey level improved the classification and helped to distinguish especially the vegetation classes with high heterogeneous canopy. For the study landscape the range of backscatter intensity, for VH and VV bands, is on average above 34 dB. The 6bits grey level quantization was consistent with the range of backscatter intensity in both VV and VH bands, and our results on grey level quantization confirms other reports [39–41]. Therefore, our land cover/use classification performance at grey level quantization of 6bits or 64 levels (GL3_B6) were optimal for discriminating different vegetation, particularly those featuring a high canopy. Other studies on heterogeneous cropland mapping recorded an accuracy of 71% using C-band SAR intensity images [26]. However, in terms of OA, our result is in line with the accuracy observed in different heterogeneous cropping landscapes using a combination of C-band SAR and optical data though [26,34]. The authors [26,34], mapped cropping lands with inherent homogeneous canopy. The landscape in the study is, however, tropical and dominated by vegetation and cropping fields with internally heterogeneous canopy.

Our texture-based classification result shows spatial fragmentation of forest cover by cocoa agroforests land use. These transition forest patches are consistent with field observations; the transition forest patches are mostly owned by families and community groups for hunting, performing traditional rituals, and serve as potential cocoa agroforests parcels. The similarity in canopy structure of cocoa agroforests and transition forests is explained by their matching class uncertainty estimates from the optical data i.e. RE1 model. Moreover, the confusion matrix in Table A1 reveals a high commission error between the two classes. Thus, by averaging the seasonal radar volume scattering, the GL3_B6 model discriminated the two vegetation cover with significantly different class uncertainties: the mean class uncertainty for transition forests was 0.28 in GL3_B6 based on GL3_B6 model, and unlike in RE1 model, this was significantly different from the 0.4 class average for cocoa agroforests. Following the high overall accuracy and a corresponding low individual class uncertainty, the multi-date texture information from SAR images provided a reliable classifier input for discriminating of perennial agroforestry land cover from transition forest.

Classification validation based on accuracy metrics as overall, user, and producer accuracies are influenced by sample class distribution in training data [64]. The differences in pixel resolution between images types (5 m for RapidEye and 10 m for SAR C-Band) resulted in a different number of reference pixels; of which lower numbers result in low estimates of misclassification probability (commission or omission errors). This explains the high overall accuracy for SAR image. The sample size is not as influential in classification validation by estimates of entropy or uncertainty. Entropy is a classic metric for biodiversity in ecology; thus, for this study, it describes the likelihood of a pixel belonging to either one of the classes considered [59]. Considering the heterogeneous landscapes, however, our results contradict other reported uncertainty in croplands mapping [59]; the trend of U and H uncertainties were different for our study landscape. At the uncertainty of 0.4 for example, the entropy or H, compared to U, revealed better the difference in the cumulative proportion of pixels between RE1 and GLE models. The slope of cumulative Shannon entropy (H) was linear for RE1 and sigmoidal for GL3. The former reflects a high likelihood of a single land cover class for each pixel in the multi-spectral image. The sigmoid curve represents a complete range of probabilities for each image pixel- a soft boundary between land cover classes. Compared to perennial agroforests, the low individual class uncertainty for transition forest is logical and consistent, as there is a high temporal stability of radar volume scattering over forest area [31].

The classification uncertainty, measured by Shannon entropy (H), for transition forests was comparable for both RE1 and GL3_B6: the likelihood of a pixel being classified only as transition forest was similar for both RE1 and GL3_B6 models. Noteworthy is the quantization level used for GLCM computation in the later, which it reduces unclassified areas to only 7.4 ha. Consequently, the forest area, estimated to cover 1706.9 ha (mean H = 0.25, class error = 0.32) using the RE1, was 500 ha (mean H = 0.28, class error = 0.06) less following the estimates from GL3_B6. Although the uncertainty for cocoa agroforests was significantly reduced by the GL3_B6, their relatively high entropy (mean H = 0.40, class error) indicate a high likelihood of classifying cocoa agroforests as one of the other classes. This distinction between the average entropy for cocoa agroforest (Af) and transition forest (Sf) for GL3_B6 model, confirms a reliable discrimination of the two vegetation types. The high variability in tree density and structure between cocoa agroforests may result in high intra-class variations in SAR backscatter intensity and confusion with other vegetation types, which might explain their high H uncertainty. Likewise, the diversity in farm subsistence crop types and stages may explain the high H for the class. As well for savannah lands, there is a frequent change of canopy volume and structure along with seasonal phonological cycles, except for areas converted to subsistence farming or previous farms that have been left fallow. Such physical changes results to unavoidable classification confusions [60]. The high class uncertainties validate the variability within land cover types and land use dynamics in the landscape; therefore, the uncertainties may not be easily reduced in a texture-based classification of SAR data.

The results of this study provide a new insight into the application of GLCM for SAR image processing for mapping tropical croplands of heterogeneous and multi-strata canopy. We highlight the important consideration of grey level quantization in heterogeneous land cover/use classification. Depending on the landscape structure and vegetation types, though, application of SAR imagery for cropland mapping may warrant different analysis procedure [26,34,60]. Thus there may be a need to customised analysis parameter depending on objective and landscape characteristics. The texture measures were pre-selected following the recommendation in reference [32], which is based on inference from optical image classification results, are tailored for vegetation mapping. This pre-selection of texture feature introduced bias in discriminating areas without vegetation. The GLCM texture measures were tailored to discriminate vegetation types, and less sensitive to non-vegetation cover. Texture feature extraction procedure are well established in the literature, with varying opinions. We highlight here, some issues which limit this study. We estimated the GLCM textures using a 5 × 5 window scale, considering the 10m pixel resolution for sentinel1 SAR. However, different window sizes may influence both texture values and classification accuracy. The analysis of different window sizes was not the aim of this study and may be a subject for further investigated for this and other landscapes. We perceive that a fusion of texture information with optical images [26,34], at a temporal scale may improve the uncertainty in discriminating vegetation classes. Also, considering the diversity of vegetation across cocoa production landscapes, the complex structure and canopy of cocoa agroforests, the continuous archive and improvement in SAR image resolution can train deep learning procedures for land cover/use classification at the scale of several landscapes.

6. Conclusions

This study, to our knowledge, is the first to explore multi-polarization and multi-temporal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

C-band SAR for discriminating cocoa agroforest cropping land in a heterogeneous landscape. We use seasonal differences in volume scattering from the dielectric (water content) status and structure of vegetation canopy as a metric to discriminate vegetation types. We make the following conclusions:

1. For the same window size and an invariant direction, reducing the grey level quantization improved classification accuracy marginally, but significantly reduced the uncertainty in discriminating cocoa agroforests from other vegetation covers.
2. Classification validation by estimates of Shannon entropy (H) reveal subtle differences in individual class prediction and provide reliable information for making inferences in heterogeneous landscape mapping.
3. The magnitude of forest fragmentation by cocoa agroforest, which is concealed by vegetation indices from spectral reflectance, is mapped using GLCM texture measures from C-band SAR images.

We suggest an approach for mapping cocoa agroforests in tropical heterogeneous cropping landscapes using C-band SAR imagery. The approach shows the reliability of Sentinel1 SAR image archive in landscape monitoring; especially for mapping cocoa agroforests expansion, and their contribution to the loss of transition and primary forest cover. Therefore, it has potential application in, for example, estimating the contribution of agroforestry to national and regional REDD+ (Reducing Emission from Deforestation and Forest Degradation and the role of conservation, sustainable management of forests and enhancement of carbon stocks in developing countries) strategies. However, there is a need to assess classification uncertainties in different agroforestry dominant landscapes – for an operational regional mapping in the Congo Basin sub-region.

Author Contributions: Conceptualization and design of the study, Frederick N. Numbisi; methodology, Frederick N. Numbisi and Frieke M.B. Van Coillie; field investigation, Frederick N. Numbisi; data curation, Frederick N. Numbisi; data analysis, Frederick N. Numbisi; original draft preparation, Frederick N. Numbisi.; review and editing, Frederick N. Numbisi, Frieke M.B. Van Coillie and Robert De Wulf; supervision, Frieke M.B. Van Coillie and Robert De Wulf.

Funding: Special Research Fund, Ghent University, contract number BOF DOS01W03314

Acknowledgments: This study was conducted under the Special Research Fund, of Ghent University, for students from developing countries, and through which RapidEye image was procured. Fieldwork was supported by the World Agroforestry Centre in Cameroon. Field inventory was assisted by local resource persons, administrative personnel and cocoa farmers who granted access to their plantations. We thank the European Space Agency for providing the freely accessible archive of Sentinel-1 SAR imagery under the Copernicus Open Data Hub. We express gratitude to developers and contributors to both R and Python programming and data mining software, which are open source and free. The authors would like to thank the editors of the ISPRS midterm Symposium and the anonymous reviewers for their valuable comments.

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

Appendix A

Table A1. Classification confusion matrix of RE1 Model. OA = Overall Accuracy, UA = User Accuracy, PA = Producer Accuracy.

nTrees = 250, mTry = 2, OOB error = 19.2%, OA = 81.0%										
		Reference								
		Bu	Es	Sv	W	Af	Fa	Sf	Class Error	PA
Predicted	Bu	1238	0	0	0	0	1	0	0.001	99.9
	Es	2	529	8	0	0	117	0	0.194	80.6
	Sv	0	0	1469	1	52	172	2	0.134	86.6
	W	0	0	9	52	16	0	2	0.342	65.8

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 179; doi:10.3390/ijgi8040179

Af	0	0	64	8	855	31	335	0.339	66.1
Fa	0	57	233	0	55	1447	3	0.194	80.6
Sf	0	0	3	0	309	4	665	0.322	67.7
UA	99.8	90.3	82.3	85	66.4	81.7	66.0		

Table A2. Classification confusion matrix of GL3_B6 Model.

nTrees = 550, mTry = 8, OOB error = 9.9%, OA = 88.8%										
		Reference							Class Error	PA
		Bu	Es	Sv	W	Af	Fa	Sf		
Predicted	Bu	244	2	5	0	23	17	1	0.164	83.6
	Es	14	89	19	0	6	21	0	0.403	59.7
	Sv	0	0	424	0	0	12	0	0.028	97.2
	W	2	0	3	0	6	7	0	1.000	0
	Af	3	0	0	0	325	8	4	0.044	95.6
	Fa	1	0	18	0	3	415	1	0.053	94.8
	Sf	0	0	0	0	16	0	246	0.061	93.9
UA		92.4	97.8	90.4	0	85.8	86.5	97.6		

Table A3. Predicted surface area (ha) of land cover/use for optical multi-spectral image and GLCM textures at four different grey level quantization

LULC Class	**Model**				
	RE1	**Gl3**	**Gl3_B8**	**Gl3_B6**	**Gl3_B4**
Bu	68.89	547.38	650.35	545.89	719.07
Es	160.73	73.41	235.14	83.14	76.15
Sv	4485.2	3118.63	3035.37	3465.95	3629.72
W	137.08	0.27274	0.15079	0.16927	0.16446
Af	2733.63	2986.84	2904.18	3355.87	3254.77
Fa	2052.11	2787.95	3081.47	3120.90	3002.42
Sf	1706.94	1202.86	1306.70	1210.16	1094.61
Un	0	647.32	562.62	7.42	12.60

References

1. FAO. 2016. *State of the World's Forests 2016. Forests and agriculture: land-use challenges and opportunities. Rome.*; 2016; ISBN 9789251092088.
2. Ordway, E. M.; Asner, G. P.; Lambin, E. F. Deforestation risk due to commodity crop expansion in sub-Saharan Africa Deforestation risk due to commodity crop expansion in sub- Saharan Africa. *Environ. Res. Lett.* **2017**, *12*, doi:10.1088/1748-9326/aa6509.
3. Payne, O.; Mann, A. S. Zomming In:"Sustainable" Cocoa Producer Destroys Pristine Forest in Peru Available online: https://www.wri.org/blog/2015/zooming-sustainable-cocoa-producer-destroys-pritine-forest-peru (accessed on Dec 16, 2018).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 January 2019 doi:10.20944/preprints201901.0050.v1

Peer-reviewed version available at ISPRS Int. J. Geo-Inf. 2019, 8, 179; doi:10.3390/ijgi8040179

4. Kongsager, R.; Napier, J.; Mertz, O. The carbon sequestration potential of tree crop plantations. *Mitig. Adapt. Strateg. Glob. Chang.* **2013**, *18*, 1197–1213, doi:10.1007/s11027-012-9417-z.
5. Somarriba, E.; Cerda, R.; Orozco, L.; Cifuentes, M.; Dávila, H.; Espin, T.; Mavisoy, H.; Ávila, G.; Alvarado, E.; Poveda, V.; Astorga, C.; Say, E.; Deheuvels, O. Carbon stocks and cocoa yields in agroforestry systems of Central America. *Agric. Ecosyst. Environ.* **2013**, *173*, 46–57, doi:10.1016/j.agee.2013.04.013.
6. Gockowski, J.; Sonwa, D. Cocoa Intensification Scenarios and Their Predicted Impact on CO2 Emissions , Biodiversity Conservation , and Rural Livelihoods in the Guinea Rain Forest of Cocoa Intensification Scenarios and Their Predicted Impact on CO 2 Emissions , Biodiversity Conser. *Environ. Manage.* **2010**, doi:10.1007/s00267-010-9602-3.
7. Sonwa, D.; Weise, S.; Tchatat, M.; Janssens, M. Diversity of plants in cocoa agroforests in the humid forest zone of Southern Cameroon. *Biodivers. Conserv.* **2007**, doi:10.1007/s10531-007-9187-1.
8. Tscharntke, T.; Clough, Y.; Bhagwat, S. A.; Buchori, D.; Faust, H.; Hertel, D.; Hölscher, D.; Juhrbandt, J.; Kessler, M.; Perfecto, I.; Scherber, C.; Schroth, G.; Veldkamp, E.; Wanger, T. C. Multifunctional shade-tree management in tropical agroforestry landscapes - A review. *J. Appl. Ecol.* **2011**, *48*, 619–629, doi:10.1111/j.1365-2664.2010.01939.x.
9. Mukete, N.; Li, Z.; Beckline, M.; Patricia, B. Cocoa Production in Cameroon : A Socioeconomic and Technical Efficiency Perspective. *Int. J. Agric. Econ.* **2018**, *3*, 1–8, doi:10.11648/j.ijae.20180301.11.
10. FAOSTAT. 2017. Crops. Available online: http://www.fao.org/faostat/en/#data/QC (accessed on Aug 3, 2017).
11. Ruf, F.; Schroth, G.; Doffangui, K. Climate change, cocoa migrations and deforestation in West Africa: What does the past tell us about the future? *Sustain. Sci.* **2015**, *10*, 101–111, doi:10.1007/s11625-014-0282-4.
12. Barima, Y. S. S.; Tamia, A.; Kouakou, M.; Bamba, I.; Charles, Y.; Godron, M.; Andrieu, J. Cocoa crops are destroying the forest reserves of the classified forest of Haut-Sassandra (Ivory Coast). *Glob. Ecol. Conserv.* **2016**, *8*, 85–98, doi:10.1016/j.gecco.2016.08.009.
13. Jagoret, P.; Michel-Dounias, I.; Snoeck, D.; Ngnogué, H. T.; Malézieux, E. Afforestation of savannah with cocoa agroforestry systems: A small-farmer innovation in central Cameroon. *Agrofor. Syst.* **2012**, *86*, 493–504, doi:10.1007/s10457-012-9513-9.
14. Dumont, S. E.; Gnahoua, G. M.; Ohouo, L.; Sinclair, F. L.; Vaast, P. Farmers in Côte d'Ivoire value integrating tree diversity in cocoa for the provision of ecosystem services. *Agrofor. Syst.* **2014**, 1047–1066, doi:10.1007/s10457-014-9679-4.
15. Gyau, A.; Smoot, K.; Diby, L.; Kouame, C. Drivers of tree presence and densities: the case of cocoa agroforestry systems in the Soubre region of Republic of Côte d'Ivoire. *Agrofor. Syst.* **2015**, *89*, 149–161, doi:10.1007/s10457-014-9750-1.
16. Sonwa, D. J.; Weise, S. F.; Nkongmeneck, B. A.; Tchatat, M.; Janssens, M. J. J. Structure and composition of cocoa agroforests in the humid forest zone of Southern Cameroon. *Agrofor. Syst.* **2016**, 1–20, doi:10.1007/s10457-016-9942-y.
17. Ordway, E. M.; Asner, G. P.; Lambin, E. F. Deforestation risk due to commodity crop expansion in sub-Saharan Africa Deforestation risk due to commodity crop expansion in sub- Saharan Africa. **2017**.
18. Norgrove, L.; Hauser, S. Carbon stocks in shaded Theobroma cacao farms and adjacent secondary forests of similar age in Cameroon. *Trop. Ecol.* **2013**, *54*, 15–22.
19. Tetteh, D. A.; Asase, A.; Ofori-frimpong, K.; Attuquayefio, D. Effect of cocoa farming intensification on

biodiversity and ecosystem properties in southern Ghana. *J. Ecol. Nat. Environ.* **2018**, *10*, 172–181, doi:10.5897/JENE2017.0673.

20. Braga, D. P. P.; Domene, F.; Gandara, F. B. Shade trees composition and diversity in cacao agroforestry ´ , Brazilian Amazon systems of southern Para. *Agrofor. Syst.* **2018**, doi:10.1007/s10457-018-0250-6.
21. FAOSTAT. 2018. Crops. Methodology – Crops Primary Available online: http://www.fao.org/faostat/en/#data/QC (accessed on Dec 12, 2018).
22. Mithöfer, D.; Roshetko, J. M.; Donovan, J. A.; Nathalie, E.; Robiglio, V.; Wau, D.; Sonwa, D. J.; Blare, T. Unpacking ' sustainable ' cocoa : do sustainability standards , development projects and policies address producer concerns in Indonesia , Cameroon and Peru ? *Int. J. Biodivers. Sci. Ecosyst. Serv. Manag.* **2018**, *13*, 444–469, doi:10.1080/21513732.2018.1432691.
23. Wood, G. A. R. Establishment. In *Cocoa*; Wrigley, G., Ed.; Blackwell Science: London, UK, 2001; p. 119–165, ISBN 978-0-632-06398-7.
24. Carodenuto, S.; Merger, E.; Essomba, E.; Panev, M.; Pistorius, T.; Amougou, J. A methodological framework for assessing agents, proximate drivers and underlying causes of deforestation: Field test results from Southern Cameroon. *Forests* **2015**, *6*, 203–224, doi:10.3390/f6010203.
25. Jia, K. U. N.; Li, Q.; Tian, Y.; Wu, B. Crop classification using multi-configuration SAR data in the North. *Int. J. Remote Sens.* **2012**, *33*, 170–183.
26. 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, doi:10.1080/22797254.2018.1455540.
27. Solberg, S.; Næsset, E.; Gobakken, T.; Bollandsås, O. M. Forest biomass change estimated from height change in interferometric SAR height models. *Carbon Balance Manag.* **2014**, *9*, 1–12, doi:10.1186/s13021-014-0005-2.
28. Stelmaszczuk-gorska, M.; Urbazaev, M.; Schmullius, C.; Thiel, C. Estimation of Above-Ground Biomass over Boreal Forests in Siberia Using Updated In Situ , ALOS-2 PALSAR-2 , and RADARSAT-2 Data. *Remote Sens.* **2018**, *10*, 1550, doi:10.3390/rs10101550.
29. Szigarski, C.; Jagdhuber, T.; Baur, M.; Thiel, C.; Parrens, M.; Wigneron, J.; Piles, M.; Entekhabi, D. Analysis of the Radar Vegetation Index and Potential Improvements. *Remote Sens.* **2018**, 1–15, doi:10.3390/rs10111776.
30. Thiel, C.; Cartus, O.; Eckardt, R.; Richter, N.; Thiel, C.; Schmullius, C. Analysis of multi-temporal land observation at C-band. *Int. Geosci. Remote Sens. Symp.* **2009**, *3*, 318–321, doi:10.1109/IGARSS.2009.5417764.
31. Schmullius, C.; Thiel, C.; Pathe, C.; Santoro, M. *Remote Sensing Time Series*; 2015; Vol. 22; ISBN 978-3-319-15966-9.
32. Hall-Beyer, M. Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *Int. J. Remote Sens.* **2017**, *38*, 1312–1338, doi:10.1080/01431161.2016.1278314.
33. Haralick, R. M. Statistical and structural approach to texture. *Proceeding IEEE vol 67 no 5* **1979**, *67*, 786–804, doi:10.1109/PROC.1979.11328.
34. Mishra, V. N.; Prasad, R.; Rai, P. K.; Vishwakarma, A. K.; Arora, A. Performance evaluation of textural features in improving land use / land cover classification accuracy of heterogeneous landscape using multi-sensor remote sensing data Performance evaluation of textural features in improving land use / land cover classif. *Earth Sci. Informatics* **2018**, doi:10.1007/s12145-018-0369-z.
35. de Siqueira, R. F.; Robson, W.; Pedrini, H. Neurocomputing Multi-scale gray level co-occurrence

matrices for texture description. *Neurocomputing* **2013**, *120*, 336–345, doi:10.1016/j.neucom.2012.09.042.

36. Zhou, J.; Guo, R. Y.; Sun, M.; Di, T. T.; Wang, S.; Zhai, J. The Effects of GLCM parameters on LAI estimation using texture values from Quickbird Satellite Imagery. *Sci. Rep.* **2017**, 1–12, doi:10.1038/s41598-017-07951-w.
37. Lan, Z.; Liu, Y. Study on Multi-Scale Window Determination for GLCM Texture Description in High-Resolution Remote Sensing Image Geo-Analysis Supported by GIS and Domain Knowledge. **2018**, doi:10.3390/ijgi7050175.
38. Marceau, D. J.; Howarth, P. J.; Dubois, J. M.; Gratton, D. J. Evaluation of the Grey-Level Co-Occurrence Matrix Method For Land-Cover Classification Using SPOT Imagery. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 513–519.
39. Clausi, D. A.; Zhao, Y. Rapid extraction of image texture by co-occurrence using a hybrid data structure. **2002**, *28*, 763–774.
40. Karthikeyan, S.; Rengarajan, N. Performance Analysis of Gray Level Co- occurrence Matrix Texture Features for Glaucoma Diagnosis. **2014**, *11*, 248–257, doi:10.3844/ajassp.2014.248.257.
41. Patel, M. B.; Rodriguez, J. J.; Gmitro, A. F. Effect of Gray-Level Re-quantization on Co-occurrence based Texture Analysis. *2008 15th IEEE Int. Conf. Image Process. San Diego, CA, USA, 12 15 Oct. 2008* **2008**, 585–588, doi:10.1109/icip.2008.4711822.
42. Silleos, N. G.; Alexandridis, T. K.; Gitas, I. Z.; Perakis, K. Vegetation indices: Advances made in biomass estimation and vegetation monitoring in the last 30 years. *Geocarto Int.* **2006**, *21*, 21–28, doi:10.1080/10106040608542399.
43. Mutanga, O.; Skidmore, A. K. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *Int. J. Remote Sens.* **2004**, *25*, 3999–4014, doi:10.1080/01431160310001654923.
44. Viña, A.; Gitelson, A. A.; Nguy-Robertson, A. L.; Peng, Y. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sens. Environ.* **2011**, *115*, 3468–3478, doi:10.1016/j.rse.2011.08.010.
45. Wiesmair, M.; Feilhauer, H.; Magiera, A.; Otte, A.; Waldhardt, R. Estimating Vegetation Cover from High-Resolution Satellite Data to Assess Grassland Degradation in the Georgian Caucasus. *Mt. Res. Dev.* **2016**, *36*, 56–65, doi:10.1659/MRD-JOURNAL-D-15-00064.1.
46. Liu, J.; Pattey, E.; Jégo, G. Assessment of vegetation indices for regional crop green LAI estimation from Landsat images over multiple growing seasons. *Remote Sens. Environ.* **2012**, *123*, 347–358, doi:10.1016/j.rse.2012.04.002.
47. Gitelson, A. A.; Kaufman, Y. J.; Merzlyak, M. N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298, doi:10.1016/S0034-4257(96)00072-7.
48. Jiang, Z.; Huete, A. R.; Didan, K.; Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845, doi:10.1016/j.rse.2008.06.006.
49. Huete, A. . A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 1988, *25*, 295–309.
50. Qi, J.; Chehbouni, A.; Huete, A. R.; Kerr, Y. H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126, doi:10.1016/0034-4257(94)90134-1.
51. Delegido, J.; Verrelst, J.; Alonso, L.; Moreno, J. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors* **2011**, *11*, 7063–7081, doi:10.3390/s110707063.
52. Abdel-Hamid, A.; Dubovyk, O.; Abou El-Magd, I.; Menz, G. Mapping Mangroves Extents on the Red Sea Coastline in Egypt using Polarimetric SAR and High Resolution Optical Remote Sensing Data.

Sustainability **2018**, *10*, 646, doi:10.3390/su10030646.
53. Breiman, L. E. O. Random Forests. **2001**, 5–32.
54. Hastie, T.; Tibsharani, R.; Friedman, J. Springer Series in Statistics The Elements of. *Math. Intell.* **2009**, *27*, 83–85, doi:10.1007/b94608.
55. Criminisi, A.; Shotton, J.; Konukoglu, E. Decision Forests for Classification , Regression , Density Estimation , Manifold Learning and Semi-Supervised Learning. **2011**.
56. Liaw, A.; Wiener, M. Classification and Regression by randomForest. **2002**, *2*, 18–22.
57. Balzter, H.; Cole, B.; Thiel, C.; Schmullius, C. Mapping CORINE land cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests. *Remote Sens.* **2015**, *7*, 14876–14898, doi:10.3390/rs71114876.
58. Zhao, K.; Suarez, J. C.; Garcia, M.; Hu, T.; Wang, C.; Londo, A. Remote Sensing of Environment Utility of multitemporal lidar for forest and carbon monitoring : Tree growth , biomass dynamics , and carbon fl ux. *Remote Sens. Environ.* **2018**, *204*, 883–897, doi:10.1016/j.rse.2017.09.007.
59. Loosvelt, L.; Peters, J.; Skriver, H.; Lievens, H.; Van Coillie, F. M. B.; De Baets, B.; Verhoest, N. E. C. Random Forests as a tool for estimating uncertainty at pixel-level in SAR image classification. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *19*, 173–184, doi:10.1016/j.jag.2012.05.011.
60. 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**, *10*, 1–22, doi:10.3390/rs10101642.
61. Unwin, D. J. Geographical information systems and the problem of “error and uncertainty.” *Prog. Hum. Geogr.* **1995**, *19*, 549–558, doi:10.1177/030913259501900408.
62. Shannon E. Claude A Mathematical Theory of Communication. *Bell Teleph. Syst. Tech. Publ.* **1948**, *27*, 379–423.
63. Vajapeyam, S. Understanding Shannon ' s Entropy metric for Information. **2014**, 1–6.
64. Heydari, S. S.; Mountrakis, G. Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. *Remote Sens. Environ.* **2018**, *204*, 648–658, doi:10.1016/j.rse.2017.09.035.