

Article

Random forest algorithm for mapping deforestation in the Ituri-Epulu-Aru landscape (Democratic Republic of Congo)

Joël Masimo Kabuanga^{1*}, Onésime Mubenga Kankonda^{1,2}, Ned Horning³, Mehdi Saqalli⁴, Nicolas Maestriperier⁵, Fils Makanzu Imwangana⁶ and Landing Mané⁷

¹ Département d'aménagement des écosystèmes, Faculté de Gestion des Ressources Naturelles Renouvelables, Université de Kisangani, P.O. Box 2012, Kisangani, Democratic Republic of Congo; mkabuang@ohmail.com

² Centre de Surveillance de la Biodiversité, Université de Kisangani, P.O. Box 2012, Kisangani, Democratic Republic of Congo; kankonda65@yahoo.fr

³ Center for Biodiversity and Conservation, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5192, United States; nedhorning@gmail.com

⁴ CNRS (UMR 5602 GEODE) Maison de la Recherche de l'Université du Mirail, 5, Allées A. Machado, 31058 Toulouse Cedex 1, France; mehdi.saqalli@univ-tlse2.fr

⁵ TerraNIS, 12 Avenue de l'Europe, 31520 Ramonville, France; nicolas.maestriperier@gmail.com

⁶ Unité de Recherche et de Formation en Gestion des Risques Naturels, Département des Géosciences, Faculté des sciences, Université de Kinshasa, B.P. :190 Kinshasa XI; nicolas.maestriperier@gmail.com

⁷ Satellite Observatory of the Forests of Central Africa, 14, Sergent Moke - Q/Socimat, Concession Safricas, Ngaliema/Kinshasa, DRC; lmene@osfac.net

* Correspondence: mkabuang@hotmail.com; Tel.: +243-826-369-021

Abstract: The evaluation of deforestation by optical remote sensing remains a challenge in the humid tropical region due to high cloud cover. This paper develops a simple and reproducible method for mapping deforestation of the old-growth forest using open access software. A map of old-growth forest depletion was created using composites from three different dates (2003, 2010, 2016). Four models were tested: the first model using spectral bands (nir, swir1, swir2 and red), the second model was based on the association of spectral bands and spectral indices (NDVI, B54R, NDWI and NBR), the third model was constructed using spectral bands and geomorphological indices (DEM, Slope and Roughness) and the last model combined spectral bands, spectral indices and geomorphological indices. The optimal random forest ntrees and Mtry parameters were determined for each model to optimize the mapping in each model. The out-of-bag error for these four models was 2.15 %, 2.05 %, 1.86 % and 1.85 %, respectively. The fourth model had the lowest error and was hence used to predict deforestation of the old-growth forest. The annual rates of deforestation amounted 0.26 % (69861 ha) and 0.66 % (145768 ha) between 2003 – 2010 and 2010 – 2016, respectively. The area of the old-growth forest in 2016 was 3601607 ha and 215629 ha of forest lost between 2003 and 2016. These results showed that the Random Forest Classification (RFC) model was able to effectively map the reduction of old-growth forests.

Keywords: Random Forest, machine learning, multispectral imagery, deforestation, PFBC landscapes.

1. Introduction

Equatorial old-growth forests represent a natural ecosystem having the most important quantity of biomass and among the highest biodiversity [1-3]. As land use in the world is undergoing profound spatial changes, old-growth forests are increasingly subjected to anthropogenic pressure [4-5].

The methods used to assess Land Use and Cover Change (LUCC) have improved considerably in recent decades [6-7; 4]. This assessment requires a rapid and reliable mapping of vegetation cover dynamics, theoretically hoping for a sustainable exploitation and management of natural resources. If forest inventory is still the main source of LUCC information, remote sensing is a powerful tool for the analysis of forest dynamics [8]. Remote sensing observations are an attractive source of information for survey and land-use mapping. For example, the global assessment of forest resources conducted by the United Nations Food and Agriculture Organization [9] through remote sensing estimated the world forest area at 3999 million ha. The same source indicates that global forest loss was 192 million ha between 1990 and 2015, that is, a rate of 2.16 million ha per year.

Much of the world's old-growth forest is in the tropics. Unfortunately, remote sensing mapping in this region remains challenging given the persistently high cloud cover [10-12]. For this reason, a deforestation assessment is more often carried out at the national [13], regional [14-15] or global level [7; 4]. Many natural resource management institutions rely on these assessments but the map products do not always meet managers expectations [16]. Few studies in the Democratic Republic of Congo have been undertaken to assess the deforestation at the local level [17].

Mapping land use dynamics of complex landscapes is a particularly difficult task as some land covers have similar spectral characteristics and some change throughout the season. Thus, a multivariate analysis is often required for mapping land cover dynamics [18]. However, when two land cover maps corresponding to two different dates are combined, the individual error of each land cover class is multiplied if one considers that the errors of the two maps are independent [19]. Thus, multi-date analysis reduces classification errors compared to comparing independent classifications produced for multiple dates [20].

In addition to the temporal dimension, studies have shown that auxiliary information can improve the accuracy of land cover classification. At the same time, the inclusion of temporal and geomorphological observations in the classification process can increase the dimensionality of the data. But, the increasing number of input variables to the classification can induce complexity resulting in increased computing time.

The selection of variables is an important step in many machine learning applications. It is not only necessary to obtain more accurate maps, but also to understand which variables are most relevant in the classification process. In this paper, we consider the application of the Random Forest (RF) algorithm for feature selection because of its interesting properties, such as high accuracy and robustness against over-fitting of training data [21]. RF is currently proposed and used to improve land cover mapping from remote sensing imagery [22]. It provides an approach for assessing the importance of characteristics or predictors which can be considered as a useful parameter to study the role of each temporal, spectral or thermal feature in the analysis of old-growth forest reduction discrimination. The purpose of the study was to develop a simple, fast and reproducible method for mapping the deforestation of the old-growth forest using Random Forest.

2. Materials and Methods

2.1. Study area

The study area is the Ituri-Epulu-Aru landscape (IEAL) is located between 2°37'022"N – 0°31'030"N and 27°34'034"E – 30°00'039"E, in Democratic Republic of Congo (Fig. 1). Most of the landscape is situated in Ituri province (in the administrative territories of Mambasa, Irumu and Djugu). Only a smaller portion of the landscape is included in the Haut-Uélé (Wamba and Watsa territories) North-Kivu provinces. The average daily temperature is very stable over the year between 23 and 25.5°C. The average annual rainfall in the landscape fluctuates between 1,600 and 2,000 mm [23]. The driest month of the year has an average rainfall of less than 50 mm in some parts of the landscape. During the dry period, the sky is absolutely cloudless, the relative humidity is low and evapotranspiration is very high [24].

The IEAL is covered by a dense semi-evergreen open-canopy forest, although there is a semi-deciduous forest in the far north-east. The IEAL includes the Okapi Wildlife Reserve (OWR) (13720

km²), the Mai-Tatu Community Reserve (proposed), Enzyme Refines Association (ENRA) private logging concession (520 km²) and three community management areas: Banana (575 Km²), Andekau (6973 Km²) and Bakwanza (2181 Km²) [23]. Also, a reducing emissions from deforestation and degradation (REDD+) pilot project has been implemented in the Mambasa region.

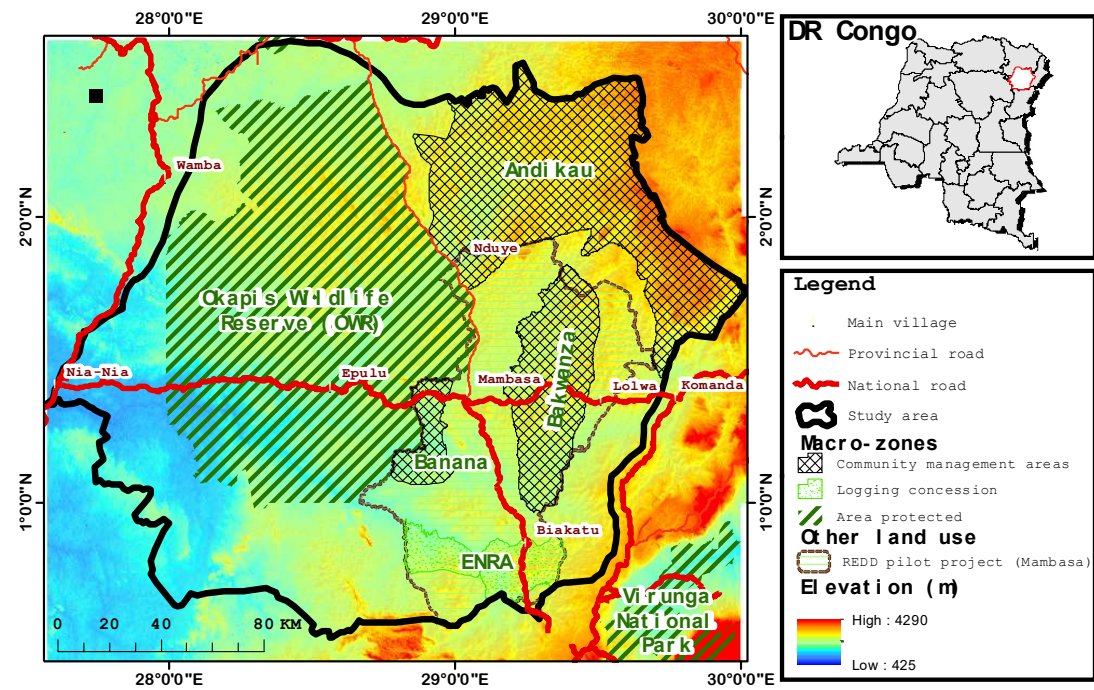


Figure 1. Location and elevation data of study area. The black lines indicate the limit of the IEAL

2.2. Methods

Figure 2 schematically illustrates the methodology used in this study to generate and test the algorithm. It presents the data and their preprocessing, the variables used for the construction of the model, the random sampling of learning datasets, the random forest classifier (RFC) and four land cover classification models.

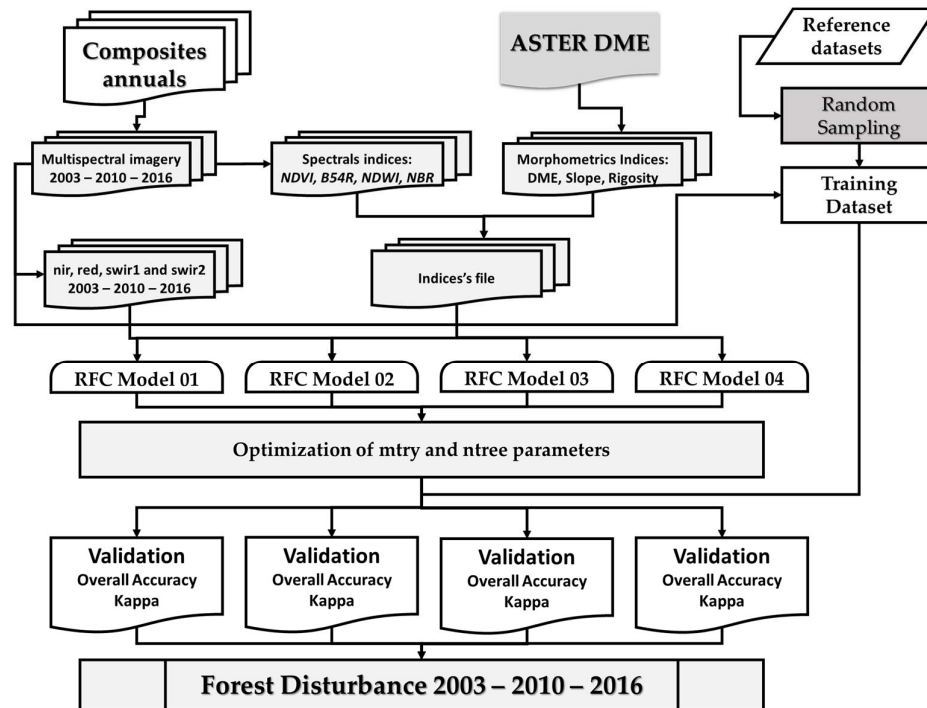


Figure 1. Overview of the analysis framework.

2.2.1. Remote sensing data and preprocessing

The satellite images used to map deforestation in the IEAL are annual image composites of 0.025-degree resolution obtained from the Central Africa Regional Program for the Environment (CARPE). The original images are from the Landsat Thematic Mapper (TM), the Enhanced Thematic Mapper (ETM+) and the Operational Land Imager (OLI) sensors. These composites were formed from four spectral bands: nir (0.845-0.885 μm), red (0.630-0.680 μm), SWIR1 (1.560-1.660 μm) and SWIR2 (2.100-2.300 μm) that had undergone atmospheric, radiometric and geometric corrections. The CARPE composites [13] were chosen as they were free of cloud cover, they cover all Congo Basin countries and can be downloaded free of charge from the CARPE website (www.carpe.umd.edu/). These images are organized in square tiles of one degree by one degree. In this study, we used 10 tiles per date. The date range for each composite was less than one year.

The Congo Basin Forest Partnership (CBFP) landscape was created in 2002 and development work began in 2003 in the IEAL [25]. Hence, 2003 was selected as the reference date. In addition, 2016 is the last year CBFP data are available in the IEAL. To ensure the multi-temporal comparability, the images were re-projected in the same reference coordinate system: WGS 84 - UTM zones 35 North. Then, for each spectral band, a mosaic of tiles was created for the selected years. Radiometric offsets due to differences in acquisition dates were minimized by applying histogram equalization. Finally, the mosaics were composited into a single 2003, 2010, 2016 multi-date file containing twelve bands (fig. 3).

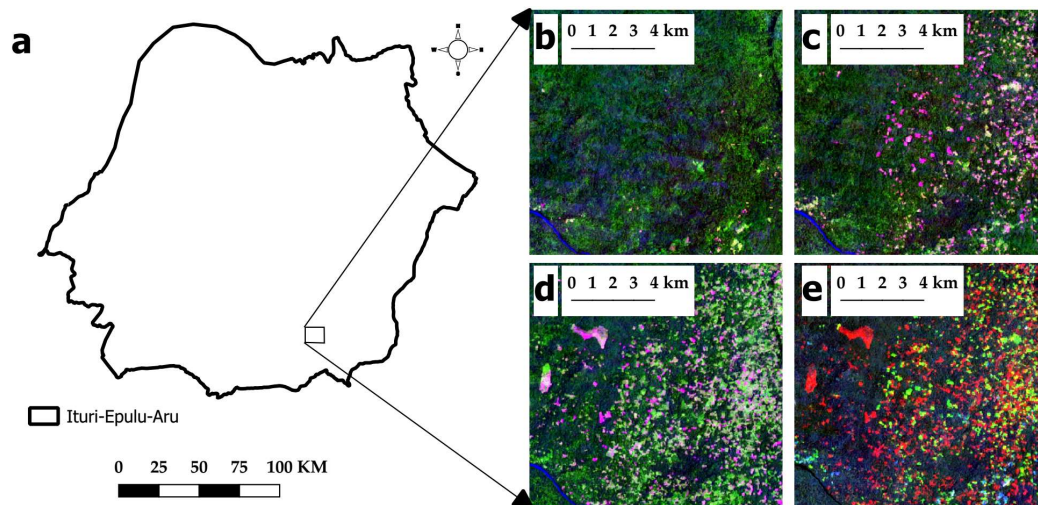


Figure 3. a) The Ituri-Epulu-Aru landscape; (b) Color composition in 2003 (R: swir1 - 2003, G: nir - 2003, B: red - 2003); (c) Color composition in 2010 (R: swir1 - 2010, G: nir - 2010, B: red - 2010); (d) color composition in 2016 (R: swir1 - 2016, G: nir - 2016, B: red - 2016); (e) Multidate color composition (R: red - 2003, G: red - 2010, B: red - 2016).

2.2.2. Spectral and geomorphological indices

The variables used to improve the mapping of old-growth forest depletion between –years were of three types: spectral bands, spectral indices and geomorphological indices. The spectral indices were calculated for each date to improve the discrimination of the different classes of vegetation. Based on previous studies of land-cover change [26-30], the following spectral indices were calculated: normalized difference vegetation index (NDVI), normalized difference moisture index (NDMI), Band5 / Band 4 ratio (B54R), normalized burn ratio (NBR) (Table 1).

Table 1. Spectral indices.

Spectral Index	Formula	Reference
Normalized difference index (NDVI)	$(Q_{NIR} - Q_{red}) / (Q_{NIR} + Q_{red})$	[31]
Normalized difference moisture index (NDMI)	$(Q_{NIR} - Q_{SWIR1}) / (Q_{NIR} + Q_{SWIR1})$	[27]
Band5/Band4 ratio (B54R)	Q_{SWIR1} / Q_{NIR}	[30]
Normalized burn ratio (NBR)	$(Q_{NIR} - Q_{SWIR2}) / (Q_{NIR} + Q_{SWIR2})$	[32]

qred= red reflectance, qNIR= near infrared reflectance, qSWIR= short-wave infrared reflectance.

In addition to the spectral indices, geomorphological indices (elevation, slope, and roughness) derived from a 30-meter Digital Terrain Model (DEM) were introduced to mitigate the relief effect on spectral bands reflectance [33].

2.2.3. Random sample

The training zones were defined on the screen by digitizing several polygons on the multi-date color composition for each spectral class. Thirty percent of the pixels from the 483 polygons was randomly selected for validation of land cover maps.

Table 2. Description of land cover.

Class name	Description	Class code	Number of polygons
Bare ground / Buildings / Rocks since 2003	Bare soil / buildings / rocks unchanged over the period of analysis.	BSNR	35
Other vegetal cover	Secondary forest young and adult / Wasteland / Savannah grass / crop and area little covered/Agricultural complex.	OVC	68
Deforestation between 2003 and 2010	Old-growth forests converted to other occupations from 2003 to 2010.	Df1	133
Deforestation between 2010 and 2016	Primary forests converted to other occupations from 2010 to 2016.	Df2	145
Stable old-growth forests between 2003 and 2016	Primary forests intact between 2003 and 2016.	Fp	24
Water	Water plant.	W	78

In this study, we randomly sampled pixels from more than 483 polygons divided into six classes (an average of 81). These samples represented more than 53995 ha and were distributed in the studied landscape space. The number of training pixels has been fixed at 3500 pixels per class (i.e., 21000 pixels). This number is sufficient to represent variation throughout the image [34].

2.2.4. Classification algorithm for mapping the deforestation

We used the Random Forest classification algorithm [35] available in the *randomForest* package [36] of the R software [37]. Random Forest (RF) is a non-parametric classification algorithm that combines the decision tree algorithm, bagging, and a model aggregation technique [22]. The samples omitted from the bootstrap sample are called samples out of bag (OOB). Out-of-bag samples help to evaluate classification error and the variable significance estimate.

In the RF classifier, two parameters must be defined [38]: the number of decision trees (Ntree) and the number of variables to select for the best distribution during tree growth (Mtry). The literature recommends that Ntree be defined at 500 or greater because the errors tend to stabilize before this number of trees is reached [39]. In this project we set Ntree to 2000. Moreover, the optimal Mtry was defined using the function “train” of the CARET package [40].

2.2.5. Land cover classification models and validation

To map the land use of the area, four models were tested: (i) the first model built on the basis of the image channels (red, nir, swir1 and swir2), (ii) the second model was constructed using image channels and vegetation indices (NDVI, B54R, NDWI, NBR), (iii) the third model was constructed using image channels and geomorphological indices (DEM, Slope and Roughnesses) and (iiii) the forth model combined image channels, vegetative indices and geomorphological indices (Table 3).

Table 3. Models of land cover classification.

Variables	Model 1	Model 2	Model 3	Model 4
Spectral bands	red, nir, swir1, swir2	red, nir, swir1, swir2	red, nir, swir1, swir2	red, nir, swir1, swir2
Spectral index		NDVI, B54R, NDWI, NBR		NDVI, B54R, NDWI, NBR,
Geomorphological index			DEM, Slope, Roughnesses	DEM, Slope, Roughnesses

The performance of each model was calculated using out-of-the-bag (OOB) error [41]. The closer the OOB error value is to zero, the better the model [20]. In addition, a confusion matrix was produced to evaluate the quality of learning. The confusion matrix estimates errors of omission and

commission. It also validates the land cover map by means of the Kappa coefficient and the estimation of the overall accuracy [42; 43].

The post-classification majority filter (3x3 pixel size) was applied to the land cover maps of each model. This majority filter has been applied to eliminate occurrences of isolated pixels and to avoid the "pixelated" effect (pepper and salt).

2.2.6. The annual deforestation rate

The annual deforestation rate is calculated as the ratio of area cleared over a period, divided by the initial forest area and the number of years of the period [44; 45]. However, several publications have recently highlighted that the evolution of forest loss cannot be obtained with this simple report [46]. The standardized formula proposed by Puyravaud [46] was adopted to calculate the annual deforestation rate for this study.

$$\theta = \frac{-1}{t_2 - t_1} \ln \left(\frac{A_2}{A_1} \right) * 100 \quad (1)$$

Where A1= forest area in the initial year; A2= forest area in the final year; t1 = initial year; t2 = final year. ln = natural logarithm

3. Results

3.1. Optimization and validation of classification models

3.1.1. Number of trees and Mtry optimization

The evolution of the OOB error was a function of the increase in the number of trees (Fig. 4). For all models, the OOB error decreases with increasing number of trees and stabilized before the number of trees reaches 500.

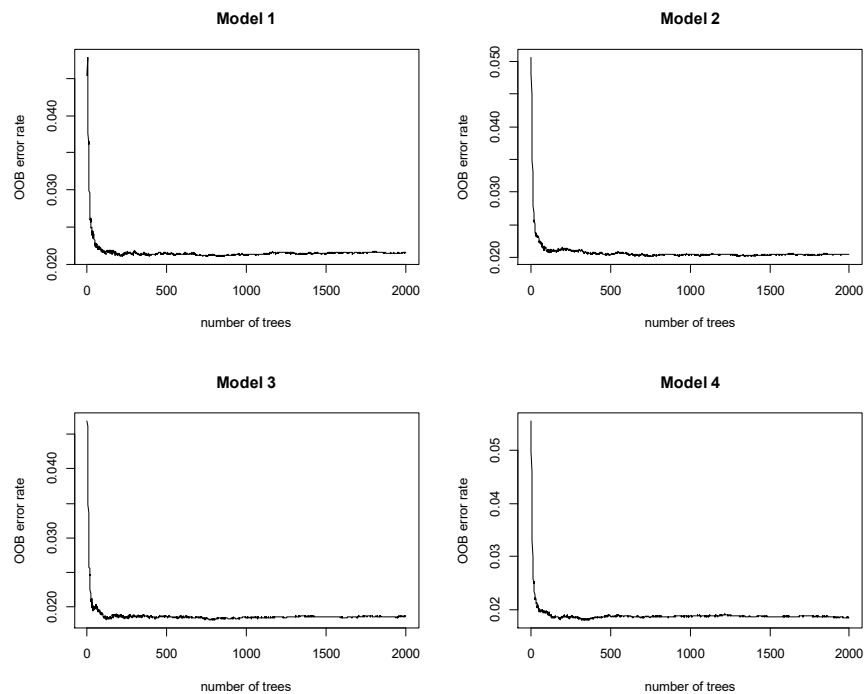


Figure 4. Evolution of OOB according to the increase of the number of trees.

Also, the four models showed the same behavior when it comes to optimizing the number of variables to select for the best distribution during tree growth. All the models proposed three

different optimal Mtry values (Table 4). The best OOB error for all four models was obtained when the Mtry value was 2. Models 2 and 4 presented the smallest values of the OOB error (0.021). The largest OOB error was obtained in the model 1 (0.028, with Mtry equal to 12) (Table 4).

Table 4. Mtry optimization.

Model	Mtry	Accuracy	OOB	Kappa
Model 1	2	0.976	0.024	0.971
	7	0.974	0.026	0.969
	12	0.972	0.028	0.966
Model 2	2	0.979	0.021	0.975
	8	0.978	0.022	0.974
	15	0.975	0.025	0.97
Model 3	2	0.977	0.023	0.973
	13	0.976	0.024	0.971
	24	0.973	0.027	0.967
Model 4	2	0.979	0.021	0.974
	14	0.978	0.022	0.974
	27	0.975	0.025	0.97

3.1.2. Validation of classification models of land cover

The results of the various tests performed on three groups of variables are presented in Table 5. The conclusion was that model 1 created using only spectral bands had a larger OOB error (2.15%). When we combined the spectral bands with the radiometric indices (NDVI, NIRI and NDWI), the model improved very slightly by 2.05%. The same was true when we combined the spectral bands with the geomorphological indices (1.86%). In addition, the comparison of model 2 and model 3 revealed that the information from geomorphological indices improved the model more than those from spectral indices. Model 4 had the lowest OOB error (1.85%). Therefore, model 4 was selected to map the dynamics of deforestation in the study period 2003 – 2009 and 2009 - 2016.

Table 5. OOB and kappa of each model.

	Model 1	Model 2	Model 3	Model 4
Variables	Spectral bands	red, nir, swir1, swir2	red, nir, swir1, swir2	red, nir, swir1, swir2
	Spectral index	NDVI, B54R, NDWI, NBR		NDVI, B54R, NDWI, NBR,
	Geomorphological index		DEM, Slope, Roughnesses	DEM, Slope, Roughnesses
	OOB	2.15	2.05	1.86
	Kappa coefficient	0.9784	0.9795	0.9814
				0.9815

Table 6 presents model 4 validation using the confusion matrix. The Kappa coefficient was 98.15% and the overall accuracy was 98.15%.

Table 6. Matrix of confusion of Model 4.

Model 4		Ground Truth ROIs						User Accuracy (%)
		OVC	BSNR	Df1	Df2	Fp	Water	
Classification	OVC	3374	3	57	19	47	0	96.40
	BSNR	7	3491	2	0	0	0	99.74
	Df1	36	1	3363	82	18	0	96.09
	Df2	13	1	48	3426	12	0	97.89
	Fp	30	0	3	0	3466	1	99.03
	Water	0	1	0	1	6	3492	99.77
Producer Accuracy (%)		97.51	99.83	96.83	97.11	97.66	99.97	Kappa coefficient =0.9778
		Overall Accuracy= 98.15%						

OVC: Other vegetal cover; BSNR: Bare ground / Buildings / Rocks since 2003; Df1: Deforestation between 2003 and 2010; Df2: Deforestation between 2010 and 2016; Fp: Stable old-growth forests between 2003 and 2016; Water: Water.

3.2. Deforestation trends in the IEAL

Statistics on forest cover trends are presented in Figure 5. The landscape was dominated by old-growth forests. They account for over 87.67% of the total landscape in 2016. Other stable vegetal covers accounted for almost 6.31% representing more than 259032 ha. The Bare ground / Buildings / Rocks land cover since 2003 represented less than one percent (16.2 ha). Water accounted for 0.38% of the studied landscape (15662 ha).

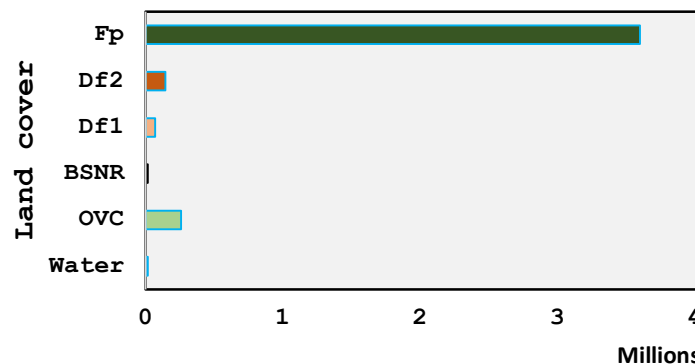


Figure 5. Land cover composition between 2003 – 2010 and 2010– 2016. OVC: Other vegetal cover; BSNR: Bare ground / Buildings / Rocks since 2003; Df1: Deforestation between 2003 and 2010; Df2: Deforestation between 2010 and 2016; Fp: Stable old-growth forests between 2003 and 2016; Water: Water.

Table 7 presents the area of old-growth forest cover and deforestation rates observed in the IEAL and in the neighboring eastern highlands. These results suggest that the total area of the old-growth forests in 2016 was 3601608 ha. In the whole, more than 69861 ha were lost during the 2003-2010 period, i.e, 9980 ha per year and approximately 145768 ha for the 2010-2016 period, i.e., 24295 ha per year. This result showed that the annual rate of deforestation increased from 0.03% –in the first period to 0.66% in the second. However, this forest loss was differently distributed within the IEAL. The Banana Community Management area had a relatively constant annual deforestation rate over the two studied periods (0.54%), in contrast with the sharp increase for the other macro zones and in the REDD + Mambasa intervention zone as shown in Table 7. While the ENRA and secondly the REDD + intervention zone had already a high deforestation rate for the 2003-2010 period (3.05% and 0.70% respectively to compare with 0.02%, 0.07% and 0.54% for the three community areas, Andikau, Bakwanza and Banana respectively), the first two communities of Andikau and Bakwanza did see

important relative growth for the 2010-2016 period with a multiplicative factor of 41.5 and 5.1 respectively only to join the average growth of the area (ENRA: 3.20%, REDD+: 1.23%; Andikau: 0.83%; Bakwanza: 0.36%; Banana: 0.54% stable).

Table 7. Area and annual rates of deforestation by macro-zones for the two periods observed: 2003 - 2010 and 2010 - 2016.

Location	Area of old-growth forests			Annual rate of deforestation % (deforested area in ha)	
	2003	2010	2016	2003 – 2010	2010 – 2016
Andikau	613751	612876	583046	0.02 (874)	0.83 (29830)
Bakwanza	210350	209310	204868	0.07 (1040)	0.36 (4442)
Banana	50752	48855	47285	0.54 (1896)	0.54 (1570)
ENRA	76974	62171	51319	3.05 (14803)	3.20 (10851)
Redd+ ¹	792437	754560	700857	0.70 (37877)	1.23 (53702)
OWR	1350113	1344970	1337097	0.05 (5143)	0.10 (7872)
Total	3817237	3747376	3601607	0.26 (69861)	0.66 (145768)

4. Discussion

4.1. Adjustment of models and importance of variables

For all models designed in this study, the Kappa coefficient is greater than 80% and the OOB does not reach 2.2% [47]. The four tested models showed little difference in value of the Kappa coefficient and the OOB error (a difference of 0.3% and 0.0031, respectively).

Many studies found that land cover classes can be confounded. For example, Kabuanga [48] showed that the young secondary forest can easily be confounded with the brownfields and old fallows. Elsewhere, AECOM [49] showed some possible confounding between savanna vegetation, areas with little vegetation and crop areas. In our study, however, the obtained results showed that the six land cover classes were not confounded. These results are corroborated by those by Mikwa et al. [50] who found that although the old-growth forests are spectrally similar to secondary forests, their respective discrimination is still possible if the sampling effort is substantial. Thus, merging of secondary forests, wastelands, savannas and agricultural complexes into one class has eliminated the confounding between classes, which in the end impacted the classification model of land use. In the same way, the built-up areas showed spectral signatures close to those of rocks. For this reason, these two classes have been merged. As usual, the water class was clearly distinguishable from other classes [20].

The temporal aspect also played a clear role. More than 90% of the landscape remained stable over the entire observation period. Of the selected six classes, only two (Df1 and Df2) showed spatio-temporal changes expressing deforestation of the old-growth forests. These two classes stood out clearly on the colorful multi-date composition. Image-to-image rectification allowed an objective analysis of the same objects.

Variable importance is presented in Figure 6. The results show that in all four models, the spectral bands play a large role in the mapping of land use. The swir1, swir2, nir and red domains are widely used in land cover mapping [51; 26-30]. The SWIR domain plays a leading role in improving classification models. From a temporal point of view, the 2016 bands occupy the first position in terms of importance. SWIR2 seems to play first in models where geomorphological indices are absent.

The addition of geomorphological indices provided supplementary information and changes the influence of the variables. Indeed, in models made up of geomorphological indices, SWIR1 took the first place in terms of importance. In contrast, in models where the spectral indices are absent, the nir was less important. Also, the geomorphological indices seemed to play a less important role. They improved the total performance of the models by bringing non-redundant information to the models as earlier found by Huo et al. [51] and Oeser et al. [26]. The unique aspect of this study lies in the use

of topography information which mitigates the relief effect on the spectral bands reflectance [33]. In this regard, the DEM was the most important geomorphological variable.

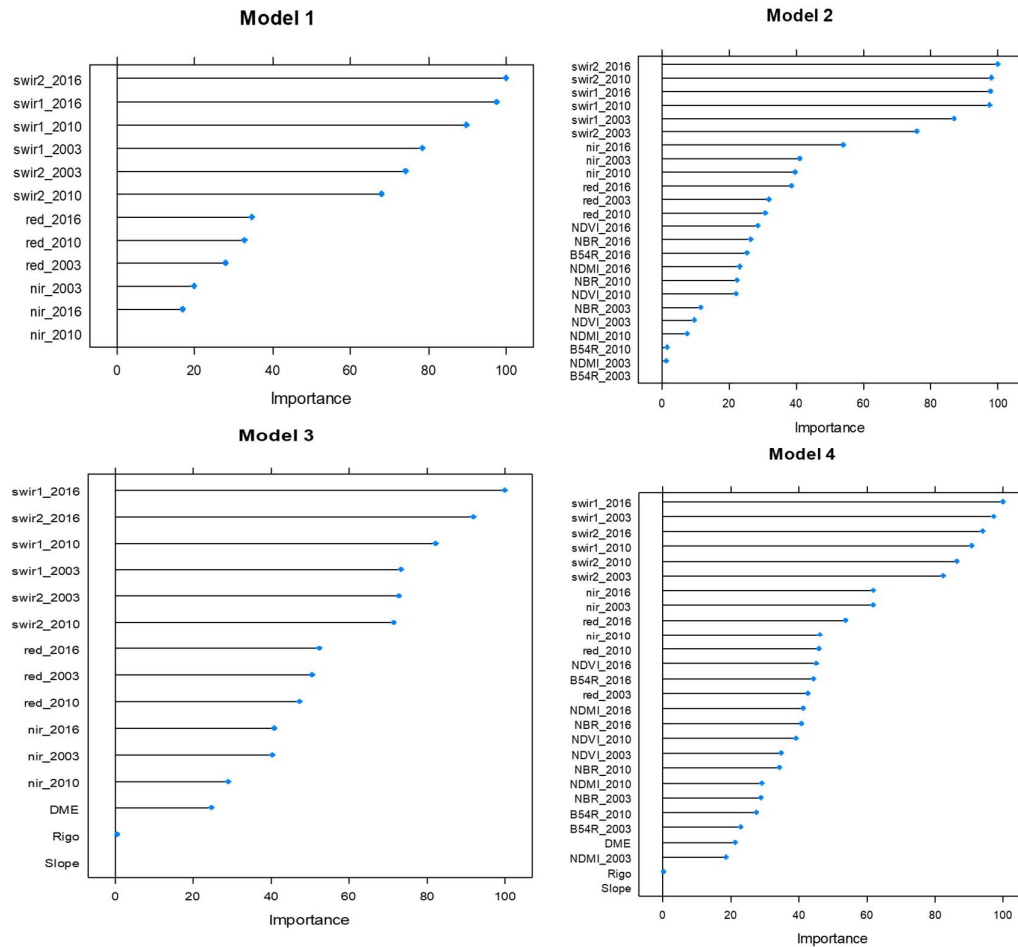


Figure 6. Importance of the variables in each model.

4.2. History of deforestation in the IEAL

The comparison of the results obtained in this study with those of other similar studies should be carried out with caution since the methodologies and data are not compatible (data, methodology, observation periods, post-classification treatments). Lusana et al. [16] estimated the old-growth forests area as 40,492 km² in 2003 and 39,977 km² in 2010, meaning a loss of 51,514 ha between 2003 and 2010, including secondary forest, which explains why their estimations are higher to ours for the same dates. The FACET [52] estimates it as 38,432 km² for 2010. However, both studies unravel rising rates of deforestation.

5. Conclusions

This paper assessed the potential of multi-spectral remote sensing observations for the classification of land use using RFC technique. The study area presents specific characteristics and various types of natural and artificial surfaces. The key lesson learnt from our study is that the combination of several types of information (spectral, temporal and topographic data) improves the results. RFC offers a wide range of possibilities regarding variable significance which guides users to important variables or features. Using the RFC importance measures reduces the dimensionality of

the input features. RFC importance variables showed how multi-date spectral and topographic characteristics have the greatest influence on class separability in the studied area.

These results showed that the RFC model was able to effectively map the reduction of old-growth forests. This study also highlighted the importance of topographic data in addition to multispectral imagery for the purpose of land cover classification. Measuring the importance of variables was another advantage of random forests that was successfully applied in this study. The optimization of the RF was relatively easy because it considers only two parameters (ntrees and Mtry).

In developing countries, several constraints make it difficult to study changes in land use. The use of open access software can allow the reproduction of this method without legal constraints. For this reason, the script used is detailed and entirely reproduced in the supplemental material to facilitate the implementation of RFC.

Author Contributions: Conceptualization, J.M.K., O.M.K., and L.M.; methodology, N.H., M.S., N.M., and J.M.K.; software, J.M.K., and N.H.; validation, N.H., O.M.K., and L.M.; formal analysis, J.M.K., and O.M.K.; investigation, J.M.K.; resources, J.M.K., and L.M.; data curation, J.M.K.; writing—original draft preparation, J.M.K., O.M.K., N.H., M.S., N.M., F.M.I., and L.M.; writing—review and editing, J.M.K., O.M.K., N.H., M.S., N.M., F.M.I., and L.M.; visualization, J.M.K.; supervision, O.M.K., L.M., and F.M.I.; project administration, O.M.K., and J.M.K.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We warmly thank Victor KADIATA for having the Satellite Observatory of the Forests of Central Africa (SOFCA) for ensuring the download of satellite images. We sincerely thank Alastair McNeilage and Toussaint Molenge (CARPE) for recommending us to SOFCA. Big thanks to Salomon Dyoma Ezadri, Alimengo Paul, Atimo Michel, Kimate Henry, Landasi Tsavurume, M'vuge Bijoux, Mamimolo Jean-Marie, Mumbere Jackson, Ngenda Elvis, Munde Oscar, Yambwa Gloire for assistance in the field. We also thank the former Chief Curator / Site Manager of the Okapis Wildlife Reserve and Directors of Parks and Reserves in DR Congo Paulin TSHIKAYA and the Curator / Head of Research and Biomonitoring Program at the Okapis Wildlife Reserve BALIMBAKI Aimé for access to the study site. We thank ROSEMARIE RUF for its support and Franchesco Escobar for proofreading the Manuscript.

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

References

1. Iyongo W.M.L.; Visser, M.; De Cannière, C.; Verheyen, E.; Dudu A.B.; Ulyel A.-P.J.; Bogaert, J. 2012. Anthropisation et effets de lisière: impacts sur la diversité des rongeurs dans la Réserve Forestière de Masako (Kisangani, R.D. Congo). *Tropical Conservation Science*. 5 (3), 270-283. <https://doi.org/10.1177/194008291200500304>
2. Gibson, L.; Lee, T.M.; Koh, L.P.; et al. 2011. Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature*, 478, 378–381 <https://doi.org/10.1038/nature10425>
3. Brown, S., Gaston, G., 1996. *Estimates of biomass density for tropical forests*. In: Levine, J.S. (Ed.), *Biomass Burning and Global Change: Remote Sensing, Modeling and Inventory Development, and Biomass Burning in Africa*. The MIT Press, Cambridge, USA, pp. 133–139.
4. Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice et J. R. G. Townshend, 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change, *Science*, 342: pp. 850-53, <https://doi.org/10.1126/science.1244693>
5. Bamba, I., M. S. Yedmel, et J. Bogaert, 2010, Effets des routes et des villes sur les forêts denses dans la province Orientale de la République Démocratique du Congo, *European Journal of Scientific Research*, 43, pp. 417-429.
6. Curtis, P.B.; Slay, C.M.; Harris, N.L.; Tyukavina, A.; Hansen, M.C., 2018. Classifying drivers of global forest loss. *Science* 6407, 361, 1108-1111. <http://dx.doi.org/10.1126/science.aau3445>

7. Song, X.P.; Hansen, M.C.; Stehman, S.V.; Potapov, P.V.; Tyukavina, A.; Vermote, E.F.; Townshend J.R., 2018. Global land change from 1982 to 2016. *Nature*, 560, 639–643. <https://doi.org/10.1038/s41586-018-0411-9>
8. Kabuanga J.M., Guguya A.B., Okito E.N., Maestripiéri N., Saqalli M., Rossi V., et Iyongo W.M.L., 2020. Suivi de l'anthropisation du paysage dans la région forestière de Babagulu, République Démocratique du Congo, *VertigO*, Vol 20 n° 2, vertigo.28347 <https://doi.org/10.4000/vertigo.28347>
9. FAO, 2015. Forest resources assessment 2015. Global Synthesis, FAO, Rome.
10. Lindquist, E. J.; Hansen, M. C.; Roy, D.P.; Justice, C.O., 2008. The suitability of decadal image data sets for mapping tropical forest cover change in the Democratic Republic of Congo: implications for the global land survey. *Int. J. Remote Sens.*, 29, 7269–7275. <https://doi.org/10.1080/01431160802275890>
11. Hansen, M. C.; Stehman, S. V.; Potapov, P. V.; Loveland, T. R.; Townshend, J. R. G.; DeFries, R. S.; Arunarwati, B.; Stolle, F.; Steininger, M.; Carroll, M.; DiMiceli, C., 2008. Humid tropical forest clearing from 2000 to 2005 quantified using multi-temporal and multi-resolution remotely sensed data. *Proceedings of the National Academy of Sciences*, 105, 9439–9444. <https://doi.org/10.1073/pnas.0804042105>
12. Achard, F.; DeFries, R.; Eva, H.; Hansen, M.; Mayaux, P.; Stibig, H-J., 2007. Improved pan-tropical observations and mid-resolution monitoring of deforestation, *Environmental Research Letters*, 2, 11. <http://dx.doi.org/10.1088/1748-9326/2/4/045022>
13. Turubanova S, Peter V Potapov, Alexandra Tyukavina and Matthew C Hansen. 2018. Ongoing primary forest loss in Brazil, Democratic Republic of the Congo, and Indonesia. *Environ. Res. Lett.* 13 074028. <https://doi.org/10.1088/1748-9326/aacd1c>
14. Duveiller, G., Defourny, P., Desclée, B. and Mayaux, P., 2008. Deforestation in Central Africa: Estimates at regional, national and landscape levels by advanced processing of systematically-distributed Landsat extracts. *Remote Sens. Environ.*, 112(5): 1969–1981 <https://doi.org/10.1016/j.rse.2007.07.026>
15. Tyukavina, A.; Hansen, M.C.; Potapov, P.; Parker, D.; Okpa, C.; Stehman, S.V.; Kommareddy, I.; Turubanova, S., 2018. Congo Basin forest loss dominated by increasing smallholder clearing. *Science advances*, 4 (11), eaat2993. <http://dx.doi.org/10.1126/sciadv.aat2993>
16. Lusuna, M.; Kibambe J.-P.; Makana, J.-R.; Mwinyihali, R., 2014. *Evaluation par Télédétection de l'Etendue de la Couverture Forestière dans le Paysage Ituri-Epulu-Aru (2000-2014)*. Technical Report WCS-PCFI.
17. Kyale Koy J., Wardell D. A., Mikwa J.-F., Kabuanga J. M., Monga Ngonga A. M., Oszwald J., Doumenge C., 2019. Dynamique de la déforestation dans la Réserve de biosphère de Yangambi (République démocratique du Congo): variabilité spatiale et temporelle au cours des 30 dernières années. *Bois et Forêts des Tropiques*, n° 341: 13–28. <https://doi.org/10.19182/bft2019.341.a31752>
18. Eastman, J.R., Sangermano, F., Ghimire, B., Zhu, H., Chen, H., Neeti, N., Cao, Y., and Crema, S., 2009. Seasonal Trend Analysis of Image Time Series, *International Journal of Remote Sensing*, 10:pp 2721–2726.
19. Fuller, R. M., Smith, G. M., and Devereux, B. J. 2003. The characterization and measurement of land cover change through remote sensing: Problems in operational applications? *International Journal of Applied Earth Observation and Geoinformation*, 4: pp 243–253.
20. Grinand, C., Rakotomalala, F., Gond, V., Vaudry, R., Bernoux, M., and Vieilledent, G. 2013. Estimating deforestation in tropical humid and dry forests in Madagascar from 2000 to 2010 using multi-date Landsat satellite images and the random forests classifier. *Remote Sensing of Environment*, 139: 68–80. <http://dx.doi.org/10.1016/j.rse.2013.07.008>
21. Díaz-Uriarte, R., Alvarez de Andrés, S. Sélection de gènes et classification des données de puces à ADN en utilisant la forêt aléatoire. *BMC Bioinformatics* 7, 3 (2006). <https://doi.org/10.1186/1471-2105-7-3>
22. Pal, M., 2005. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* 26, 217–222. <http://dx.doi.org/10.1080/01431160412331269698>
23. Brown, E.L.; Mwinyihali, R.; Hart, J.; Ngungu, B.; Wells, M.; Makambo, E., 2008. *Ituri-Epulu-Aru*. In: de Wasseige, C.; de Marcken P.; Bayol, N.; Hiol, F.; Mayaux, Ph.; Desclée, B.; Nasi, R.; Billand, A.; Defourny,

- P.; Eba'a. Les forêts du bassin du Congo – Etat des forêts, Office des publications de l'Union Européenne, Luxembourg. . http://www.observatoire-comifac.net/docs/edf2008/FR/EDF_2008_FR_25.pdf
24. Brown, E.L., 2010. *Okapi Faunal Reserve, Ituri-Epulu-Aru landscape, Democratic Republic of Congo*. In: Yanggen, D.; Angu, K.; Tchamou, N. *Landscape-scale Conservation in the Congo Basin: Lessons Learned from the Central African Regional Program Environment (CARPE)*. Gland Switzerland: IUCN, xiv+262p. <https://portals.iucn.org/library/efiles/documents/2010-037.pdf>
 25. Yanggen, D.; Angu, K.; Tchamou, N., 2010. *Landscape-scale Conservation in the Congo Basin: Lessons Learned from the Central African Regional Program Environment (CARPE)*. Gland Switzerland: IUCN, xiv+262p.
 26. Oeser, J.; Pflugmacher, D.; Senf, C.; Heurich, M.; Hostert, P., 2017. Using intra-annual Landsat time series for attributing forest disturbance agents in Central Europe. *Forests.*, 8, 251 <http://dx.doi.org/10.3390/f8070251>
 27. Coops, N.C.; Johnson, M.; Wulder, M.A.; White, J.C., 2006. Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. *Remote Sens. Environ.* 103, 67–80. <http://dx.doi.org/10.1016/j.rse.2006.03.012>
 28. Wilson, E.H.; Sader, S.A., 2002. Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sens. Environ.* 80, 385–396. [http://dx.doi.org/10.1016/S0034-4257\(01\)00318-2](http://dx.doi.org/10.1016/S0034-4257(01)00318-2)
 29. Cohen, W.B.; Fiorellad, M.; Gray, J.; Helmer, E.; Anderson, K., 1998. An efficient and accurate method for mapping forest clearcuts in the Pacific Northwest using Landsat imagery. *Photogramm. Eng. Remote Sens.* 62, 1025–1036. http://myweb.facstaff.wvu.edu/~wallin/envr442/pdf_files/cohen_etal_1998_apr_293-300.pdf
 30. Vogelmann, J.E., 1990. Comparison between 2 vegetation indexes for measuring different types of forest damage in the north-eastern United States. *Int. J. Remote Sens.* 11, 2281–2297. <http://dx.doi.org/10.1080/01431169008955175>
 31. Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8 127–150. [http://dx.doi.org/10.1016/0034-4257\(79\)90013-0](http://dx.doi.org/10.1016/0034-4257(79)90013-0)
 32. Key, C.H.; Benson, N.C., 2006. *Landscape Assessment: Ground Measure of Severity, the composite Burn Index; and Remote Sensing of Severity, the Normalized Burn Ratio*; General Technical Report RMRS-GTR-164-CD; USDA Forest Service, Rocky Mountain Research Station: Ogden, UT, USA.
 33. Tachikawa, T.; Manabu K.; Akira I., 2011. *ASTER GDEM Version 2 Validation Report*. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1 24.
 34. Tatsumi, K.; Yamashiki, Y.; Torres, M.A.C.; Taïpe, C.L.R., 2015. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Comput. Electron. Agric.*, 115, 171–179 <http://www.sciencedirect.com/science/journal/01681699>
 35. Breiman, L., 2001. Random Forests. *Mach. Learn.*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
 36. Liaw, A.; Wiener, M., 2002. Classification and Regression by randomForest. *R News.* 2, 18–22.
 37. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
 38. Strobl, C.; Hothorn, T.; Zeileis, A., 2009. Party on! A new, conditional variable importance measure for random forests available in the party package. *R Journal*, 1, 14–17. https://journal.r-project.org/archive/2009-2/RJournal_2009-2_Strobl-et-al.pdf
 39. Belgiu, M.; Dragut, L., 2016. Random forests in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>

40. Kuhn M (2020). *caret: Classification and Regression Training*. R package version 6.0-86. <https://CRAN.R-project.org/package=caret>
41. Bylander, T. Estimating Generalization Error on Two-Class Datasets Using Out-of-Bag Estimates. *Mach. Learn.* 48, 287–297 (2002). <https://doi.org/10.1023/A:1013964023376>
42. Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognit. Lett.* 27, 861–874. <http://dx.doi.org/10.1016/j.patrec.2005.10.010>
43. Landis, J.R.; Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics.*, 33, 159–174. <http://dx.doi.org/10.2307/2529310>
44. Menon, S., and Bawa, K.S., 1997. Application of geographic information systems, remote sensing, and landscape ecology approach to biodiversity conservation in Western Ghats. *Curr.Sci.*, 75:pp 228 -285
45. Ramesh, B.R., Menon, S., and Bawa, K.S., 1997. A Vegetation based approach to biodiversity gap analysis in the Agastyamalai region, Western Ghats, India. *Ambio*, 26: pp 536-539
46. Puyravaud, J. P., 2002, Standardizing the calculation of the annual rate of deforestation, *Forest Ecology and Management*, 177: pp 593–596, [https://doi.org/10.1016/S0378-1127\(02\)00335-3](https://doi.org/10.1016/S0378-1127(02)00335-3)
47. Pontius, R.G.Jr. 2000. Quantification error versus location in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing*, vol. 66, 8 :pp 1011-1016.
48. Kabuanga, JM., 2015. *Analyse prospective de la déforestation estimée par télédétection dans la réserve de biosphère de Yangambi (Tshopo, RDC)*. Travail de fin d'étude. Faculté de gestion des ressources naturelles renouvelables, Université de Kisangani, RD Congo 68p.
49. AECOM, 2017. *Suivi de la déforestation le long des routes nationales*. Rapport 3. AECOM/MITPR, Kinsaha, 97p.
50. Mikwa, N.; Kabuanga M.; Anitambua S.; Kahindo M.; Nshimba SM., 2016. Analyse prospective de la déforestation estimée par télédétection dans la réserve de biosphère de Yangambi (Tshopo, RDC). *Int. J. of Innovation and Scientific Research.* 24, 236-254. <http://www.issr-journals.org/links/papers.php?journal=ijisr&application=pdf&article=IJISR-16-024-10>
51. Huo, L.-Z.; Boschetti, L; Sparks, A. M., 2019. Object-Based Classification of Forest Disturbance Types in the Conterminous United States. *Remote Sens.* 11, 477; <http://dx.doi.org/10.3390/rs11050477>
52. FACET (Forêts d'Afrique Centrale Evaluées par Télédétection). 2010. *Étendue et perte du couvert forestier en République démocratique du Congo de 2000 à 2010*. Université d'État du Dakota du Sud, Brookings, Dakota du Sud, USA, 66 p.