Remote Sensing Applications: Society and Environment 26 (2022) 100746

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

Mapping Cashew Orchards in Cantanhez National Park (Guinea-Bissau)

Sofia C. Pereira [a,*], Catarina Lopes [b], João Pedro Pedroso [c]

[a] Faculty of Sciences, University of Porto, Rua do Campo Alegre, 4169-007, Porto, Portugal
[b] Remote Sensing Consultant, Lisboa, Portugal
[c] INESC TEC, Faculty of Sciences, University of Porto, Rua do Campo Alegre, 4169-007, Porto, Portugal

ARTICLE INFO

Keywords:
Land cover
Sentinel-2
Sustainable development
Deforestation

ABSTRACT

The forests and woodlands of Guinea-Bissau are a biodiversity hotspot under threat, which are progressively being replaced by cashew tree orchards. While the exports of cashew nuts significantly contribute to the gross domestic product and support local livelihoods, the country's natural capital is under significant pressure due to unsustainable land use. In this context, official entities strive to counter deforestation, but the problem persists, and there are currently no systematic or automated means for objectively monitoring and reporting the situation. Furthermore, previous remote sensing approaches failed to distinguish cashew orchards from forests and woodlands due to the significant spectral overlap between the land cover types and the highly intertwined structure of the cashew tree patches. This work contributes to overcoming such difficulty. It develops an affordable, reliable, and easy-to-use procedure based on machine learning models and Sentinel-2 images, automatically detecting cashew orchards with a dice coefficient of 82.54%. The results of this case study designed for the Cantanhez National Park are proof of concept and demonstrate the viability of mapping cashew orchards. Therefore, the work is a stepping stone towards wall-to-wall operational monitoring in the region.

1. Introduction

Guinea-Bissau is home to globally significant forest, woodland and savanna woodland patches. However, these rich and diverse ecosystems are under severe threat: deforestation has been reported as a major ecological sustainability problem in the country. Overall, the country lost 77% and 10% of its closed-canopy and open-canopy forests, respectively, between 2001 and 2018; since Guinea-Bissau relies heavily on the agricultural sector, deforestation is largely due to the uncontrolled conversion of woodlands into new agricultural land, especially for permanent cashew tree plantations (UN-FCCC, 2019). The selling of high quality cashew nuts to installed commercial networks that export to processing factories, is a means of expeditiously improving the economic situation of both the rural families and of public finances. Cashew nuts are the main source of fast cash for the local population and the country's most exported product, representing a large proportion of its gross domestic product (Hanusch, 2016). It should be noted that the quality and quantity of cashew nuts produced in a given stand starts decreasing after 25 years, while the hydrological equilibrium and productivity of the land may become seriously compromised. Thus, the rampant uncontrolled plantation of cashew, which has been converting the country into a large tree orchard with patches of unknown extent, age, or state, threatens food security in the short-term; decreases land availability and suitability for agriculture in the medium term; and drains natural resources and biodiversity

* Corresponding author.
E-mail address: up201804301@edu.fe.up.pt (S.C. Pereira).

https://doi.org/10.1016/j.rsase.2022.100746
Received 31 December 2021; Received in revised form 25 March 2022; Accepted 27 March 2022
Available online 8 April 2022
2352-9385/

in the not-so-long term (International Finance Corporation, 2010).

Despite its extreme poverty, Guinea-Bissau has invested quite a lot of effort in attempting to conserve its biodiversity and its forests, and the country is part of the United Nations Framework Convention on Climate Change (UNFCCC) (UN-FCCC, 1994). Nevertheless, the country's low levels of education, combined with its political instability makes these policies very hard to implement, and thus Guinea-Bissau has not been able to adequately control, much less halt deforestation and move towards sustainability. Official information regarding cashew production in Guinea-Bissau is made available by the Food and Agriculture Organization (FAO) (FAO, 2020a), but since this information is based on registered transactions (such as exported tons of cashew), it is unreliable for determining the area occupied by orchards. In fact, due to the prevailing unregistered selling of the product and to the lack of objective assessment of the areas occupied by the crop, the numbers provided by official agencies are likely to be severely underestimated. Given these circumstances, land cover monitoring based on satellite remote sensing technology is an essential aid supporting a better assessment of Guinea-Bissau's cashew plantations and production.

Past research was unable to map Guinea-Bissau's cashew orchards using publicly available remote sensing data. More specifically, distinguishing cashew plantations from woodlands was considered unfeasible using Landsat data Temudo and Abrantes (2014). Motivated by the difficulties faced in solving this task in the past, this study aims at assessing the feasibility of mapping cashew orchards using publicly available remote sensing data with better spatial and spectral resolution, exploring the usefulness of textural information, and relying on more advanced classification algorithms than those used by past studies. In order to assess the feasibility of this methodology, our goal was to develop a regional land cover map that accurately depicts cashew orchards in the Cantanhez National Park (Guinea-Bissau). Such a product allows a detailed analysis of the extent and spatial distribution of the orchards, while indirectly providing valuable information on the severity of deforestation in the region. The proposed method is more practical than extensive field data collection; considerably less costly; faster; and more complete, also allowing wall-to-wall frequent coverage, while holding great potential for automation (Talukdar et al., 2020). Importantly, we advance the existing state-of-the-art by showing that mapping this type of land cover with publicly available data is possible, opening the door for multi-year multi-region mapping of cashew orchards in the future.

2. Study area and data

This study was developed in the Cantanhez National Park (Fig. 1), a protected area in the Tombali region, covering an extent of more than 1000 km^2 and located in the southern region of Guinea-Bissau. The region contains patches of forest, mangrove, savanna woodland, agricultural fields and settlements (Sousa et al., 2017). Like many other places in the country, it is home to a significant amount of biodiversity that includes endemic and/or endangered species. Deforestation for agricultural purposes greatly endangers

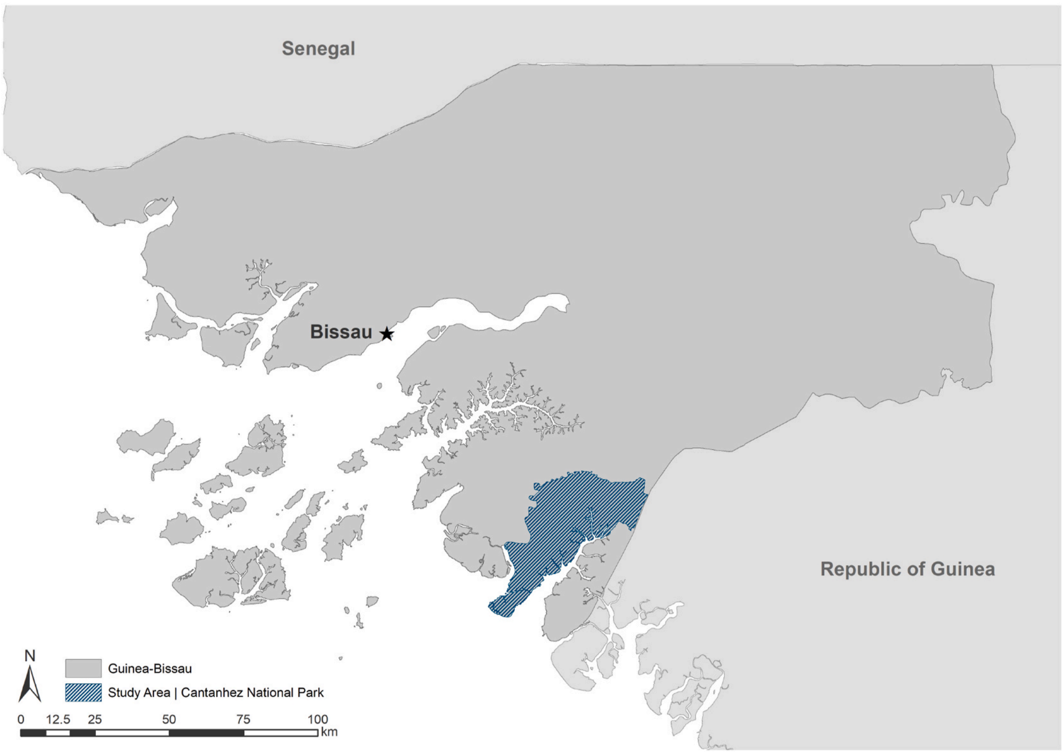

Fig. 1. Cantanhez National Park (highlighted in blue) in Guinea-Bissau.(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the conservation of these species and the fact that these agricultural areas are intertwined with forest patches can make the monitoring of this situation difficult (Sousa et al., 2017). Sentinel-2 (S-2) (Sentinel-2 MSI, 2020) provides a global coverage of the Earth's land surface every 5 days using the two similar polar-orbiting satellites *Sentinel 2A* (launched on the 23^{rd} June 2015) and *Sentinel 2B* (launched on the 7^{th} March 2017), placed in orbits with an 180° lag. The use of two satellites aims at minimizing the satellite's revisiting time, which is approximately 5 days using this setup. The mission offers high-resolution optical imagery with a spatial resolution of 10 m, 20 m and 60 m and multi-spectral data with 13 bands in the visible, near-infrared and short-wave infrared parts of the spectrum.

For this study, S-2 raster data were obtained using Google Earth Engine (GEE) (Gorelick et al., 2017), a cloud-based platform that combines open-source data catalogs of satellite imagery with computing power optimized for parallel processing of geospatial data. The 20 m and 60 m bands were resampled to a 10 m resolution. Bands 1, 9 and 10 were discarded because they are not meant to be used in remote sensing of the land (they target atmospheric conditions) and are located outside of atmospheric windows. Band 8A was also excluded due to its spectral overlap with band 8, which could cause redundancy. The processing Level-2A of S-2, available in GEE, is used in this study to generate the land cover classification. The Level-2A product output is composed of 100×100 km ^{2} tiles, with surface reflectance data. During the drier months of the year, the amount of cloud coverage is smaller. Also, the spectral overlap between different types of vegetation decreases, which facilitates the distinction between cashew and forest areas. In Guinea-Bissau, this period ranges from late November to mid May (Vasconcelos et al., 2015). To ensure that the satellite images used do in fact correspond to the dry season of the year, they were all selected from the period ranging from January 2019 to April 2019.

The reference data set made available for this study was produced through the delineation of polygons over high resolution imagery available in Google Earth. The information was collected for the same time period of the used satellite imagery, over the Region of Interest (ROI), using visual interpretation and expert field knowledge. The land cover classes considered for labelling the polygons are those shown in Table 1.

3. Methods

3.1. Satellite imagery

Multiple S-2 tiles can be combined into a mosaic, so that the full ROI is covered. Tiles representing the same location can also be combined (composited) into a single image using an aggregation function (such as the mean, median, minimum or maximum). Depicting the temporal dynamics of the vegetation often helps the classifier to distinguish between the different classes. For this reason, instead of considering a single image composite, combining all the images ranging from January 2019 to April 2019, four composites (one per each of the months being considered) were produced for this study. The four monthly mosaics (January 2020, February 2020, March 2020 and April 2020) were generated by compositing all S-2 Level 2A images, of each of the months being considered, spanning over the study area and that contained a percentage of cloudy pixels lower than 30%. The compositing criteria was the median value of all cloud free images taken at different dates of each month, for each band and each pixel. This option was preferred over a time series approach because, due to the high frequency of clouds in the region, a time series would be either noisy (if cloudy data were kept) or contain a lot of missing data (if cloudy data were removed). Compositing with the median creates images with less clouds (which have high intensity values) and shadows (which have low intensity values). Additionally, it provides more stability between composites than using a maximum value (Lopes et al., 2019; Simonetti et al., 2015).

3.2. Ground-truth information

The labeling of the data set includes the following land cover classes: Forest, Cashew, Sparse Vegetation/Savanna, Mangroves, Urban/Bare Soil and Water. This multi-class setup allows the understanding of the spatial distribution of cashew orchards in relation to other types of land cover, especially in terms of its interpenetration with forest patches. Moreover, if a binary setup were to be adopted, the non-cashew class would be extremely heterogeneous, resulting in a high variance, which could impair the classification task.

A 10 m regular grid was applied to the polygons to extract corresponding pixels from the S-2 bands stack. (Table 1). GEE was also used to intersect the reference data set with the S-2 mosaic in order to create labeled data (Fig. 2). This results in a subset of the pixels of the mosaic being labeled, which represents the training data used to develop the classifier. Fig. 2 summarizes the steps described in sections 3.1 to 3.3.

3.3. Textural metrics

Cashew trees are often planted in a row fashion. Because of this, at a certain stage of their development, cashew orchards very often begin to present a visible row pattern (Fig. 3 a and b). This pattern may help to distinguish cashew orchards from forests (Fig. 3 c) upon

Table 1
Number of pixels of each class in the reference data set.

Class	No. Pixels
Forest	8254
Cashew	3196
Sparse Vegetation	2037
Mangrove	2595
Urban/Bare Soil	3829
Water	4341

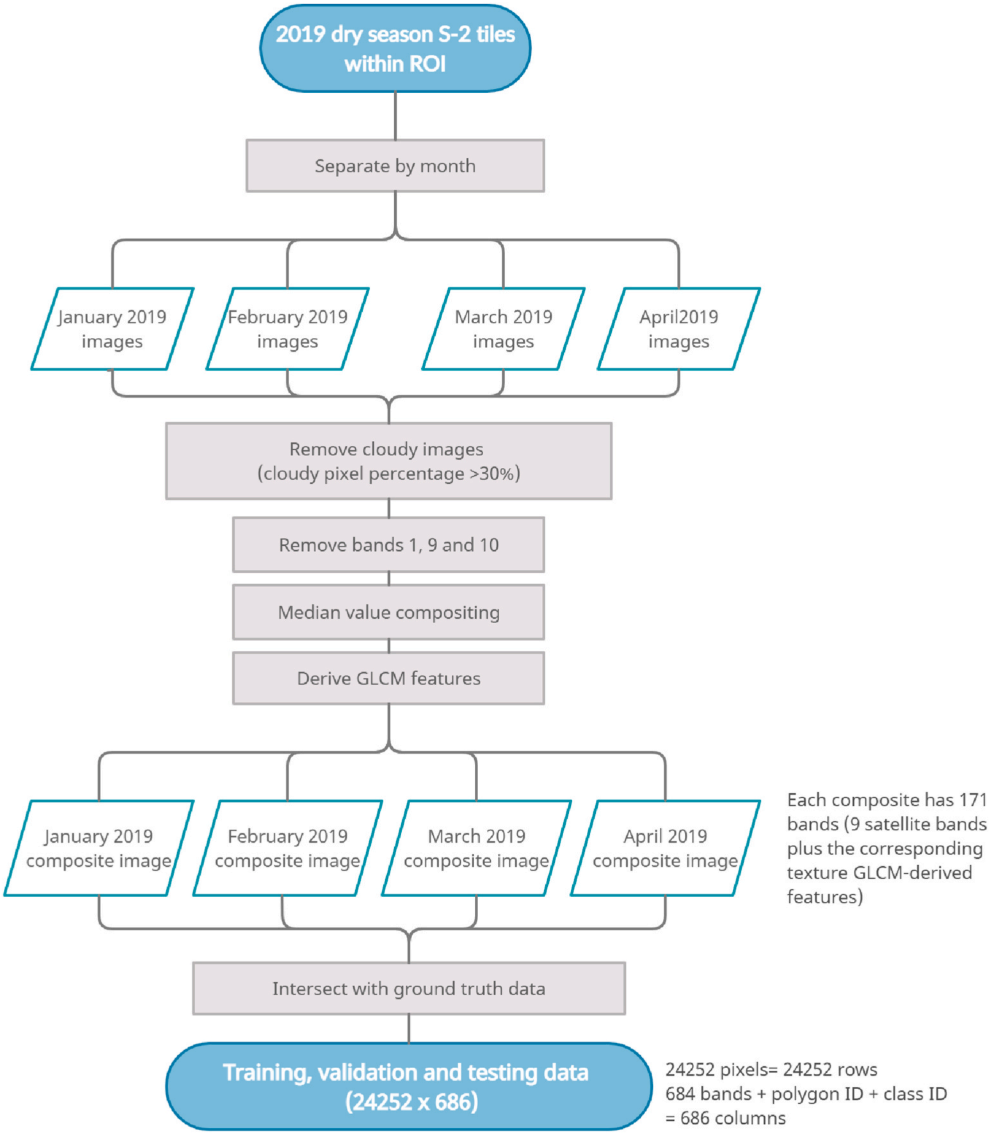

Fig. 2. Schematic overview of the satellite imagery collection and curation.

classification. Distinguishing these two types of land cover using S-2 imagery is inherently hard due to the spectral overlap of these two types of land cover. With this in mind, textural metrics derived from the Gray-Level Co-Occurrence Matrix (GLCM) (Haralick et al., 1973; Conners et al., 1984) were included in this study. Eighteen textural metrics per band were included. Variables encoding these metrics can be fed to the machine learning model together with the satellite bands in order to provide additional information that might be useful to distinguish between the two types of land cover. Bands representing each of the GLCM-derived variables were added to the mosaics, rendering four mosaics (one for each month) with 171 bands each (9 satellite bands plus the corresponding texture GLCM-derived features). Finally, the four mosaics were stacked creating a single mosaic comprised of 684 bands. A detailed description of each of the GLCM-derived features can be found on (Haralick et al., 1973) and (Conners et al., 1984).

3.4. Data preparation and image classification

Data preparation and image classification were performed using the Python programming language. The labeled data were standardized and split into training (70%) and testing (30%) sets. Pixels from the same polygon were kept together in the same set of data (avoiding spatial autocorrelation). Simultaneously, since the data is imbalanced (ie, number of pixels in each class varies), the sets were stratified by class, ensuring their correct representativeness. Variable selection was performed with recursive feature elimination in order to retain only the predictors that are relevant for the problem at hand.

A Random Forest (RF) and a Support Vector Machine (SVM) were applied to the training set using cross-validation and compared. These are two state-of-the-art, widely used algorithms for land cover mapping (Pirotti et al., 2016; Talukdar et al., 2020). While the

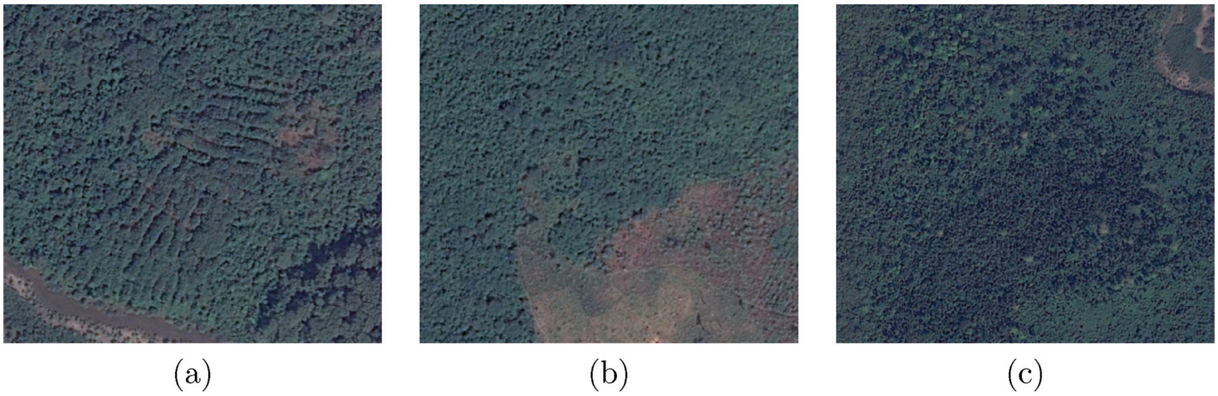

Fig. 3. High resolution satellite images (November 2016) from Google Earth (CNES/Airbus). Fig. 3a (location: 11∘13′40.81′′N, 14∘59′44.53′′W) and 3b (location: 11∘13′49.62′′N, 15∘0′25.86′′W) represent cashew orchards inside the study region. Fig. 3c (location: 11∘12′48.37′′N, 15∘1′3.53′′W) shows a forest area. Note the prominent row pattern present in some portions of the orchards (especially in Fig. 2a) and the difference in texture between Fig. 3a and b, and Fig. 3c.

first uses a non-parametric recursive partitioning approach, the latter relies on the maximum margin principle. Since their inner workings are different, the two algorithms can pick up on different aspects and patterns of the data. Given the imbalanced nature of the data, the balanced accuracy (mean of the proportion of correctly classified pixels of each individual class) was used as the scoring metric at this stage. The best performing classifier (based on the cross-validation results) was adopted. Hyper-parameter optimization was performed for the best performing algorithm using the Dice Similarity Coefficient (DSC) of the class “cashew” relatively to all the other classes as a scoring metric. The resulting algorithm will therefore have optimal performance in this specific class, but not necessarily over all classes.

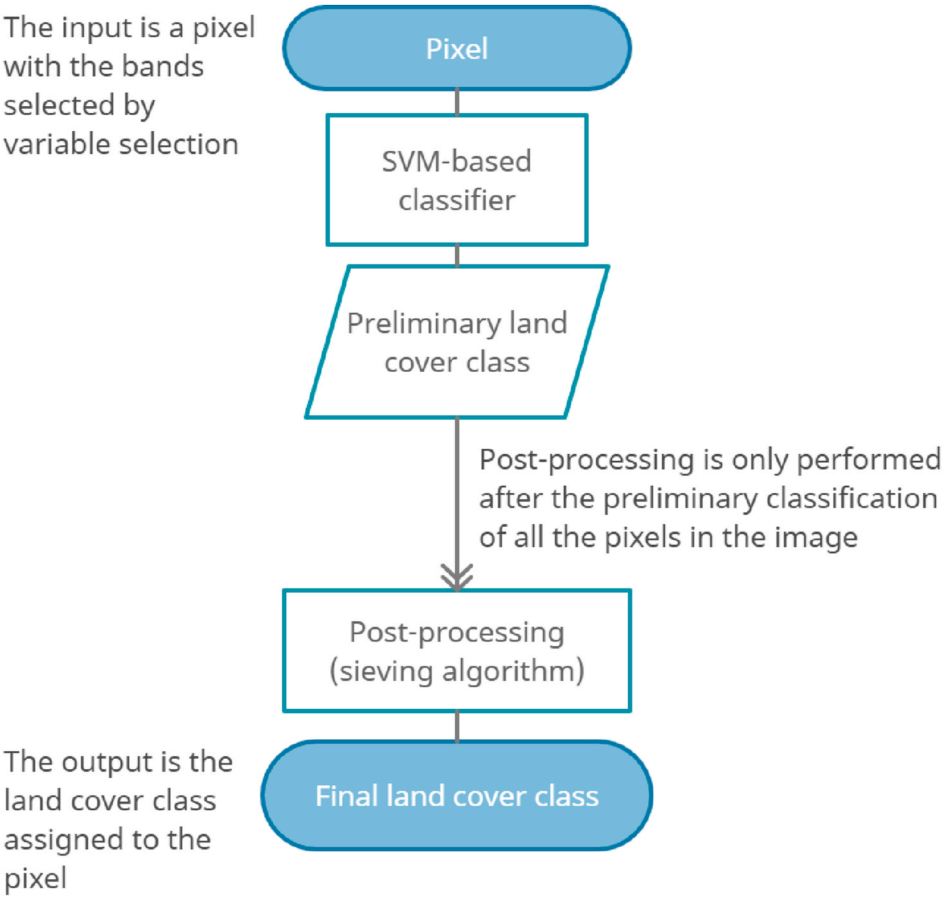

Fig. 4. Schematic overview of the land cover mapping procedure.

3.5. Land cover map production, post-processing and accuracy assessment

The classifier developed using the training data was applied to every pixel of a mosaic containing only the bands selected by the variable selection algorithm. Then, a land cover map was generated based on the resulting classification. Currently, Guinea-Bissau applies the Food and Agriculture Organization (FAO) official forest definition for international reports. FAO considers Forest as: "Land spanning more than 0.5 ha with trees higher than 5 m and a canopy cover of more than 10 percent, or trees able to reach these thresholds in situ. It does not include land that is predominantly under agricultural or urban land use." (FAO, 2020b). For this reason, the resulting map was processed with a 50-pixel sieve filter. This step ensures that the minimum mapping unit is in line with the country's forest definition. Upon applying the sieving filter, raster polygons smaller than 0.5 ha (50 pixels) were replaced with the pixel value of the largest neighboring polygon. The steps described in this section are summarized in Fig. 4.

The final accuracy was assessed by comparing the label of the testing portion of the data with the label for those same pixels in the land cover map resulting from the developed algorithm. For the visualization and color coding of the final land cover map, ArcGIS, a geographical information systems software, was used (Redlands, 2011).

4. Results

In this section, results regarding the production of the monthly composites, the selected variables, the classification model and the land cover map will be presented. Some additional results regarding variable selection and the sieving process are presented in Appendices A and B, respectively.

4.1. Monthly composites

An (artificial) RGB of each monthly composite is shown in Fig. 5. Bands 12- Short-Wave Infrared (SWIR), 08- Near Infrared (NIR) and 04-Red were used to produce the figures. Later during the dry season (Fig. 5 d), the vegetation is at a drier state (notice the more

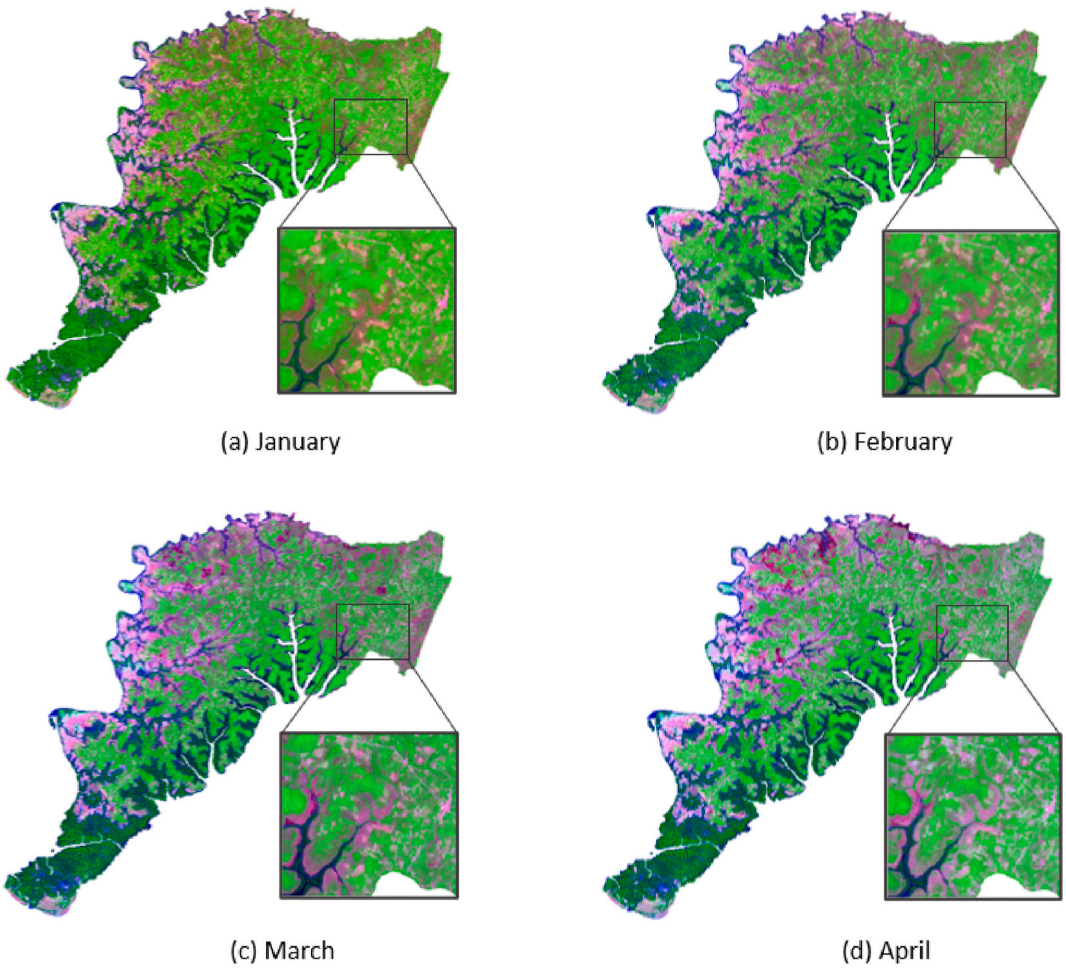

Fig. 5. SWIR-NIR-Red RGB color composite images of the study area corresponding to the four monthly median composites produced, one per each month being considered (Jan–Apr). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

pinkish appearance of this sub-figure). The spectral signatures of each class in each composite are depicted in Fig. 6. The overall spectral overlap, and especially between "Forest" and "Cashew", is smaller during towards the end of the dry season (Fig. 6d).

4.2. Classification model

After fitting and optimizing both models using the training portion of the reference data set, the balanced accuracy and the DSC of the class "Cashew" were used as scoring metrics to chose between the models. Given the superior performance of the SVM regarding such metrics (84.56% DSC and 92.27% balanced accuracy versus 81.88% DSC and 90.75% balanced accuracy with the RF algorithm), this classifier was chosen for the classification task and optimized, to develop the final land cover map. The final set of bands selected by variable selection and used for classification is presented in Table A1.

4.3. Land cover map and accuracy assessment

The final land cover map was developed using the selected classifier, followed by the application of a 50-pixel sieve (Fig. 7). The results of the final accuracy assessment, performed by comparing the labeled test data's ground truth with the output of the final map (after applying the sieving filter), are depicted on Tables 2 and 3. These are a confusion matrix and the per-class DSC of the classification system, respectively. Table 2 also includes the overall accuracy and balanced accuracy of the system. When considering all the land cover classes, the balanced accuracy of the classifier is 91.92%. Although the Dice coefficient of the class Cashew is fairly high (82.54%), the confusion matrix (Table 2) is still capable of highlighting the spectral overlap that exists between Cashew and Forest: most of the misclassified Forest pixels are being classified as Cashew and vice-versa. In order to assess the effect of sieving the map in regard to this fact, Figure B1 depicts spatial distribution of the pixels whose classification turned from cashew into forest and from forest into cashew due to the application of the sieving filter.

As the scale of the land cover map in Fig. 7 may be too coarse to get a proper appreciation of the SVM classifier performance, a

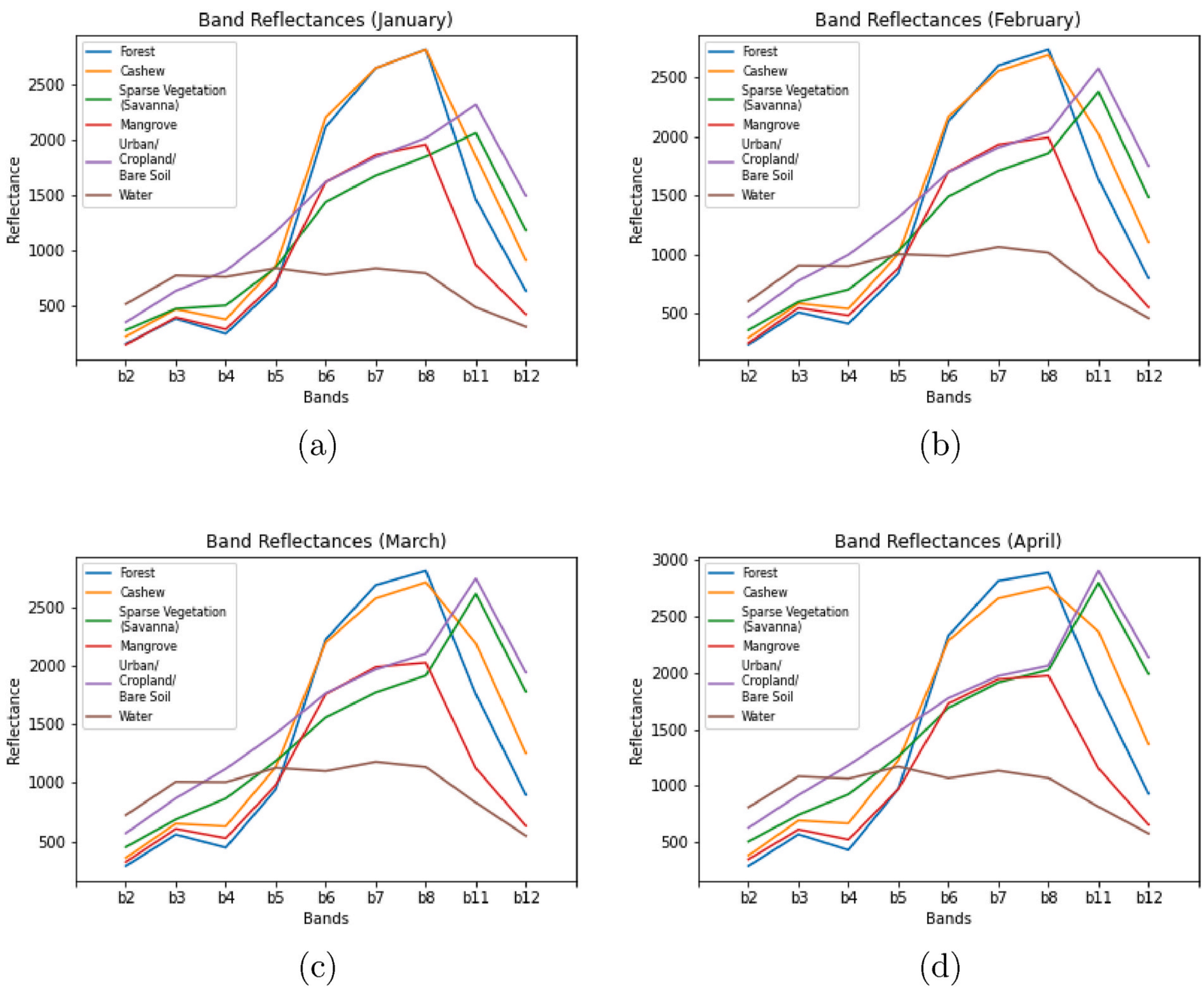

Fig. 6. Spectral signature of each land cover class used in the study. The y-axis is the mean reflectance value. Each line plot corresponds to a median composite of all the 2019 S-2 surface reflectance images regarding the corresponding month.

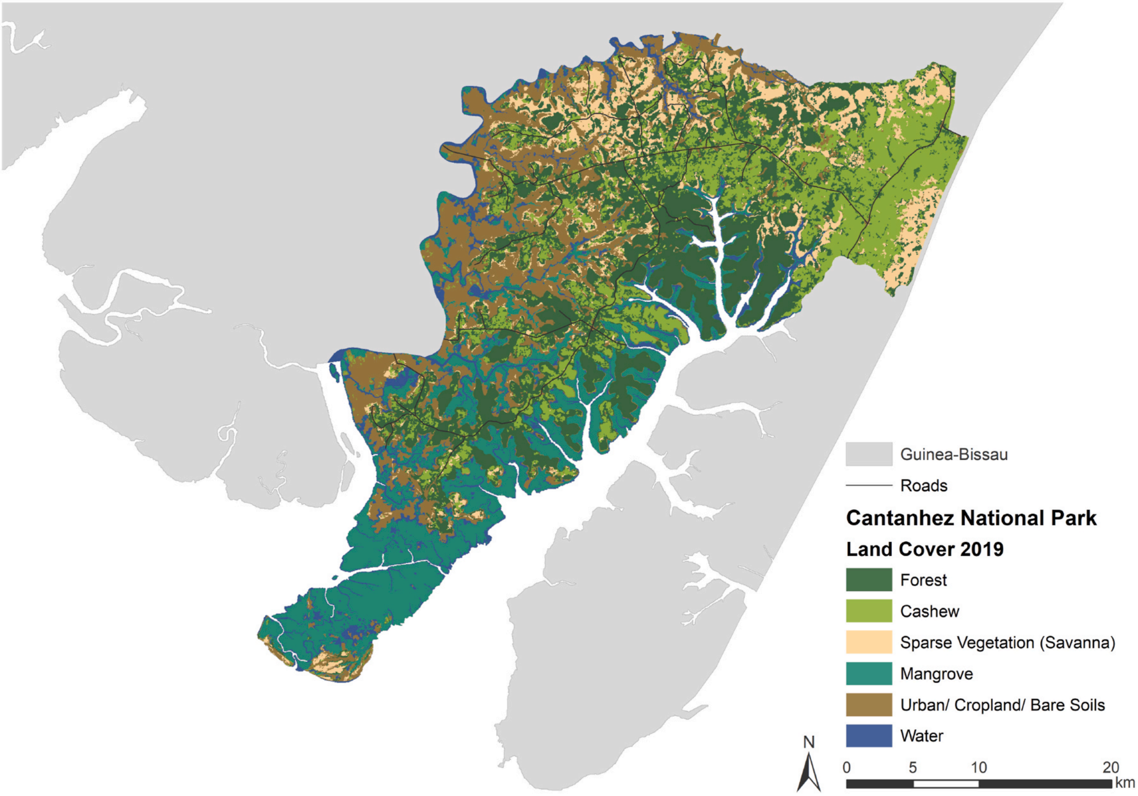

Fig. 7. Land cover map obtained with SVM-based classification of the 2019 mosaic over the ROI, after application of a 50-pixel sieve filter. The roads were added from OpenStreetMap (OpenStreetMap contributors, 2017).

Table 2
Confusion Matrix corresponding to the performance on the test data for the algorithm used to produce Fig. 7.

	Classification (number of pixels)						
Ground Truth (number of pixels)	Forest	Cashew	Sparse Vegetation	Mangrove	Urban/Cropland Bare Soil	Water	**User's Accuracy (%)**
Forest	2293	172	4	0	10	0	92.50
Cashew	130	803	11	0	11	0	84.08
Sparse Vegetation	51	0	550	0	2	0	91.21
Mangrove	0	0	0	746	0	3	99.60
Urban/Cropland/Bare Soil	3	1	24	0	1068	51	93.11
Water	6	0	0	178	2	1885	91.02
Producer's Accuracy (%)	92.35	82.27	93.38	80.74	97.71	97.22	

Accuracy = 91.77%.
Balanced Accuracy = 91.92%.

Table 3
Per-Class Dice similarity coefficient corresponding to the performance on the test data for the algorithm used to produce Fig. 7 and the corresponding number of pixels per class.

Class	Dice Similarity Coefficient (%)
Forest	92.78
Cashew	82.54
Sparse Vegetation	90.34
Mangrove	86.01
Urban/Cropland/Bare Soil	93.61
Water	91.42

detailed view over a selected region within the study area of the output produced by the SVM classifier, the output produced by the SVM classifier after sieving (final map) and the corresponding high resolution RGB image are shown in Fig. 8.

5. Discussion

The results presented in the previous section will now be discussed, starting with some notes on the monthly composites and their spectral signatures. Next, we provide some comments on the classification results and finally on the land cover map produced and respective accuracy assessment.

5.1. Vegetation phenology

Automating cashew orchard detection using satellite imagery may pose some difficulties: forests and cashew plantations are spectrally similar, and hard to discriminate when using the currently available free optical satellite imagery, such as that from *Sentinel*. Although forests and cashew plantations have similar spectral signatures, cashew is a perennial tropical tree, while forest is made up of several species, including some deciduous trees. This means that during the dry season of the year, part of the forest's trees shed their leaves, resulting in an altered spectral signature. In addition, during the dry season the herbaceous plants underneath the forest and cashew orchards dry out, which can enhance the contrast between tree canopies and background, ultimately decreasing the amount of noise in the signal reaching the satellite.

The results presented in Figs. 5 and 6 show that the spectral overlap between cashew and forest is smaller towards the end of the dry season, making it the most informative time of the season. The choice of variables (Table A1) is also in line with this finding, as the number of selected variables corresponding to April is significantly superior to that of the other months, meaning that there is more useful information in the latest month of the season. The regions of the spectrum in which more bands are selected are essential to detect and characterize vegetation, due to the very distinctive low reflectance in the visible and high reflectance in the NIR. However, Fig. 6 shows that although overall spectral separability is high in the NIR, the SWIR bands (11 and 12) best discriminate between forest and cashew plantations. The fact that the herbaceous vegetation underneath the cashew plants is drier during this time might also enhance the cashew orchard's row pattern, which is consistent with the fact that more GLCM-derived features corresponding to April are being considered useful for the classification task.

5.2. Classifier performance and land cover map visual assessment

The DSC of the class Cashew is the lowest among all the classes, even though its value is acceptable (Table 3). The classifier results in very well-balanced user's and producer's accuracy for most of the classes (Table 2) except for the class Mangrove; some Water pixels are being classified as Mangrove, although the opposite does not happen. Typically, mangroves grow in areas periodically flooded by salt water. Depending on when the images were taken, in each image the tide may be higher or lower and the water signal in the mangrove may be more or less intense, which can exacerbate confusion between the two classes. It is also important to notice that the total number of Water pixels is much larger than that of Mangrove pixels (Table 1). This means that even though only a small portion of the water pixels are being classified as mangrove, this will greatly impact the producer's accuracy of the class. As expected, the main difficulty was in distinguishing the spectral signatures of cashew orchards and forests. Despite the larger confusion between these two classes, the classifier is still clearly able to tell them apart in the majority of the pixels. This represents an advance comparatively to Temudo and Abrantes (2014), who considered this distinction unfeasible.

Visual analysis of the output is coherent with the high balanced accuracy obtained in the test set; when visually analyzing the land cover map in detail and comparing its output with high resolution satellite imagery, the map is correctly identifying the orchards and the remaining land cover classes. Even though sieving the map may have introduced minor errors, it also results in a much less noisy output (Fig. 8). More importantly, sieving the map is necessary so that the minimum mapping unit is compliant with the FAO's forest definition, adopted by the country. Even though it might not be very clear for the inexperienced eye, the classifier is correctly identifying the cashew orchards and labeling them as such, even in regions where the vegetation is still short and sparse, most likely because the textural pattern of the orchards is already being detected. Fig. 8 highlights this aspect. The Figure also shows that the

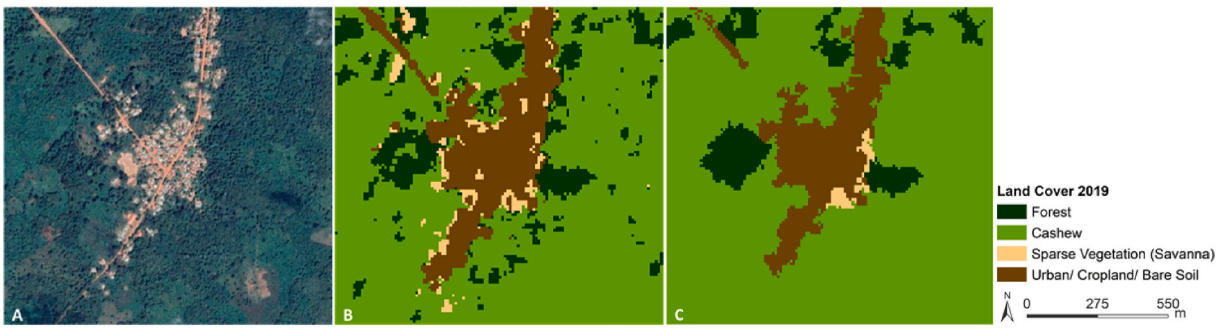

Fig. 8. Detail over a selected region within the study area (location: 11∘18′33.21′′N 14∘50′52.64′′W). A– High resolution image from Google Earth (CNES/Airbus) from November 10, 2019; B– Classification produced by the SVM-based classifier; C– Classification produced by the SVM-based classifier with the sieving filter applied (final map).

classifier is correctly identifying some forest patches, a small village and roads (classified as "Urban/Cropland/Bare Soil"), and a small portion of sparse vegetation right next to the village. Note that in the upper left area of the region, part of the road is not being correctly identified, most likely because the sieving filter is incorrectly converting those from "Urban/Cropland/Bare Soil" into "Cashew", the locally dominant land cover class.

Although Fig. 8 might convey that the area of cashew orchards is being overestimated and that of forests underestimated, the patches of forest pixels that are attributed to cashew are smaller than 0.5ha and hence cannot be considered, by definition, as forest. Since the neighboring dominant class is cashew, these forest patches are attributed to this land cover class. This is an unavoidable consequence of the sieving filter, given the abundance, spatial distribution and intertwining of these two classes. Appendix B further explores this topic by showing the spatial pattern of these transitions, as well as the total area affected by them. As expected, the transitions between the two types of land cover happen in small patches (<0.5 ha), scattered across the entire ROI, in areas where the locally dominant class is the one to which the pixels are being attributed. The area of cashew that is attributed to forest is slightly larger than the area of forest that is attributed to cashew, although, in percentage, the difference is quite small. This results in a sieved map that is more generalized than the original, containing larger continuous patches of each type of land cover without severely affecting the total area of each class.

6. Conclusions

This work shows that when ground truth land cover data are available for a given year, a classifier can be successfully trained using those data to produce accurate land cover maps that depict cashew orchards. Even though the performance on the "Cashew" class is slightly lower than that of the other classes, satisfactory results can be obtained (82.54% Dice coefficient).

This regional study serves as a viability test to prove that monitoring of this situation is in fact possible, contributing a valuable tool for the evaluation of the country's food, climate, and ecosystem sustainability outlook. This work represents the first step towards the monitoring of cashew orchards in Guinea-Bissau using satellite imagery. The possibility of deriving sufficiently accurate maps is very relevant for supporting the country's climate action, which aims at reducing deforestation and improving sustainable forest management. Given the current lack of reliable information on this topic, this type of product is especially relevant, considering that the country has been working towards improving forest monitoring and transparency. Remote sensing is thus crucial to quickly obtain data that can assist better land use decision-making and contribute to improve sustainability in Guinea-Bissau. A broader map comprising the entire country may be developed upon collection of additional data spanning the entire country, without the need of substantial alteration to the methodology. Further development of this approach could be used to develop a change detection system.

Author statement

Sofia Cardoso Pereira: Conceptualization, Methodology, Software, Investigation, Formal analysis, Visualization, Writing - Original Draft. Catarina Lopes: Methodology, Data Curation, Visualization, Writing - Review & Editing. João Pedro Pedroso: Supervision, Writing - Review & Editing.

Ethical Statement for Remote Sensing Applications: Society and Environment

Hereby, I Sofia Cardoso Pereira consciously assure that for the manuscript Mapping Cashew Orchards in Cantanhez National Park (Guinea-Bissau) the following is fulfilled:

1) This material is the authors' own original work, which has not been previously published elsewhere.
2) The paper is not currently being considered for publication elsewhere.
3) The paper reflects the authors' own research and analysis in a truthful and complete manner.
4) The paper properly credits the meaningful contributions of co-authors and co-researchers.
5) The results are appropriately placed in the context of prior and existing research.
6) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.
7) All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

The violation of the Ethical Statement rules may result in severe consequences.

To verify originality, your article may be checked by the originality detection software iThenticate. See also http://www.elsevier.com/editors/plagdetect.

I agree with the above statements and declare that this submission follows the policies of Remote Sensing Applications: Society and Environment as outlined in the Guide for Authors and in the Ethical Statement.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Variable selection

One hundred variables were selected by the variable selection algorithm. Table A1 displays key aspects regarding variable selection. The regions of the spectrum in which more bands are selected are the visible (green and red, bands 3 and 4, respectively) and the NIR (band 8) (Table A1). The sum average (savg) statistic of the GLCM matrix appears to be very important for the classification, as it has been selected across every month and for every original S-2 band (Table A1). This variable is one of the measures proposed in (Haralick et al., 1973), taking high values for lighter (high intensity) pixels, magnified when the surrounding pixels are different. Ultimately, it is one of many possible measures of contrast, which turned out to be effective in our setting.

Table A1
Set of variables selected by the variable selection algorithm (in gray). A detailed description of each of the texture-related features can be found in Haralick et al. (1973) and Conners et al. (1984).

<table>
<tr><th>Month</th><th>January</th><th>February</th><th>March</th><th>April</th></tr>
<tr><td>B2</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B2_savg</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B2_contrast</td><td></td><td></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B2_diss</td><td></td><td></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B3</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B3_savg</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B3_contrast</td><td></td><td style="background:#eee"></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B3_diss</td><td></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B3_dvar</td><td></td><td style="background:#eee"></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B3_inertia</td><td></td><td style="background:#eee"></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B4</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B4_savg</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B4_contrast</td><td style="background:#eee"></td><td style="background:#eee"></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B4_diss</td><td></td><td style="background:#eee"></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B4_inertia</td><td></td><td style="background:#eee"></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B5</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B5_savg</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><th>Month</th><th>January</th><th>February</th><th>March</th><th>April</th></tr>
<tr><td>B6</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B6_savg</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B7</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B7_savg</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B8</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B8_savg</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B8_contrast</td><td style="background:#eee"></td><td></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B8_diss</td><td style="background:#eee"></td><td></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B8_inertia</td><td></td><td style="background:#eee"></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B8_dvar</td><td></td><td></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B8_var</td><td></td><td></td><td></td><td style="background:#eee"></td></tr>
<tr><td>B11</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B11_savg</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B11_dvar</td><td style="background:#eee"></td><td></td><td></td><td></td></tr>
<tr><td>B12</td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
<tr><td>B12_savg</td><td></td><td style="background:#eee"></td><td style="background:#eee"></td><td style="background:#eee"></td></tr>
</table>

Appendix B Sieving filter

After applying the sieving algorithm, 5% (1394.1 ha of a total of 28616.7 ha) of the area classified as forest was assigned to cashew and 8% (1645.9 ha of a total of 19709.7 ha) of the area classified as cashew turned into forest. Overall, the area classified as forest increased 3% (from 28616.7 ha to 29503.7 ha), while the area classified as cashew decreased 0.1% (from 19709.7 ha to 19680.9 ha). Figure B1 shows the spatial distribution of these transitions.

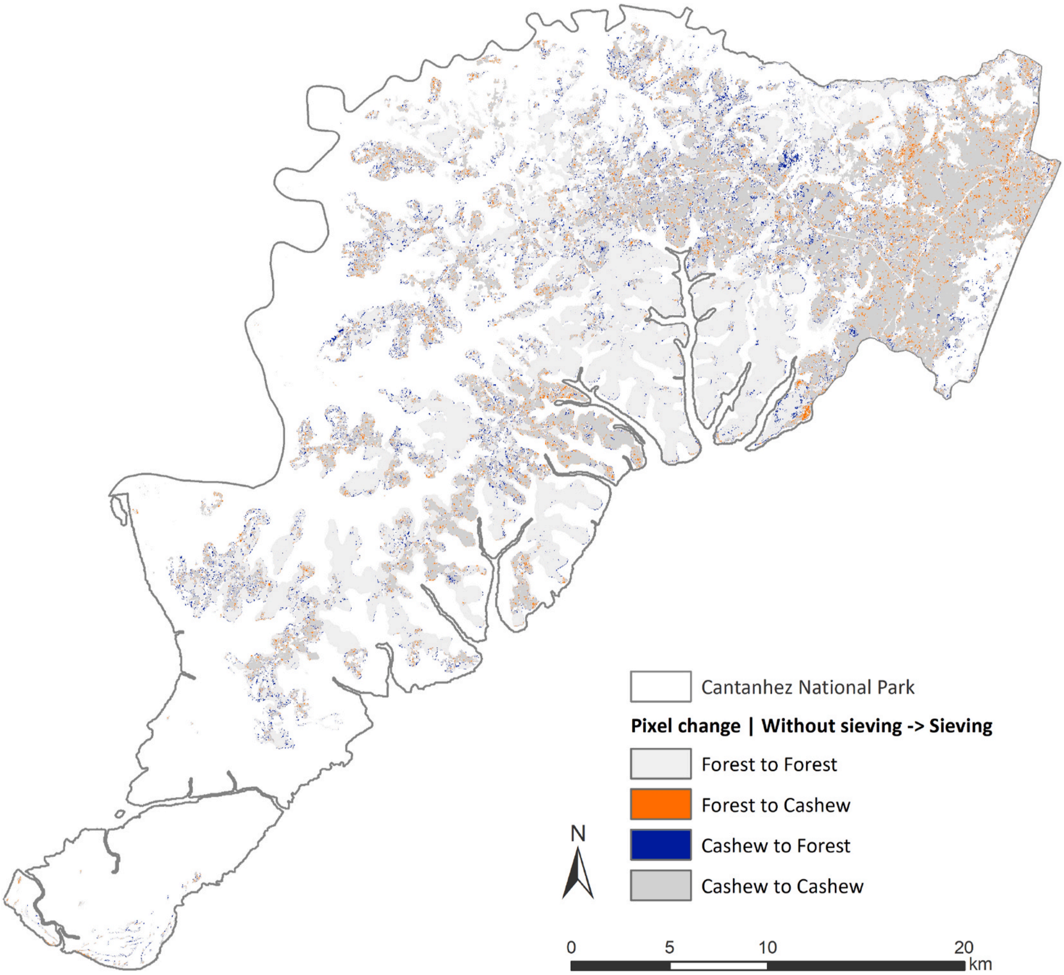

Fig. B1. Forest and cashew regions. The pixels whose classification transitioned from forest to cashew and from cashew to forest after application of the sieving filter are shown in orange and blue, respectively.

References

Conners, R.W., Trivedi, M.M., Harlow, C.A., 1984. Segmentation of a high-resolution urban scene using texture operators (Sunnyvale, California). Comput. Vis. Graph Image Process 25, 273–310. https://doi.org/10.1016/0734-189X(84)90197-X.

FAO, 2020a. Food and Agriculture Organization of the United Nations. URL: https://www.fao.org/faostat/en/#data/QCL. Accessed 2020-06-17 by choosing country "Guinea Bissau", item "cashew nuts, with shell" and year "2019"].

FAO, 2020b. FRA 2020 Terms and Definitions. Technical Report. Food and Agriculture Organization of the United Nations. URL: http://www.fao.org/3/I8661EN/i8661en.pdf.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2017.06.031.

Hanusch, M., 2016. Guinea-Bissau and the cashew economy. Technical Report 11. World Bank. URL: http://documents.worldbank.org/curated/en/443831467999735473/102933-REVISED-PUBLIC-MFM-Practice-Note-11.pdf.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. 3, 610–621. https://doi.org/10.1190/segam2015-5927230.1.

International Finance Corporation, 2010. Prospects for Cambodia 's Cashew Sub-sector. International Finance Corporation. Technical Report 1.

Lopes, C., Leite, A., Vasconcelos, M.J., 2019. Open-access cloud resources contribute to mainstream REDD+: the case of Mozambique. Land Use Pol. 82, 48–60. https://doi.org/10.1016/j.landusepol.2018.11.049.

OpenStreetMap contributors, 2017. Planet dump retrieved from. https://planet.osm.org. https://www.openstreetmap.org.

Pirotti, F., Sunar, F., Piragnolo, M., 2016. Benchmark of machine learning methods for classification of a Sentinel-2 image. Int. Arch. Photogrammetr. Rem. Sens. Spat. Inf. Sci.- ISPRS Arch. 41, 335–340. https://doi.org/10.5194/isprsarchives-XLI-B7-335-2016.

Redlands, C.E.S.R.I., 2011. Arcgis Desktop: Release 10.

Sentinel-2, M.S.I., 2020. User Guides - Sentinel-2 MSI - Overview - Sentinel Online. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/overview. Accessed 2020-05-20.
Simonetti, D., Simonetti, E., Szantoi, Z., Lupi, A., Eva, H.D., 2015. First results from the phenology-based synthesis classifier using Landsat 8 imagery. Geosci. Rem. Sens. Lett. IEEE 12, 1496–1500. https://doi.org/10.1109/LGRS.2015.2409982.
Sousa, J., Ainslie, A., Hill, C., 2017. Sorcery and nature conservation. Environ. Conserv. 1–6. https://doi.org/10.1017/S0376892917000327.
Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.A., Rahman, A., 2020. Land-use land-cover classification by machine learning classifiers for satellite observations — a review. Rem. Sens. 12, 1135. https://doi.org/10.3390/rs12071135.
Temudo, M.P., Abrantes, M., 2014. The cashew frontier in Guinea-Bissau, west Africa: changing landscapes and livelihoods, 2014 Hum. Ecol. 42 (2 42), 217–230. https://doi.org/10.1007/S10745-014-9641-0.
UN-FCCC, 1994. UNFCCC - united Nations Framework convention on climate change — knowledge for policy. URL: https://ec.europa.eu/knowledge4policy/organisation/unfccc-united-nations-framework-convention-climate-change_en. Accessed 2020-05-31.
UN-FCCC, 2019. Guinea-Bissau first biennial update report. Technical Report. URL: https://unfccc.int/sites/default/files/resource/FINAL_GNB_BUR1.pdf.
Vasconcelos, M.J., Cabral, A.I., Melo, J.B., Pearson, T.R., Pereira, H.d.A., Cassamá, V., Yudelman, T., 2015. Can blue carbon contribute to clean development in West-Africa? The case of Guinea-Bissau. Mitig. Adapt. Strategies Glob. Change 20, 1361–1383. https://doi.org/10.1007/s11027-014-9551-x.