

## Article

# Cropland Mapping over Sahelian and Sudanian Agrosystems: A Knowledge-Based Approach Using PROBA-V Time Series at 100-m

Marie-Julie Lambert \*, François Waldner and Pierre Defourny

Earth and Life Institute, Université catholique de Louvain, 2 Croix du Sud, 1348 Louvain-la-Neuve, Belgium; francois.waldner@uclouvain.be (F.W.); pierre.defourny@uclouvain.be (P.D.)

\* Correspondence: marie-julie.lambert@uclouvain.be; Tel.: +32-10-478192

Academic Editors: Clement Atzberger and Prasad S. Thenkabail

Received: 6 January 2016; Accepted: 29 February 2016; Published: 11 March 2016

**Abstract:** Early warning systems for food security require accurate and up-to-date information on the location of major crops in order to prevent hazards. A recent systematic analysis of existing cropland maps identified priority areas for cropland mapping and highlighted a major need for the Sahelian and Sudanian agrosystems. This paper proposes a knowledge-based approach to map cropland in the Sahelian and Sudanian agrosystems that benefits from the 100-m spatial resolution of the recent PROBA-V sensor. The methodology uses five temporal features characterizing crop development throughout the vegetative season to optimize cropland discrimination. A feature importance analysis validates the efficiency of using a diversity of temporal features. The fully-automated method offers the first cropland map at 100-m using the PROBA-V sensor with an overall accuracy of 84% and an F-score for the cropland class of 74%. The improvements observed compared to existing cropland products are related to the hectometric resolution, to the methodology and to the quality of the labeling layer from which reliable training samples were automatically extracted. Classification errors are mainly explained by data availability and landscape fragmentation. Further improvements are expected with the upcoming enhanced cloud screening of the PROBA-V sensor.

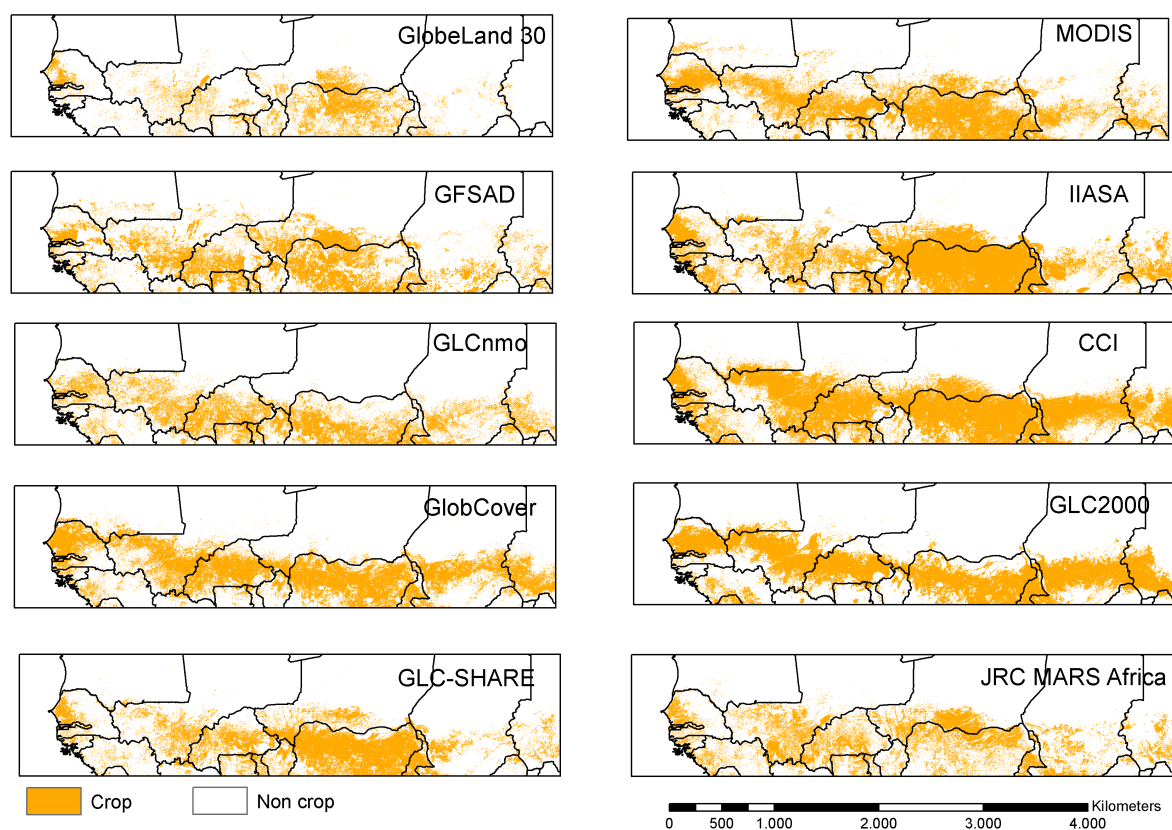
**Keywords:** cropland; Sahelian and Sudanian agrosystems; knowledge-based; PROBA-V; time series

## 1. Introduction

In West Africa, about 70% of the population depends on cropping systems to meet their basic needs [1]. Rainfed and smallholder agriculture are the two major components of the food production system [2–4]. Nevertheless, inter-annual crop production variability, mainly due to the variability of the rainfall distribution and poor soil fertility, has led to chronic food insecurity [5]. Recurrent severe regional droughts occurred in 1972–1973 and 1983–1984 and led to famines that affected millions of people [6]. Even though no regional drought has occurred since 1984, localized famines recently hit some Sahelian countries [7]. In addition to natural variability, socio-economic and political factors, such as population growth, social conflict and low economic development, put pressure on cropping systems [8,9]. It is of primary importance to monitor land use change [10–12] and to develop an in-depth understanding of the detailed spatial patterns and the temporal dynamics of cropland [13]. Early warning systems *e.g.*, the Famine Early Warning Systems Network (FEWS-NET) and the Global Information and Early Warning System (GIEWS), for food security require accurate and up-to-date spatial information about the cropland to monitor food production [7]. Nevertheless, mapping cropland remains challenging in this region, especially because of the agricultural landscape fragmentation, the spatial heterogeneity of the cropland, the diversity of the cropping systems and the mosaic of cropland, fallow and natural grassland [14,15].

Many global land cover products delivered at different resolutions ranging from 30-m to 10-km contain cropland classes. High resolution Landsat-based (30-m) cropland maps, such as Cropland-Use Intensity (CUI), Africover and GlobeLand 30, are consistent, but the methodology prevents regular updates [16,17]. Medium resolution land cover maps, such as MODIS product, ESA Climate Change Initiative (CCI) Land cover, GlobCover, Global Land Cover 2000 (GLC 2000), Global Land Cover by National Mapping Organizations (GLCnmo), Global Food Security Analysis-support Data (GFSAD), International Institute for Applied Systems Analysis (IIASA) cropland, EcoClimap II and GLC-Share [15,18–25], offer current cropland information. Yet, most of these global land cover maps do not focus on agriculture and tend to neglect the regional specificities of Western Africa. Specific cropland maps do exist, such as the global map of rainfed cropland (GMRCA) and the global irrigated area map (GIAM), but their resolutions of 10-km are not suitable for regional applications [26,27] and are too coarse for monitoring and planning purposes.

Several studies have highlighted discrepancies in the extent and the spatial distribution of cropland within global products [16,28,29]. For Africa, the MODIS product contains 30% more crop area than the GlobCover product [30]. The global cropland extent fluctuates between 1.11 and 3.62 billion hectares according to different global products [26]. The complexity of the Sahelian and Sudanian agrosystems increases the inconsistencies between global croplands (Figure 1). Some of the differences are explained by the spatial resolution, the year of production, as well as differences in the thematic legend classes and the cropland definition (Table 1).



**Figure 1.** Cropland over Sahelian and Sudanian agrosystems on ten global products showing the high variability of cropland extent.

**Table 1.** Diversity of agriculture-related classes for the 10 previous global land cover maps. GlobeLand 30 is based on Landsat images (30-m); GlobCover and CCI are based on MERIS images (300-m); MODIS and GLCnmo are based on MODIS images (500-m); GLC 2000 is based on SPOT-Vegetation images (1 km); the JRC MARS map hybrid product at 250-m; and IIASA cropland, GLC Share and Global GFSAD hybrid products at 1-km.

Product	Cropland	Product	Cropland
<b>GlobeLand30</b>	Cultivated land	<b>MODIS</b>	Cropland
<b>GFSAD</b>	Cropland, irrigation major		Mosaic cropland/natural vegetation
	Cropland, irrigation minor	<b>IIASA</b>	>25% of probability of crop
	Cropland, rainfed	<b>CCI</b>	Cropland rainfed
	Cropland rainfed minor fragments		Cropland irrigated or post flooding
	Cropland rainfed very minor fragments		Mosaic cropland (>50%)/natural vegetation
<b>GLCnmo</b>	Cropland: herbaceous crop		Mosaic natural vegetation (>50%) / cropland
	Cropland/other vegetation mosaic	<b>GLC2000</b>	Cultivated and managed areas
	Paddy field: graminoid crops/non graminoid crop		Mosaic cropland/shrubland or grass cover
<b>GlobCover</b>	Rainfed cropland		Mosaic cropland/tree cover/natural vegetation
	Mosaic cropland (50%–70%)/vegetation (20%–50%)	<b>JRC MARS</b>	// GlobCover
	Mosaic vegetation (50%–70%)/cropland (20%–50%)		
	Cultivated and managed areas		
	Post-flooding or irrigated croplands		
<b>GLC Share</b>	Cropland		

In this comparison, the mosaic classes of cropland and natural vegetation are considered as cropland. The GlobeLand 30 cropland class considers land used for agriculture, horticulture and gardens, including paddy fields, irrigated and dry farmland, vegetation and fruit gardens [17], while MODIS cropland is defined by land covered with temporary crops followed by harvest and a bare soil period [18]. Bontemps *et al.* [31] highlighted high year-to-year instabilities in labels, especially between cropland and natural vegetation for GlobCover and MODIS products.

Many classification algorithms have been developed to discriminate land cover from satellite data. Parametric classifiers, such as the maximum likelihood classifier, and non-parametric classifiers, such as the neural network classifier, decision tree classifiers and machine learning algorithms, are among the most popular [32,33]. Due to their discriminating power, many types of neural networks have been developed [34]. Decision tree classifiers transform very complex classification problems into simpler decision making processes [35]. The use of machine learning algorithms is becoming increasingly important, yielding high accuracies for land cover classification and often outperforming other algorithms [36–40].

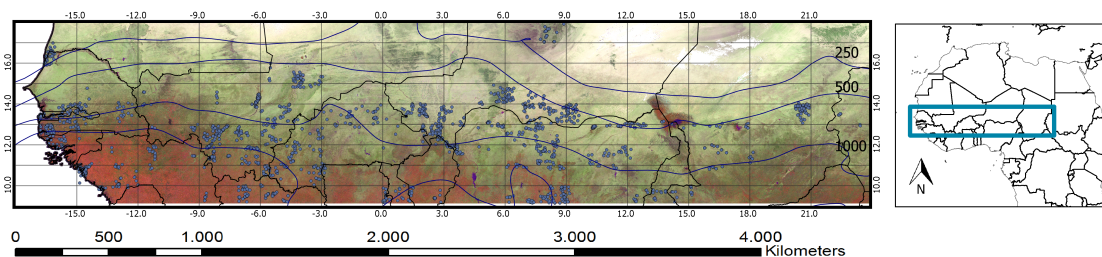
Taking advantage of the spatial resolution of PROBA-V, this research aims at developing a new method to map cropland in the Sahelian and Sudanian agrosystems where global products remain inconsistent and cropland mapping challenging [29]. The fully-automated methodology limits the use of field data and could produce regional cropland on a regular basis for crop monitoring. The relative importance of spectral-temporal features used in the method is studied. Then, the resulting 2014 *Sudano-Sahelian Cropland map* is validated with photo interpretation on very high resolution imagery and compared to other global products. The spatial distribution of errors is assessed, and potential explanatory variables are studied. Finally, an error analysis focuses on challenging areas where most discrepancies between global products occurred.

## 2. Materials

### 2.1. Study Site

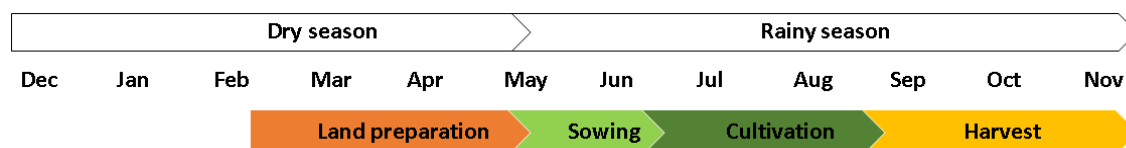
This research focuses on the Sahelian and Sudanian agrosystems in West Africa with the area of interest ranging from 17°W–23°E to 9°N–18°N, comprising a region more than 4300-km long and 1000-km wide (Figure 2). This region is known for a strong north-south gradient linked with the movements of the Intertropical Convergence Zone (ITCZ). Climatic conditions range from arid and desert regions in the north to sub-humid areas in the south through semi-arid areas. The study area is

characterized by two clearly distinct seasons: a rainy season and a dry season. A flat topography with low plateaus and valleys characterizes the landscape.



**Figure 2.** Study site with the validation sample (blue points) and regional grid (in grey) with isohyets of 250-mm, 500-mm and 1000-mm. The background image corresponds to the surface reflectances in infrared, mid-infrared and red bands for the dates corresponding to the maximum of red reflectance (maximum red feature).

The Sahelian region is the area between isohyets 250-mm and 500-mm. Sowing occurs at the beginning of the rainy season. The growing period is short (60–90 days). The start and the length of the growing period are positively correlated with the onset of precipitation, which is more variable than the ending date of the rainy season [5,41]. The Sudanian region is considered in this study as bounded by isohyets 500-mm and 1000-mm, although there is no broad agreement on its definition. The length of the rainy season in the Sudanian region equals the length of the dry season, resulting in growing period of more than 150 days between May and November [42] (Figure 3). The fields are usually very small and of various shapes, with many fields with an area less than 1 ha.



**Figure 3.** Calendar of seasons and agricultural management. The study region contains two clearly distinct seasons with crop cultivation during the rainy season and land preparation at the end of the dry season.

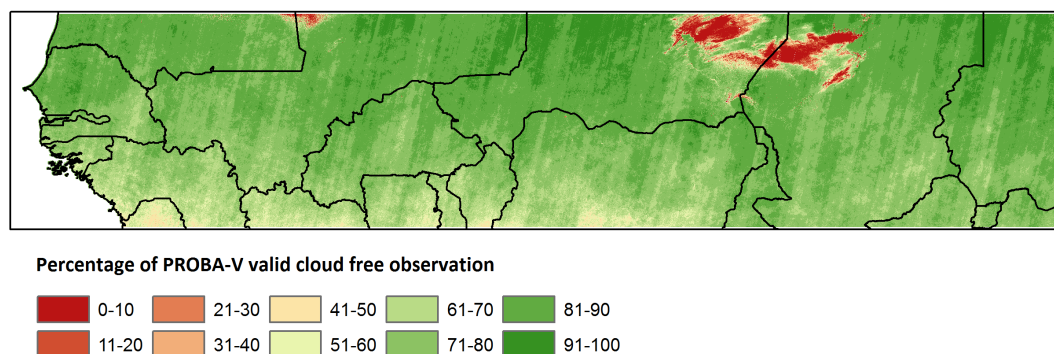
## 2.2. Data

PROBA-V is the successor satellite of SPOT-Vegetation and was launched in May 2013 in order to fill the gap between SPOT-Vegetation and Sentinel-3. While it was not in the initial specification of the mission, this sensor currently offers global coverage every five days at a 100-m spatial resolution in four wavebands, blue, red, near-infrared (NIR) and mid-infrared (SWIR), thanks to its central camera. PROBA-V 100-m daily surface reflectance has been composited on a weekly basis using a mean composite approach. The time series of eleven months (48 composites) from 19 March 2014 to 12 February 2015 was prepared and includes the growing season. The data availability mainly depends on cloud/shadow contamination and the non-five-day global coverage at 100-m in the lower latitudes due to the 517-km swath [43]. False cloud detections are systematically observed over some high brightness areas, such as desert (Figure 4). However, these areas do not contain crops and will not degrade the classification results. In the study area, 37 out of the 48 seven-day composites were on average usable (77%), making it possible to understand the temporal behavior of the main crop types. However, clouds are more frequent in lower latitudes and during the vegetation growth season. These factors could limit classification quality in those regions (Figure 5).

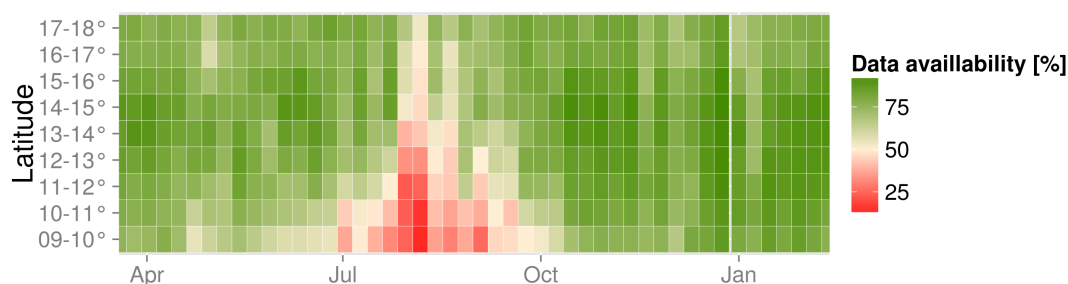
A labeling layer map was built from GlobeLand 30 [17] with the specific contribution of ESA CCI Land Cover [19] for the irrigated cropland. The labeling layer is the reference map from which the



training dataset is extracted for the supervised cropland classification. GlobeLand 30 offers a global land cover of 10 classes at 30-m based on Landsat images for 2010. The accuracy of this global dataset was first computed with an independent validation dataset of Zhao *et al.* [44] and results in an OA of 88%. GlobeLand 30 was then validated with a set of 7 high resolution images. A Pareto boundary analysis showed the performance of the 30-m spatial resolution and the better balance between omission errors and commission errors for GlobeLand 30. The area under the Pareto boundary defines the performance of the spatial resolution, as the smaller the area, the more performant the resolution (for more details, see [45]). This area is two-fold smaller for GlobeLand 30 than for the JRC product and three-fold smaller than for the MODIS and the CCI products.



**Figure 4.** Percentage of PROBA-V valid cloud-free observation for the study area during the whole time series. False detection of clouds is systematically observed over the desert area. The valid cloud-free observation percentage decreases with latitude.



**Figure 5.** Percentage of data available for each degree of latitude ranging from 9° to 18°. Low data availability is observed in low latitudes and during vegetation growth (from July to October).

The CCI Land Cover project delivers three global land cover products, the most recent relating to the 2008–2012 period. In the study area, agriculture is mainly rainfed, while irrigation occurs at specific locations, e.g., the Niger basin. The semi-arid conditions lead to strong spectral contrast between rainfed cropland and irrigated cropland. Therefore, a separate class for irrigated cropland has been created with all single pixels belonging to the crop class in GlobeLand 30 and irrigated crop in the ESA CCI Land Cover Map. The final labeling layer contains eleven classes: the ten of GlobeLand 30 plus the irrigated crop class from the CCI Land Cover.

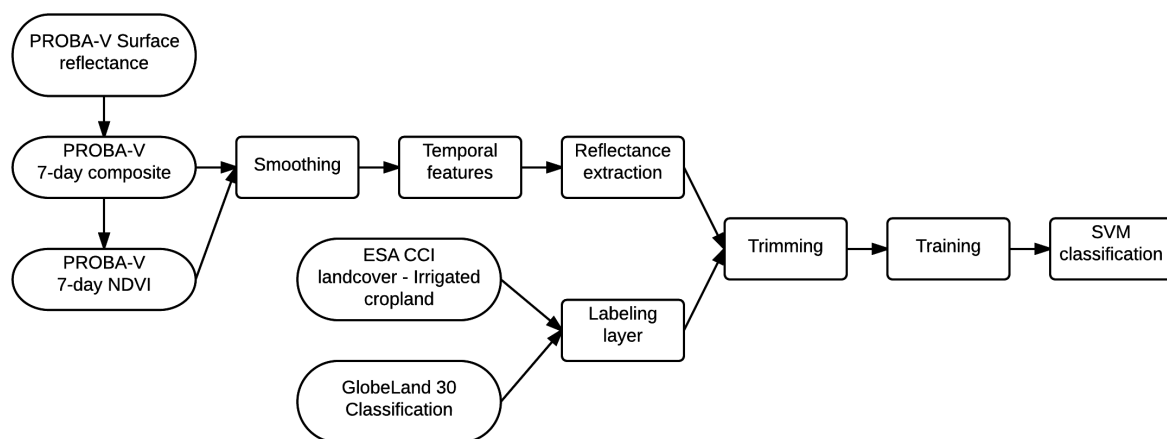
### 3. Methodology

The challenge of cropland mapping in West Africa is dealing with the very strong latitudinal gradient in terms of seasonal dynamics, as well as land cover. A knowledge-based approach identifies relevant temporal features throughout the growing period to optimize cropland capture. A support vector machine (SVM) classifier was selected for its good performances compared to other classification algorithms' classifiers [36,38]. First, the 2014 *Sudano-Sahelian Cropland map* was compared to existing

global products to study discrepancies. Then, classification errors were related to potential explanatory variables, such as location, valid cloud-free observation and landscape characteristics. Finally, classification errors were also studied in regions where global products disagreed on cropland. The method developed is fully automated, *i.e.*, meaning no human action during the whole processing from images to cropland map.

### 3.1. Cropland Classification

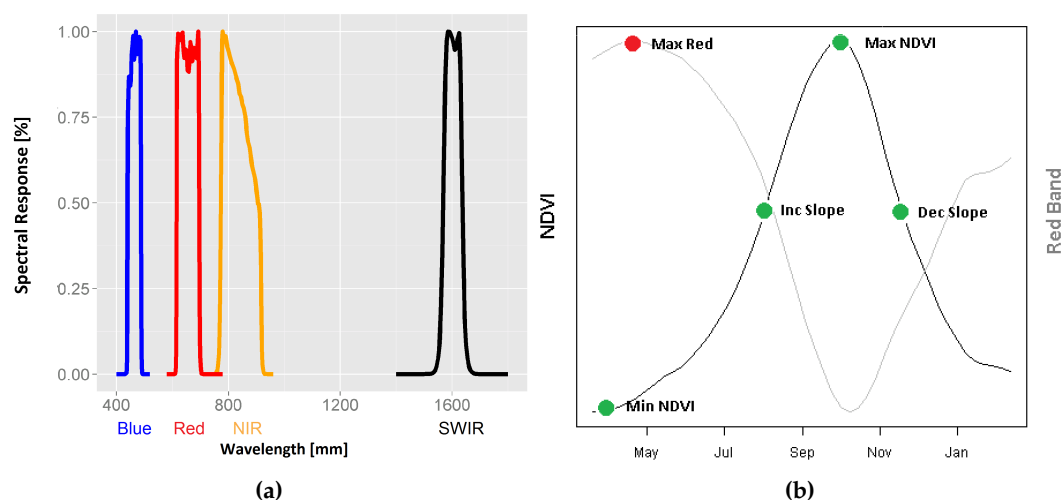
The development of cropland mapping in Sahelian and Sudanian regions proceeds in a three-step method: (i) extraction of the temporal features from filtered PROBA-V time series; (ii) efficient local training based on trimmed data; and (iii) classification using SVM (Figure 6).



**Figure 6.** Three-step methodology for cropland classification: (i) extraction of the temporal features; (ii) local training based on trimmed data; (iii) classification using SVM.

#### 3.1.1. Extracting Temporal Features from PROBA-V

A knowledge-based approach permits a better classification of cropland based on relevant temporal and spectral features. The features were surface reflectance composited at specific events of the growing season when crops are expected to behave differently than other land covers do. Among the cropland discrimination studies [46–49], five temporal features were selected: the maximum of the red band (max red), the minimum and maximum of the Normalized Difference Vegetation Index (NDVI) and the increasing and decreasing slopes of the NDVI profile (Figure 7b). Soil preparation practices, such as tillage and sowing, clear the land surface contrasting with naturally-vegetated areas and resulting in higher reflectance in the red band. This timing corresponds to the maximum of red reflectance. During the rainy season, the vegetation development and the growth rate are expected to be superior in cropland than in non-cropland areas. This was captured by the NDVI maximum and minimum and the increasing slope of the NDVI. At the end of the season, crops are harvested, leading to a sharp decrease in vegetation, which is captured by the decreasing slope of NDVI. These features are specific to the vegetation cycle and do not relate to a specific time, thus making it possible to take into account the high latitudinal gradient and the diversity of agriculture practices (crop type, management). PROBA-V reflectances were smoothed with a Whittaker filter to fill the gaps and remove the residual noise [50]. Some studies confirmed the adequacy of this smoother for time series filtering [51,52]. These smoothed reflectance values were extracted at a specific stage of temporal features for each pixel of the study area. The max red feature for instance was derived from the reflectances (4 bands) of the date at which the red reflectance reached a maximum for each pixel. Hence, neighboring pixels can belong to different dates, but correspond to the same phenological event. This led to twenty inputs for classification (five temporal features and four spectral bands) called spectral-temporal features (Figure 7).



**Figure 7.** (a) Spectral response of the four bands (blue, red, NIR, SWIR) of the PROBA-Vegetation instrument; (b) Representation of the five temporal features (minimum NDVI, maximum NDVI, increasing slope, decreasing slope and maximum red).

### 3.1.2. Trimming and Local Training

Due to the ability of SVM to handle small training sets [53], a trimming procedure was performed to remove misleading data due to the absence of an error-free and up-to-date labeling layer and recent land cover change. The trimming used spectral values of all classes to identify mislabeled pixels in a specific class. This was done by clustering so as to select pixels with the highest probability of belonging to a specific cluster. A random sample was used to select pixels from two sets of pixels: (i) pixels of class  $i$ ; (ii) pixels from all classes except  $i$ . The clustering classification identified clusters of class  $i$  that belonged to a different class. Following the clustering, the probability of each cluster to be of class  $i$  was assessed. The probability of each cluster to be of class  $i$  corresponds to the purity of the cluster, *i.e.*, the number of pixels of class  $i$  divided by the total number of pixels for a given cluster. Samples of group  $i$  belonging to the cluster with higher than a 75% probability of being classified as  $i$  were selected as the training dataset for this specific class  $i$ . This approach was repeated for each class. The class proportion in the training set was maintained similar to that initially found in the labeling layer. Additional cleaning was applied to discard pixels belonging to a class that deviated strongly from the distribution (95% of the confidence interval) of that class in at least one of the four bands of the five features.

### 3.1.3. SVM Classification

The trimmed data training set was used to calibrate the SVM classifier (for more details, see [54]). SVM searches the optimal separating hyperplane in a multidimensional space that maximizes the margin between the defined plan and the data. The one-against-one method was used to perform multiclass-classification with  $i$  classes on the SVM. Therefore,  $i(i-1)/2$  binary classifiers were trained, and the final class was obtained through a voting procedure. The SVM classifier identifies the hyper-plan using support vectors and margins. The cost-support vector classification was trained with a Gaussian radial basis kernel function to ensure a high level of automation. The  $\gamma$  parameter of the SVM was set at ten to optimize performance and computation time [40,55].

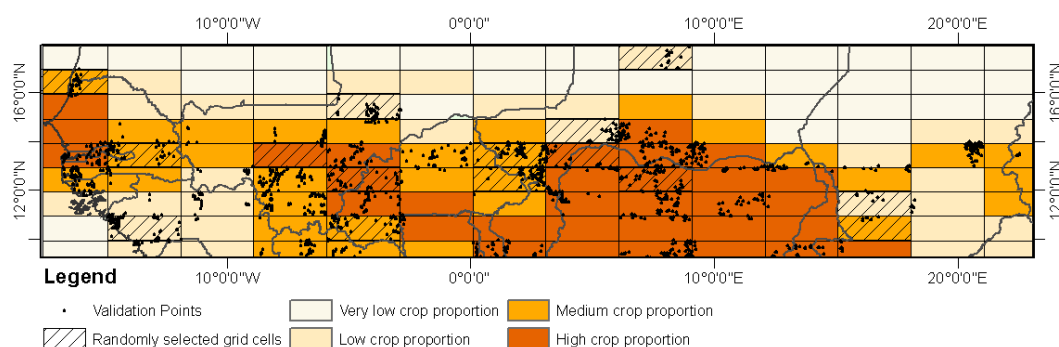
## 3.2. Handling the Spatial Gradient and the Landscape Diversity

The combination trimming-training-classification was applied on a moving window of  $1 \times 3^\circ$  to handle the latitudinal gradient. In fact, the vegetation dynamic in cropland and natural vegetation can fluctuate strongly with latitudes [56–59]. To limit the boundary effect between grid cells, the training

was performed on a larger grid cell. A buffer of  $1^\circ$  in longitudinal direction and of  $0.3^\circ$  in latitudinal direction was chosen as a compromise between processing time, high latitudinal gradient and recovery rate (40%). To mosaic smoothly the moving window results, five staggered cropland classifications were compiled by majority voting to obtain regional consistent cropland.

### 3.3. Relative Importance of Spectral-Temporal Features

A preliminary analysis of the respective spectral-temporal features' contribution was completed to identify the subset of the twenty spectral-temporal features that maximizes the classification accuracy. Twenty spectral-temporal features could increase both noise and processing time. A random forest algorithm was run to calculate the spectral-temporal feature importance by turning off one feature (while others remained constant) and computing the resulting loss in accuracy through the mean decrease in the Gini index (GI) [60]. This analysis carried out for four classes of crop proportions calculated on quantiles: very low crop proportion ( $<1.4\%$ ), low crop proportion ( $1.4\%–4.4\%$ ), medium crop proportion ( $4.4\%–17.8\%$ ) and high crop proportion ( $17.8\%–72.2\%$ ) (Figure 8). The distribution of crops varies across a latitudinal gradient with a higher proportion of crops between  $10^\circ$  and  $13^\circ$  of latitude.



**Figure 8.** Crop proportion of each grid cell and randomly-sampled grid cells for SVM classification with the best selected features. A higher crop proportion is observed in lower latitudes.

The feature importance analysis was computed for each grid cell within a crop proportion class. Then, a random sample of five grid cells per crop proportion class was used to perform the classification with the best selected features (Figure 8). The best features were selected according to a decision rule; the mean decrease in GI must exceed the mean of the mean GI decrease for a given grid cell. The accuracy of SVM classification with the best selected features was compared to SVM classification with all features using a Student test computed on the F-score.

### 3.4. Validation

A two-stage sampling was designed to collect the validation sample. The first sampling units were 40 grid cells within the 126 grid cells of the study area. The second sampling unit was a stratified random sampling, as recommended by Wagner and Stehman [61], within the selected grid cells. The strata corresponded to the eleven land cover classes of the labeling layer. The sample size allocated to each stratum was twenty pixels. All of the random 100-m by 100-m pixels automatically generated were visually assessed with 2014 images in Google Earth. Each random pixel was labeled by visual interpretation in one of the two classes: crop and non-crop, depending on whether more than 50% of the pixel was covered by the specific class. The availability of 2014 images was the main limiting factor in building this dataset. A set of 2315 points with 1616 non-crop and 699 crop validation samples had finally been collected. In total, 30% of the validation points belong to the crop class. This proportion (30%) matches the magnitude of the crop proportion in this region observed on the global product (crop proportion ranging from 13% to 42%).



Typically, the accuracy of a map is assessed by measuring the degree of agreement between the classification output and the validation dataset. A confusion matrix was used to derive the metrics of classification accuracy [62]. Liu *et al.* [63] recommended amongst fourteen category-level and twenty map-level accuracy measures the user's accuracy (UA), producer's accuracy (PA) and overall accuracy (OA) as primary accuracy indices. The OA was computed as the ratio of the sum of all correctly-classified pixels on the sum of all validation pixels. The F-score was calculated as a combination of PA and UA for a specific land cover class (Equation (1)). The UA for a specific class is the ratio between correctly-classified pixels and all pixels classified as the specific class. The PA for the specific class is the fraction of correctly-classified pixels and all ground truth-specific class pixels. Because of this research focus on cropland, the F-score was only computed for crop class and is called the F-score in the following sections.

$$\text{F-score} = \frac{2 \times \text{PA} \times \text{UA}}{\text{PA} + \text{UA}} \quad (1)$$

### 3.5. Error Analysis

The accuracy assessment described in Section 3.4 provides a global performance evaluation, although it is well established that classification accuracy varies across space and that errors are not equally distributed spatially [59,64–66]. Eight potential explanatory variables were proposed to explain the classification accuracy computed by OA and the F-score: latitude and longitude of the grid cell center, availability of cloud-free data and five landscape metrics indices. Within the five landscape metrics, some are specific to crops, such as crop proportion, Matheron index and crop fragmentation, while others are related to the overall landscape, such as fragmentation and entropy. These landscape features were measured on the labeling layer in order to explain the spatial distribution of errors. The correlation and the importance rank computed with the random forest algorithm were calculated for all of the potential explanatory variables.

The perimeter area ratio for all classes of land cover is used as a proxy of landscape fragmentation:

$$\text{Perimeter Area Ratio} = \frac{\sum_{i=1}^n \text{Perimeter}_i}{\sum_{i=1}^n \text{Area}_i} \quad (2)$$

where  $i$  represents a specific class of land cover and  $n$  the total number of land cover classes.

The Shannon entropy index measures the diversity of the landscape due to the number of land cover classes and the proportions of those classes (Equation (3)) [67]. In general, this index increases when the number of classes is higher and when the proportions of all existing classes are equal.

$$\text{Shannon Entropy Index} = \sum_{i=1}^n P_i \times \ln(P_i) \quad (3)$$

where  $P_i$  is the proportion of a class  $i$ .

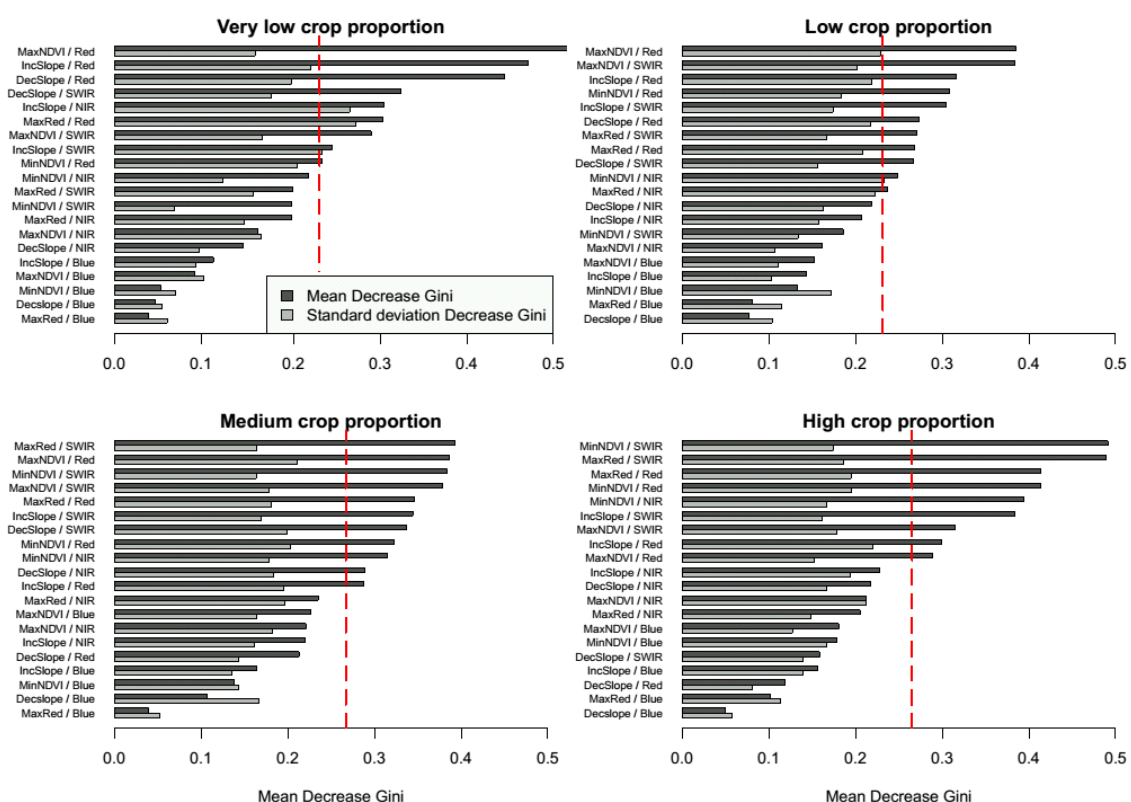
The Matheron index measures the ratio between the total outer perimeter of crop cells edges (sum for all patches) and the product between the area of crop and the total area [68]:

$$\text{Matheron Index} = \frac{\text{Crop Perimeter}}{\sqrt{\text{Crop Area}} \times \sqrt{\text{Total Area}}} \quad (4)$$

## 4. Results

### 4.1. Spectral-Temporal Feature Importance

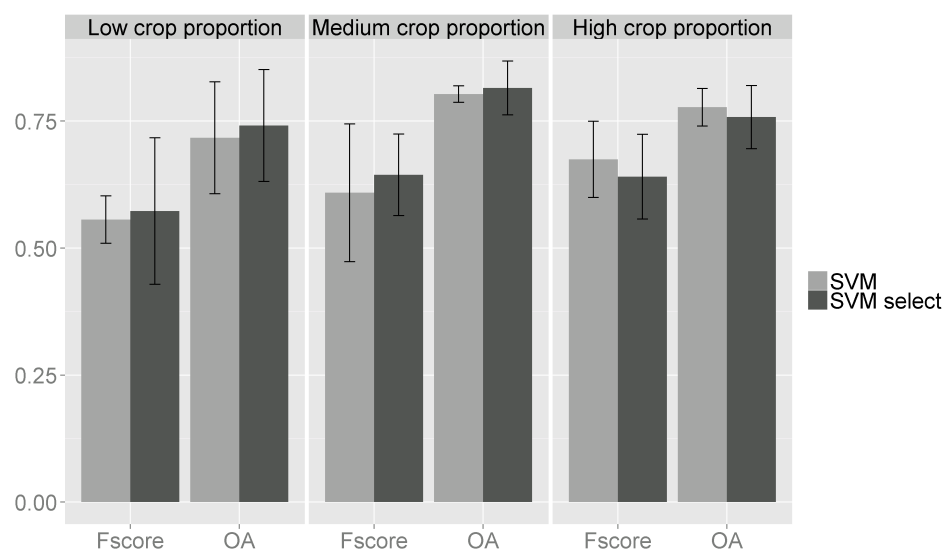
First, the diversity of the features seems to complement them according to the cropland proportion. The maximum in red reflectance and the minimum NDVI seem the two most discriminant features in higher crop proportion regions (Figure 9). These features refer to the start of the growing period when differences between cropland and natural vegetation are high due to land preparation. In fact, high crop proportion areas correspond to regions with higher amounts of rainfall, which makes natural vegetation greener at the beginning of the season when crops are considered as bare soils due to land preparation. All temporal metrics seem important in one or another crop proportion class without a specific distinction for any one of them. The mean decrease in GI was computed for the 20 spectral-temporal features for each of the four crop proportion classes. When a given spectral-temporal feature is left out, the higher mean decrease in GI compared to the GI obtaining using all 20 features leads to the higher importance of this spectral-temporal feature. Regardless of the crop proportion classes, as expected, the blue band contributes the least in cropland discrimination due to the impacts of aerosols and atmospheric scattering [69] (Figure 9). The SWIR band is of higher importance than the NIR band, while the red and the SWIR bands are the two most important bands for the classification.



**Figure 9.** Respective contribution of each spectral-temporal feature for crop discrimination for the different cropland density classes. The dotted red line corresponds to the decision rule for selecting the best features.

Following the feature importance analysis, the best features were selected (SVM select) to compare classification accuracy with regards to the classification using all of the features (SVM all). Due to the absence of validation points in the very low crop proportion area, only three crop proportion classes were considered, and five grid cells were randomly selected in each crop proportion stratum. OA and F-score were computed for the two classifications: SVM select and SVM, both by crop proportion

classes. SVM select classification slightly outperforms SVM for low and medium crop proportion strata, while SVM excels in the high crop proportion landscape (Figure 10).



**Figure 10.** Comparison of the accuracy for SVM classification with all features (SVM) and selected features (SVM select). The whiskers represent the standard deviation of the five selected grid cells for each crop proportion class.

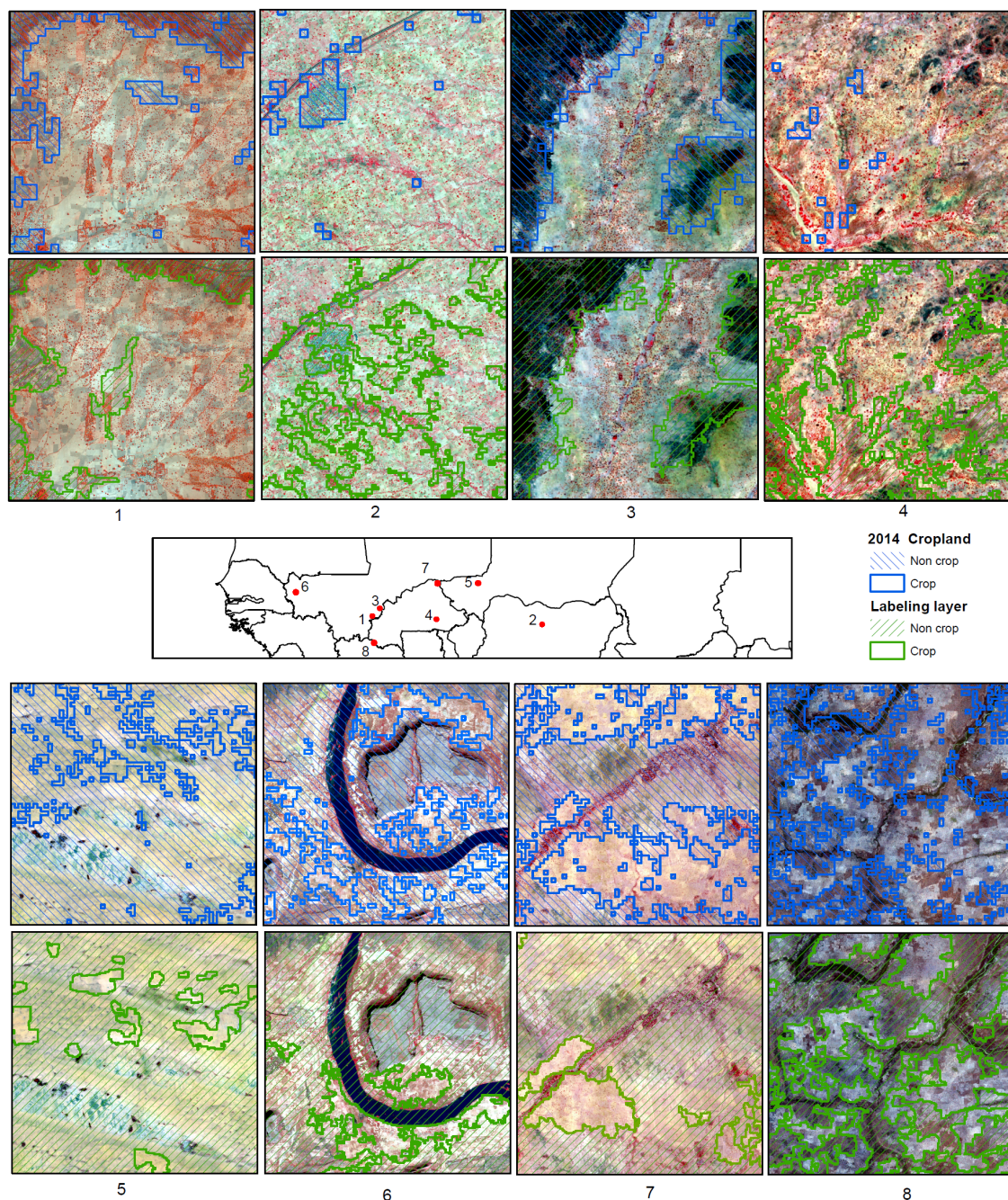
Standard deviations are rather larger for SVM select than for SVM. The two classifications were compared using a Student test on the F-score. The assumptions of the test were verified: the normality of the variable and the equality of the variances of both samples were confirmed by a Shapiro-Wilk test ( $p$ -value = 0.735) and a Fisher test ( $p$ -value = 0.872), respectively. The two F-score values are considered equal ( $p$ -value = 0.873), which avoids rejecting the null hypothesis of the equality of samples. Based on these results and to make the method as generic as possible for all agricultural landscapes, the SVM classification was applied at the regional scale using all twenty spectral-temporal features.

#### 4.2. Qualitative Analysis of 2014 Sudano-Sahelian Cropland map

The visual comparison of the 2014 *Sudano-Sahelian Cropland map* and the labeling layer, mostly derived from GlobeLand 30, highlights the actual contribution of the PROBA-V classification. The images used for this analysis are a combination of Digital Globe images acquired during 2014, Spot 5 and RapidEye imagery for the years 2012 and 2013. Images 1 and 2 are respectively a World View 2 of 26 June 2014 and a GeoEye of 14 June 2014. Images 3–7 are Spot 5 of the 2013 season, while Image 8 is a RapidEye image acquired on 29 January 2012 (Figure 11).

Cropland boundaries are more precise in the labeling layer, probably due to higher spatial resolution, while the non-cultivated lithosols are better delineated by the 2014 *Sudano-Sahelian Cropland map* (Image 1, in Figure 11). Roads are better defined by the labeling layer, which, however, includes more noise and incorrect non-crop area. The village patches are larger in the 2014 *Sudano-Sahelian Cropland map*, but unlike the labeling layer, it does display the full extent of the village (Image 2 in Figure 11). The contrasted landscape shown in image 3 with cultivated valleys and rocky plateaus yields an accurate classification in both cases. The labeling layer with the higher spatial resolution even captures the forest in the very bottom of the valley as non-crop (Image 3 in Figure 11).





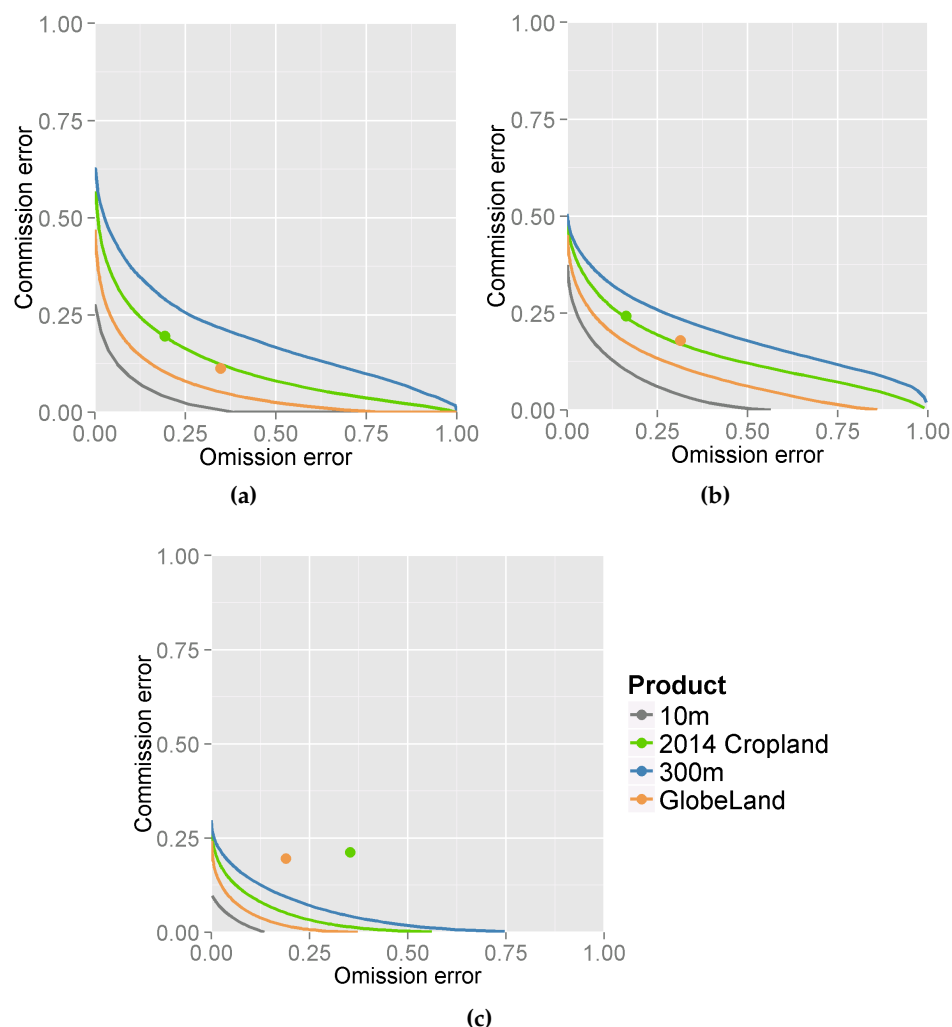
**Figure 11.** Some examples of cropland in the 2014 *Sudano-Sahelian Cropland map* (at 100-m) and the labeling layer (at 30-m). Background images are: (1) World View 2 of 26 June 2014; (2) GeoEye of 14 June 2014; (3–7) Spot 5 of the 2013 season; and (8) RapidEye of 29 January 2012.

In a more complicated landscape, the 2014 *Sudano-Sahelian Cropland map* failed to discriminate bare soil areas (in pink in Image 4) and dark surfaces (in black), while some of the cropland is mapped as non-crop in the labeling layer (Image 4 in Figure 11). In arid areas, errors observed in both datasets are probably due to the labeling layer, which drives the classification training. Crops in these arid areas are very sparse, and discrimination between bare soil and crop is challenging (Image 5, in Figure 11). This is confirmed by Image 7 in the north of Burkina Faso, where crops alternate with bare soil. The labeling layer missed considerable cropland areas, while these are correctly mapped by the 2014 *Sudano-Sahelian Cropland map*. Hence, in this area, spectral-temporal features allow the



better capture of crop areas thanks to the temporal dimension of the PROBA-V time series (Image 7, in Figure 11). In irrigated regions, the labeling layer failed to capture all cropland areas. The 2014 *Sudano-Sahelian Cropland map* captures more crops, but also more noise (Image 6, in Figure 11). Finally, Image 8 sums up what is often found in these illustrations. The 2014 *Sudano-Sahelian Cropland map* seems to better capture all of the cropland areas, but has more difficulty in drawing exact boundaries and in capturing small landscape features, such as forest in the bottom valley. Conversely, the labeling layer is sometimes too restrictive and misses some of the cropland (Image 8, Figure 11). To conclude, difficulties in differentiating between cropland and fallows, as well as misclassification of shrubland, bare soil and desert area as cropland remain in the final product.

Supervised classification with visual determination of the region of interest (ROI) was performed on each of these sites. Wall to wall validation between high resolution images and GlobeLand 30 or 2014 *Sudano-Sahelian Cropland map* was assessed. Proximity between the commission-omission point of the 2014 *Sudano-Sahelian Cropland map* and the Pareto boundary implies a high performance of the algorithm classification (Figure 12).



**Figure 12.** Three examples of Pareto boundary, commission and omission errors for GlobeLand 30 and 2014 *Sudano-Sahelian Cropland map*. Pareto boundaries were also computed for a 10-m product and a 300-m product. (a) Site in center Mali ; (b) Site in Sikasso region (Mali South East); (c) Site in South Mali.

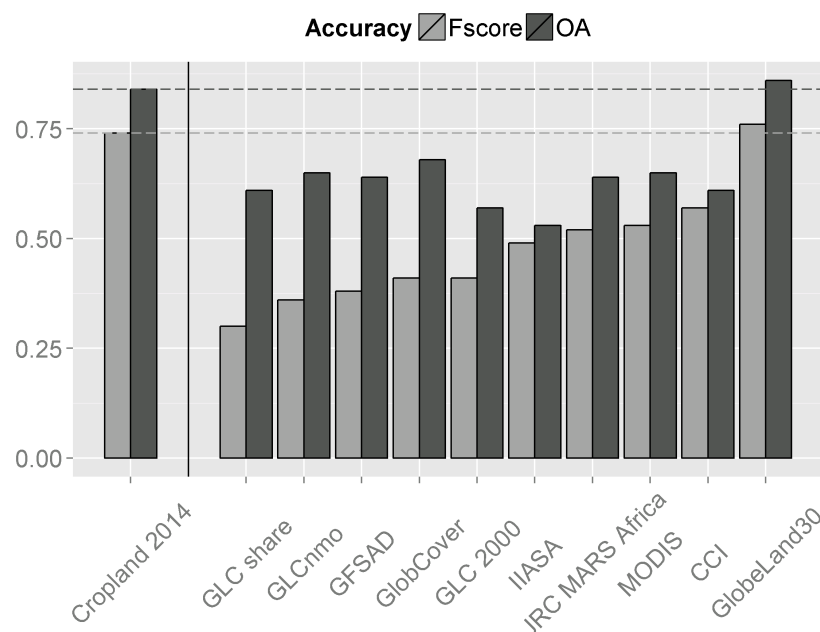
To better quantify the visual interpretation of the 2014 *Sudano-Sahelian Cropland map* and the labeling layer, quantitative analysis of the accuracy of the 2014 *Sudano-Sahelian Cropland map* was assessed. The produced cropland was also compared to other existing global products.

#### 4.3. Accuracy of the Cropland Map and Comparison with Existing Global Products

The overall accuracy computed from the 2315 validation points reaches 84% and the F-score for crop class 74%. The confusion matrix shows better PA and UA for the non-crop class: 89% compared to 74% for crop classes. UA and PA are identical for both classes (Table 2). The same validation dataset was applied on the crop/non-crop maps derived from the 10 previous global products. All products were resampled to the 100-m resolution using the majority method. A high variability of accuracies was observed regardless of the production year of the map (Figure 13).

**Table 2.** Confusion matrix for the regional cropland map using the 2315 validation points. Overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA) are computed in percent.

	Non Crop	Crop	UA [%]
Non crop	1431	180	89
Crop	185	519	74
PA [%]	89	74	OA[%] = 84

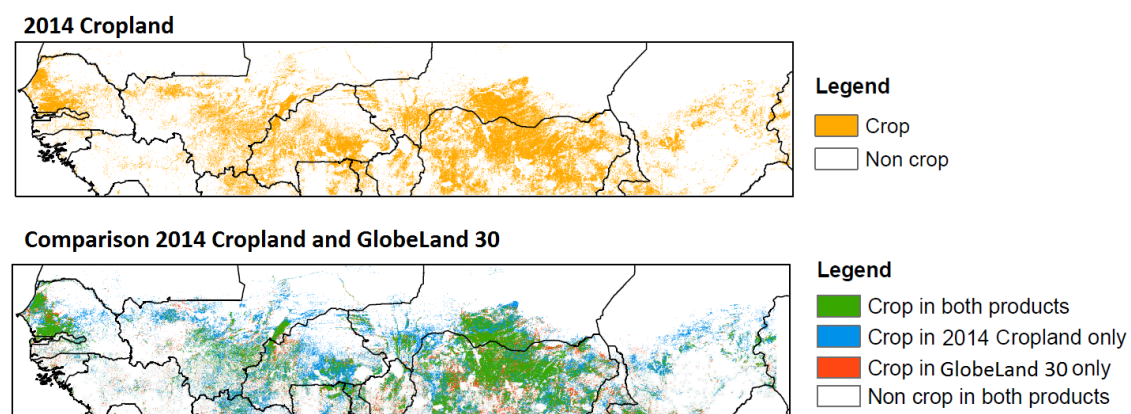


**Figure 13.** Comparison of the 2014 *Sudano-Sahelian Cropland map* with the ten previous global products showing high variability in the accuracy indices and better performance for the 2014 *Sudano-Sahelian Cropland map* and GlobeLand 30 used as the labeling layer.

The 2014 *Sudano-Sahelian Cropland map* and the GlobeLand 30 product almost reach the same performance levels (OA: 84% vs. 86%; and F-score: 74% vs. 75%, respectively). Other products studied show substantially lower accuracy indices. This confirms the adequacy of the labeling layer and the efficiency of the methodology in exploiting recent annual remote sensing datasets. Similarities of the accuracy indices of GlobeLand 30 and our cropland product must be balanced against disagreement between those products (Figure 14). Eighty five percent of pixels are labeled identically by GlobeLand 30 and the 2014 *Sudano-Sahelian Cropland map*, while 41% of pixels classified as crops in one or another

are mapped as crops in both products. As expected, because of the higher OA and F-score of GlobeLand 30 for the entire validation dataset (Figure 13), the OA in the conflict region (414 validation pixels) reaches 54% for GlobeLand 30 and 46% for the 2014 *Sudano-Sahelian Cropland map*.

However, one might argue that these measures are excessively harsh, as the size of the validation pixels for the >100-m products is smaller than the products' pixel sizes. They might be revised when considering a validation pixel size that matches the products. Nonetheless, the variation in OA when validating a product at 300-m in place of one at 100-m with the 100 squared meter validation pixels is only of 2%. This number rise to 6% when a 1-km product is used, which still cannot explain the total discrepancies between global land cover products. It was computed by aggregating the labeling layer at 100-m, 300-m and 1-km, defining the label of each pixel as crop or non-crop when more than 50% is covered by the specific class and, finally, computing accuracy.



**Figure 14.** The 2014 *Sudano-Sahelian Cropland map* derived from PROBA-V at 100-m and the comparison with GlobeLand 30 derived from Landsat imagery acquired around 2010.

#### 4.4. Spatial Distribution of Errors

Grid cells including more than twenty validation points, *i.e.*, 38 grid cells, were selected to estimate locally the OA and the F-score. The eight possible explanatory variables were also computed for each grid cell. In total, 41% of the variance of the OA is explained by the eight potential explanatory indices for only 21% of the F-score (Table 3). OA and F-score are strongly linked to fragmentation; the lower the fragmentation, the more accurate the classification. Entropy is negatively correlated with both F-score and OA. This means that the lower the number of classes and the higher the proportion in the classes, then the higher is the accuracy. As expected, the accuracy is highest when few classes are present with small fragmentation and a high proportion in each class. Spatial heterogeneity is confirmed as an important factor in driving land cover map accuracy [70,71].

Looking at the spatial distribution of latitude impacts both the F-score and the OA: it increases the accuracy in northern regions where less clouds are present. Conversely, longitude is correlated with OA, so that accuracy rises when going west (Table 3). This might be due to the fact that isohyets move south when going east, resulting in greener and more vegetated areas in the west of the study area. Hence, misclassification issues in less vegetated areas, such as desert or bare soil, could lead to lower accuracy indices. Specific cropland characteristics do not affect the F-score more than OA in terms of correlation, which was unexpected. Nevertheless, the importance rank computed by random forest highlighted the relevance of crop-specific indices for cropland classification (Table 3). The Matheron index and crop proportion are negatively related to accuracy because a higher fragmentation and a higher number of crop patches make classification difficult due to the number of mixed pixels at the edges [64]. Data availability affects OA and F-score. The encountered errors were partially explained by the high fragmentation of the landscape, and the low latitude linked with lower data availability.

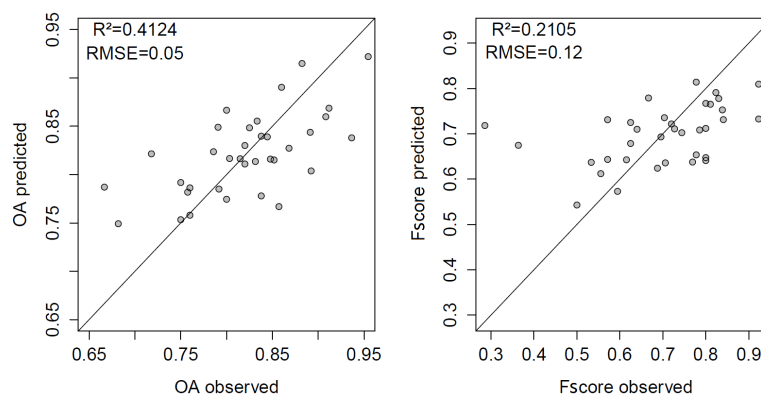
Very few validation points were available in high latitudes due to the absence of high resolution imagery in these regions.

**Table 3.** Correlation and importance rank for Overall accuracy (OA) and F-score computed on eight potential explanatory variables.

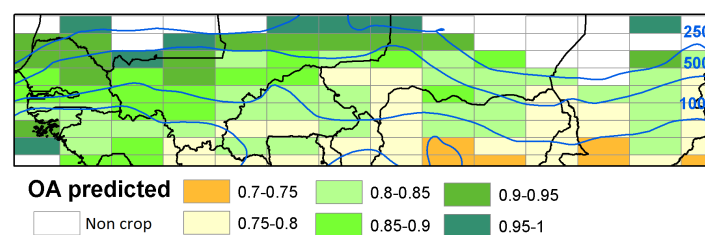
	OA		F-score	
	Correlation	Ranking	Correlation	Ranking
<b>Location</b>				
Latitude	0.44	2	0.33	3
Longitude	−0.28	5	−0.09	7
<b>Time-series</b>				
Data availability	0.25	3	0.22	1
<b>Landscape characteristics</b>				
Fragmentation	−0.39	1	−0.2	6
Entropy	−0.13	6	−0.1	8
Matheron Index	−0.29	4	−0.05	5
Crop proportion	−0.05	8	0.09	2
Crop fragmentation	−0.24	7	0.01	4
Total variance explained [%]	41.24		21.05	

#### 4.5. Multiple Linear Regression to Explain OA and F-Score

A multiple linear regression based on eight explanatory variables explains 41.24% of OA variance and 21.05% of F-score variance (Figure 15). This regression ( $R^2 = 0.4124$ ) was then applied to the entire study area to predict the OA (Figure 16). According to the model, a degradation of OA is observed in the southeast of the study region.



**Figure 15.** OA and F-score predicted by the multiple regression compared to the OA and F-score observed.

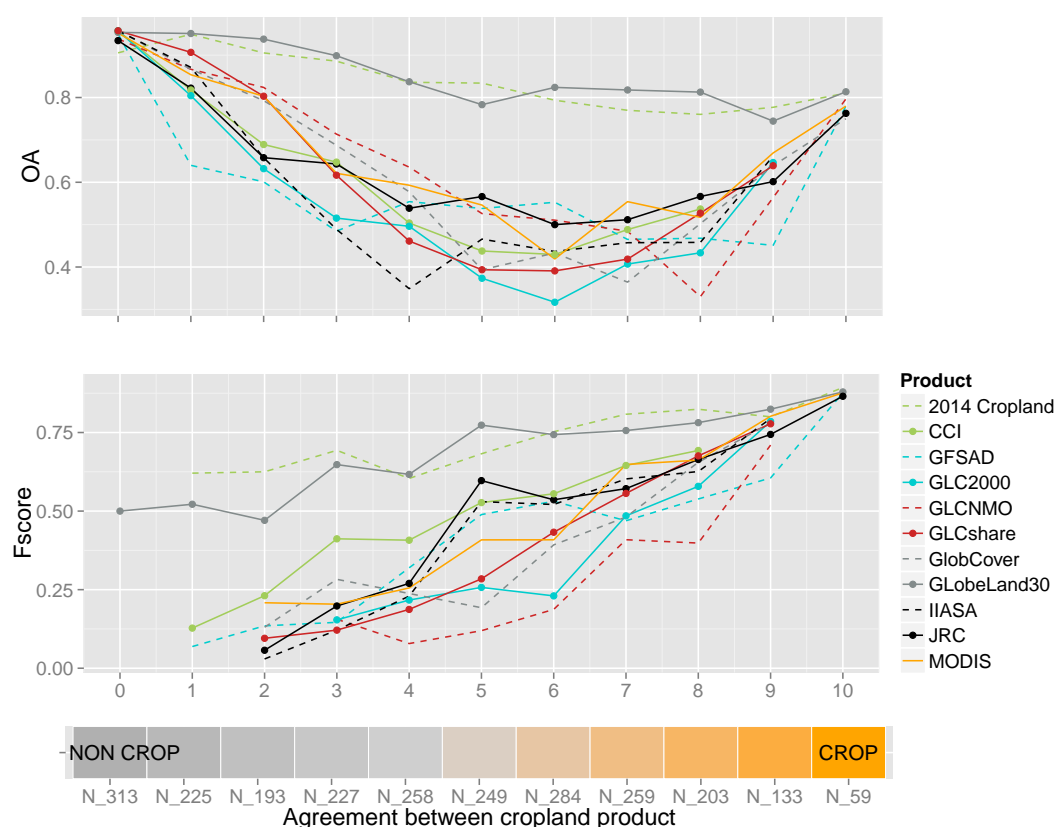


**Figure 16.** Prediction of OA over the study area based on a multiple regression using eight explanatory variables related to spatial localization, data availability and landscape characteristics.



#### 4.6. OA and F-Score in the Disagreement Region of Global Products

To further identify the accuracy of the products, it is of paramount interest to focus the analysis on challenging areas where most discrepancies between maps are observed. A global agreement map built from the global products (see Table 1) ranges from zero, where all products agree on the non-crop class, to ten, where all products agree on the crop class. A value of five represents the highest disagreement between products as five products classified the pixel as crop and five classified the pixel as non-crop. OA and F-score were computed for each of the eleven classes of agreement using the validation points belonging to that class. The OAs of all global products, except GlobeLand 30 and the 2014 Sudano-Sahelian Cropland map, have a similar behavior (Figure 17). The highest accuracy of the majority of products is met in regions labeled as non-crop by all products, and accuracy then drops until values around 0.45, when most products disagree. Finally, OA increases to yield an accuracy of 0.8 when all products agree on crop class. GlobeLand 30 and the cropland map produced in this study have a nearly identical profile, which could be summarized as a slight decrease in accuracy when agreement on crop arises. F-score increases when agreement on crop occurs. GlobeLand 30 and the 2014 Sudano-Sahelian Cropland map behave very differently to others. The F-score of the new product and GlobeLand 30 starts with a higher value in regions where many products agree on non-crop. This means that these two products correctly label crops in regions where almost all global products mislabeled crop as non-crop.



**Figure 17.** Evolution of OA and F-score with agreement computed on ten global products. The X axis corresponds to agreement between global products with 0 = all products classify pixels as non-crop and 10 = all products classify pixels as crop; N corresponds to the number of validation points corresponding to a given class of agreement.

#### 4.7. Fragmentation of the Landscape

The proportion error defined by [72] represents the proportion by which individual classes are over- or under-estimated at a lower resolution than the initial one (Equation (5)):

$$E_r = (p_r - p_0) / p_0 \quad (5)$$

where  $p_0$  is the original proportion of the crop class at 30-m and  $p_r$  is the estimated crop proportion at resolution  $r$  (90-m or 300-m in this study). The 90-m cropland map is obtained by aggregating the 30-m for three pixels and the 300-m for 10-pixels. The label of this obtained pixel is crop when more than 50% of the area is classified as crop. The 90-m products are considered to be very close to the 100-m resolution, as available for the PROBA-V sensor. In low and medium crop proportion classes, the underestimation of crops is of 5% for the 90-m products and more than 25% for the 300-m products (Table 4). The advantage of 100-m in a highly fragmented landscape is useful and can fill the gap between high resolution at medium resolution.

**Table 4.** Proportion error for two cases (initial resolution = 30-m; and lower resolution = 90-m or 300-m) within the four crop proportion classes.

Crop Proportion	Very Low	Low	Medium	High
Proportion error (30 m–90 m) (%)	NA	−5.2	−5.2	−0.6
Proportion error (30 m–300 m) (%)	NA	−29.8	−27.0	−5.7

## 5. Discussion

The proposed classification method of PROBA-V 100-m time series yields an overall accuracy of 84% similar to that of GlobeLand 30 from which training data were derived. Both products overpass all other global products in terms of accuracy. The accuracy of GlobeLand 30 ought to be linked to its spatial resolution of 30-m and a highly interactive verification procedure [17]. GlobeLand 30 is produced with a POK-based (pixel object knowledge) methodology, a pixel and object-based method with knowledge. The knowledge-based interactive verification procedure of GlobeLand 30 was developed through a web service and relies on intensive manual corrections. The fully-automated method presented aims at overcoming this limitation by enabling the production of cropland maps on an annual basis. Three main factors can explain the accuracy of the 2014 *Sudano-Sahelian Cropland map*: the 100-m spatial resolution of the time series, the methodology itself and the quality of the labeling layer. The hectometric spatial resolution makes it possible to resolve fine landscape patterns, especially in those regions where field size often does not exceed 1 ha. Similarly, a recent study has highlighted significant better performance using 100-m PROBA-V data than 300-m data for crop type classification [73].

Misclassifications are mainly explained by low data availability due to high cloud cover during vegetation growth (25% of OA explained by data availability) and by landscape fragmentation (39% of OA explained by fragmentation). Remaining errors in the training dataset are minimized by the stringent trimming procedure limits, from which 30% of the best points were retained. Highly fragmented landscape is better captured by high resolution images as demonstrated with Pareto boundaries. The Sentinel 2, 5-day coverage at 10-m in 13 spectral bands, will probably offer a good combination of spatial-spectral and temporal resolution. Landsat 8, 16-day coverage at 30-m in 11 spectral bands, offers a spatial resolution able to catch the landscape fragmentation of the study region. Nevertheless, the temporal coverage might be too small to catch vegetation behavior during the vegetation growth. Remaining errors in the training dataset are minimized by the strong trimming procedure limits. Discrepancies between the GlobeLand 30 product and the 2014 *Sudano-Sahelian Cropland map* can partly be linked to the time production of the map, the spatial resolution, the input data and the interactive verification for GlobeLand 30.

Spectral-temporal features were used to fully exploit the PROBA-V time series. Nevertheless, remaining noise in the PROBA-V surface reflectance image could partly explain some misclassification. The noise might be related to undetected clouds and shadows. PROBA-V cloud screening uses both blue and SWIR bands. Due to the time lag between bands (12 s between NIR and SWIR), two separate masks are built and then merged [74]. Threshold rules are applied on the SWIR (blue) band to detect cloud with an additional check on the blue (SWIR) band. These threshold values are static and applied globally, leading to misclassifications of bright surfaces as clouds [43]. Furthermore, borders of large clouds are poorly detected, and the remaining haze effect affects the reflectances. Some cases of cloud omission, others of cloud commission, are also present. In addition, part of the noise is related to the spectral temporal features themselves. Spectral-temporal features are based on extreme values and are thus more sensitive to noise, as noise itself is characterized by extreme values [66]. Region-dependent and dynamic cloud screening are currently being investigated by the PROBA-Vegetation team and could significantly improve the cloud mask [75]. This could improve feature extraction and thus classification by significantly decreasing noise in the time series. If no significant improvements are made to the cloud product delivered, we believe that a stronger temporal filter should be applied prior to performing the Whittaker filter in order to identify undetected clouds.

## 6. Conclusions

This fully-automated method has delivered the first cropland at 100-m for the Sahelian and Sudanian region using the PROBA-V sensor. The product reaches an overall accuracy of 84% and an F-score of 74% for the crop class. This accuracy indices are significantly larger than major global products (except GlobeLand 30 which was used as labeling layer) thanks to the spatial resolution of 100-m, the crop-specific knowledge-based methodology and the adequate choice of the labeling layer. The diversity of the spectral-temporal features chosen to take advantage of the entire PROBA-V time series and its full spectrum seems to adequately complement each others for cropland discrimination in heterogeneous crop proportion landscapes. Misclassifications observed in the *2014 Sudano-Sahelian Cropland map* are partially explained by undetected clouds, haze and shadows, data availability and fragmentation of the landscape.

Further improvements are expected with the upcoming enhanced cloud screening. The recent launch of the European Sentinel-2 sensor offers new possibilities for monitoring land and vegetation with its spatial resolution of 10-m and its 13 spectral bands. The methodology is scale independent and could be applied on Sentinel-2 time series to better capture the landscape specificities in the Sudano-Sahelian regions. Annual cropland maps could be updated on a regular basis using this fully automated method in order to monitor cultivated area extension and abandonment over a large area.

**Acknowledgments:** This research was funded by a FNRS (Fonds National de Recherche Scientifique belge) grant for M-J Lambert. The authors thank the PROBA-V team for free access to the PROBA-V 100-m time series. The authors also would like to thank the STARS project for access to Digital Globe imagery and ESA (European Space Agency) for access to RapidEye and SPOT 5 imagery.

**Author Contributions:** Marie-Julie Lambert and François Waldner designed and carried out the all the present research under the supervision of Pierre Defourny.

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

## References

1. CORAF/WECARD. *Strategic Plan 2007–2016*; CORAF/WECARD: Dakar, Senegal, 2007.
2. Brevault, T.; Renou, A.; Vayssieres, J.; Amadji, G.; Assogba Komlan, F.; Dalanda Diallo, M.; De Bon, H.; Diarra, K.; Hamadoun, A.; Huat, J.; *et al.* DIVECOSYS: Bringing together researchers to design ecologically-based pest management for small-scale farming systems in West Africa. *Crop Prot.* **2014**, *66*, 53–60.

3. Lahmar, R.; Bationo, B.; Dan Lamso, N.; Guero, Y.; Tiftonell, P. Tailoring conservation agriculture technologies to West Africa semi-arid zones: Building on traditional local practices for soil restoration. *Field Crops Res.* **2012**, *132*, 158–167.
4. Jalloh, A.; Roy-Macauley, H.; Sereme, P. Major agro-ecosystems of West and Central Africa: Brief description, species richness, management, environmental limitations and concerns. *Agric. Ecosyst. Environ.* **2012**, *157*, 5–16.
5. Akponikpe, P.; Minet, J.; Gerard, B.; Defourny, P.; Biielders, C. Spatial fields dispersion as a farmer strategy to reduce agro-climatic risk at the household level in pearl millet-based systems in the Sahel: A modeling perspective. *Agric. For. Meteorol.* **2011**, *151*, 215–227.
6. Dardel, C.; Kergoat, L.; Hiernaux, P.; Mougin, E.; Grippa, M.; Tucker, C. Re-greening Sahel: 30 years of remote sensing data and field observations (Mali, Niger). *Remote Sens. Environ.* **2014**, *140*, 350–364.
7. Genesio, L.; Bacci, M.; Baron, C.; Diarra, B.; Di Vecchia, A.; Alhassane, A.; Hassane, I.; Ndiaye, M.; Philippon, N.; Tarchiani, V.; *et al.* Early warning systems for food security in West Africa: Evolution, achievements and challenges. *Atmos. Sci. Lett.* **2011**, *12*, 142–148.
8. Karlson, M.; Ostwald, M. Remote sensing of vegetation in the Sudano-Sahelian zone: A literature review from 1975 to 2014. *J. Arid Environ.* **2015**, *124*, 257–269.
9. Potts, M.; Zulu, E.; Wehner, M.; Castillo, F.; Henderson, C. *Crisis in the Sahel: Possible Solutions and the Consequences of Inaction*; The OASIS Initiative: Berkeley, CA, USA, 2013.
10. Bailis, R.; Ezzati, M.; Kammen, D.M. Mortality and greenhouse gas impacts of biomass and petroleum energy futures in Africa. *Science* **2005**, *308*, 98–103.
11. Manning, A.D.; Gibbons, P.; Lindenmayer, D.B. Scattered trees: A complementary strategy for facilitating adaptive responses to climate change in modified landscapes? *J. Appl. Ecol.* **2009**, *46*, 915–919.
12. Mertz, O.; D'haen, S.; Maiga, A.; Moussa, I.B.; Barbier, B.; Diouf, A.; Diallo, D.; Da, E.D.; Dabi, D. Climate variability and environmental stress in the Sudan-Sahel zone of West Africa. *Ambio* **2012**, *41*, 380–392.
13. Liu, J.; Liu, M.; Tian, H.; Zhuang, D.; Zhang, Z.; Zhang, W.; Tang, X.; Deng, X. Spatial and temporal patterns of China s cropland during 1990 2000: An analysis based on Landsat {TM} data. *Remote Sens. Environ.* **2005**, *98*, 442–456.
14. Delrue, J.; Bydekerke, L.; Eerens, H.; Gilliams, S.; Piccard, I.; Swinnen, E. Crop mapping in countries with small-scale farming: A case study for West Shewa, Ethiopia. *Int. J. Remote Sens.* **2013**, *34*, 2566–2582, doi:10.1080/01431161.2012.747016.
15. Vancutsem, C.; Marinho, E.; Kayitakire, F.; See, L.; Fritz, S. Harmonizing and combining existing Land Cover/Land Use datasets for cropland area monitoring at the African continental scale. *Remote Sens.* **2012**, *5*, 19–41.
16. Fritz, S.; See, L.; Rembold, F. Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. *Int. J. Remote Sens.* **2010**, *31*, 2237–2256, doi:10.1080/01431160902946598.
17. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27.
18. Pittman, K.; Hansen, M.C.; Becker-Reshef, I.; Potapov, P.V.; Justice, C.O. Estimating global cropland extent with multi-year MODIS data. *Remote Sens.* **2010**, *2*, 1844–1863.
19. Defourny, P. *Land Cover CCI: Product user Guide: Version 2*; Technical Report; UCL, ESA: Louvain-la-Neuve, Belgium, 2014.
20. Teluguntla, P.G.; Thenkabail, P.S.; Xiong, J.; Gumma, M.K.; Giri, C.; Milesi, C.; Ozdogan, M.; Congalton, R.; Tilton, J.; Sankey, T.T.; *et al.* *Global Cropland Area Database (GCAD) Derived from Remote Sensing in Support of Food Security in the Twenty-First Century: Current Achievements and Future Possibilities*; Taylor & Francis: Boca Raton, FL, USA, 2015, in press.
21. Bartholomé, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sens.* **2005**, *26*, 1959–1977.
22. Ryutaro, T.; Nguyen, T.H.; Toshiyuki, K.; Bayan, A.; Gegen, T.; Dong, X.P. Production of global land cover data GLCNMO2008. *J. Geogr. Geol.* **2014**, doi:10.5539/jgg.v6n3p99.
23. Bontemps, S.; Defourny, P.; Van Bogaert, E. *GLOBCOVER 2009: Product Description and Dalidation Report*; Technical Report; UCL, ESA: Louvain-la-Neuve, Belgium, 2011.



24. Tchuente, A.T.K.; Roujean, J.L.; Faroux, S. ECOCLIMAP-II: An ecosystem classification and land surface parameters database of Western Africa at 1 km resolution for the African Monsoon Multidisciplinary Analysis (AMMA) project. *Remote Sens. Environ.* **2010**, *114*, 961–976.
25. Latham, J.; Cumani, R.; Rosati, I.; Bloise, M. *FAO Global Land Cover (GLC-SHARE) Beta-Release 1.0 Database*; FAO: Rome, Italy, 2014.
26. Biradar, C.; Thenkabail, P.; Noojipady, P.; Li, Y.; Dheeravath, V.; Turrall, H.; Velpuri, M.; Gumma, M.; Gangalakunta, O.; Cai, X.; *et al.* A global map of rainfed cropland areas (GMRCA) at the end of last millennium using remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 114–129.
27. Prasad, S.T.; Biradar, C.; Noojipady, P.; Dheeravath, V.; Li, Y.; Velpuri, M.; Gumma, M.; Gangalakunta, O.; Turrall, H.; Cai, X.; *et al.* Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. *Int. J. Remote Sens.* **2009**, *30*, 3679–3733, doi:10.1080/01431160802698919.
28. Wu, W.; Shibasaki, R.; Yang, P.; Zhou, Q.; Tang, H. Remotely sensed estimation of cropland in China: A comparison of the maps derived from four global land cover datasets. *Can. J. Remote Sens.* **2008**, *34*, 467–479, doi:10.5589/m08-059.
29. Waldner, F.; Fritz, S.; Di Gregorio, A.; Defourny, P. Mapping priorities to focus cropland mapping activities: Fitness assessment of existing global, regional and national cropland maps. *Remote Sens.* **2015**, *7*, 7959–7986, doi:10.3390/rs70607959.
30. Fritz, S.; See, L.; You, L.; Justice, C.; Becker-Reshef, I.; Bydekerke, L.; Cumani, R.; Defourny, P.; Erb, K.; Foley, J.; *et al.* The need for improved maps of global cropland. *Eos Trans. Am. Geophys. Union* **2013**, *94*, 31–32.
31. Bontemps, S.; Herold, M.; Kooistra, L.; Van Groenestijn, A.; Hartley, A.; Arino, O.; Moreau, I.; Defourny, P. Revisiting land cover observation to address the needs of the climate modeling community. *Biogeosciences* **2012**, *9*, 2145–2157.
32. Swain, P.; Davis, S. *Remote Sensing: The Quantitative Approach*; Advanced book program; McGraw-Hill International Book Co.: New York, NY, USA, 1978.
33. Hansen, M.; Dubayah, R.; Defries, R. Classification trees: An alternative to traditional land cover classifiers. *Int. J. Remote Sens.* **1996**, *17*, 1075–1081, doi:10.1080/01431169608949069.
34. Lippmann, R. An introduction to computing with neural nets. *IEEE ASSP Mag.* **1987**, *4*, 4–22.
35. Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **1991**, *21*, 660–674.
36. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259.
37. Szuster, B.W.; Chen, Q.; Borger, M. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Appl. Geogr.* **2011**, *31*, 525–532.
38. Pal, M.; Mather, P. Assessment of the effectiveness of support vector machines for hyperspectral data. *Future Gener. Comput. Syst.* **2004**, *20*, 1215–1225.
39. Paneque-Gálvez, J.; Mas, J.F.; Moré, G.; Cristóbal, J.; Orta-Martínez, M.; Luz, A.C.; Guèze, M.; Macía, M.J.; Reyes-García, V. Enhanced land use/cover classification of heterogeneous tropical landscapes using support vector machines and textural homogeneity. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 372–383.
40. Huang, C.; Davis, L.; Townshend, J. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **2002**, *23*, 725–749.
41. Sivakumar, M. Exploiting rainy season potential from the onset of rains in the Sahelian zone of West Africa. *Agric. For. Meteorol.* **1990**, *51*, 321–332.
42. Kang, B.T. *Moist Savannas of Africa—Potentials and Constraints for Crop Production*; ITA: Ibadan, Nigeria, 1995.
43. Dierckx, W.; Sterckx, S.; Benhadj, I.; Livens, S.; Duhoux, G.; Van Achteren, T.; Francois, M.; Mellab, K.; Saint, G. PROBA-V mission for global vegetation monitoring: Standard products and image quality. *Int. J. Remote Sens.* **2014**, *35*, 2589–2614.
44. Zhao, Y.; Gong, P.; Yu, L.; Hu, L.; Li, X.; Li, C.; Zhang, H.; Zheng, Y.; Wang, J.; Zhao, Y.; *et al.* Towards a common validation sample set for global land-cover mapping. *Int. J. Remote Sens.* **2014**, *35*, 4795–4814.
45. Boschetti, L.; Flasse, S.P.; Brivio, P.A. Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: The Pareto Boundary. *Remote Sens. Environ.* **2004**, *91*, 280–292.
46. Waldner, F.; Lambert, M.J.; Li, W.; Weiss, M.; Demarez, V.; Morin, D.; Marais-Sicre, C.; Hagolle, O.; Baret, F.; Defourny, P. Land cover and crop type classification along the season based on biophysical variables retrieved from multi-sensor high-resolution time series. *Remote Sens.* **2015**, *7*, 10400–10424.

47. Matton, N.; Canto, G.S.; Waldner, F.; Valero, S.; Morin, D.; Inglada, J.; Arias, M.; Bontemps, S.; Koetz, B.; Defourny, P. An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series. *Remote Sens.* **2015**, *7*, 13208–13232.
48. Waldner, F.; Canto, G.S.; Defourny, P. Automated annual cropland mapping using knowledge-based temporal features. *ISPRS J. Photogramm. Remote Sens.* **2015**, *110*, 1–13.
49. Akbari, M.; Mamanpoush, A.R.; Gieske, A.; Miranzadeh, M.; Torabi, M.; Salemi, H. Crop and land cover classification in Iran using Landsat 7 imagery. *Int. J. Remote Sens.* **2006**, *27*, 4117–4135.
50. Eilers, P. A perfect smoother. *Anal. Chem.* **2003**, *75*, 3631–3636.
51. Atkinson, P.M.; Jeganathan, C.; Dash, J.; Atzberger, C. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sens. Environ.* **2012**, *123*, 400–417.
52. Atzberger, C.; Eilers, P.H. Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements. *Int. J. Remote Sens.* **2011**, *32*, 3689–3709.
53. Foody, G.; Mathur, A. Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification. *Remote Sens. Environ.* **2004**, *93*, 107–117.
54. Wang, L. *Support Vector Machines: Theory and Applications*; Springer Science & Business Media: Singapore, 2005; Volume 177.
55. Vapnik, V.N.; Vapnik, V. *Statistical Learning Theory*; Wiley: New York, NY, USA, 1998; Volume 1.
56. Hiernaux, P.; Mougin, E.; Diarra, L.; Soumaguel, N.; Lavenu, F.; Tracol, Y.; Diawara, M. Sahelian rangeland response to changes in rainfall over two decades in the Gourma region, Mali. *J. Hydrol.* **2009**, *375*, 114–127.
57. Bégué, A.; Vintrou, E.; Saad, A.; Hiernaux, P. Differences between cropland and rangeland MODIS phenology (start-of-season) in Mali. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *31*, 167–170.
58. Bégué, A.; Vintrou, E.; Ruelland, D.; Claden, M.; Dessay, N. Can a 25-year trend in Soudano-Sahelian vegetation dynamics be interpreted in terms of land use change? A remote sensing approach. *Glob. Environ. Chang.* **2011**, *21*, 413–420.
59. Renier, C.; Waldner, F.; Jacques, D.C.; Babah Ebbe, M.A.; Cressman, K.; Defourny, P. A dynamic vegetation senescence indicator for near-real-time desert locust habitat monitoring with MODIS. *Remote Sens.* **2015**, *7*, 7545–7570.
60. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
61. Wagner, J.E.; Stehman, S.V. Optimizing sample size allocation to strata for estimating area and map accuracy. *Remote Sens. Environ.* **2015**, *168*, 126–133.
62. Foody, G.M. Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *Int. J. Remote Sens.* **2005**, *26*, 1217–1228, doi:10.1080/01431160512331326521.
63. Liu, C.; Frazier, P.; Kumar, L. Comparative assessment of the measures of thematic classification accuracy. *Remote Sens. Environ.* **2007**, *107*, 606–616.
64. Leroux, L.; Jolivot, A.; Bégué, A.; Seen, D.L.; Zoungrana, B. How reliable is the MODIS land cover product for crop mapping Sub-Saharan agricultural landscapes? *Remote Sens.* **2014**, *6*, 8541–8564.
65. Foody, G. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201.
66. Waldner, F.; Ebbe, M.A.B.; Cressman, K.; Defourny, P. Operational monitoring of the desert locust habitat with earth observation: An assessment. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2379–2400.
67. Turner, M.G.; O'Neill, R.V.; Gardner, R.H.; Milne, B.T. Effects of changing spatial scale on the analysis of landscape pattern. *Landsc. Ecol.* **1989**, *3*, 153–162.
68. Mayaux, P.; Lambin, E.F. Estimation of tropical forest area from coarse spatial resolution data: A two-step correction function for proportional errors due to spatial aggregation. *Remote Sens. Environ.* **1995**, *53*, 1–15.
69. Jia, K.; Liang, S.; Wei, X.; Yao, Y.; Su, Y.; Jiang, B.; Wang, X. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. *Remote Sens.* **2014**, *6*, 11518–11532.
70. Smith, J.H.; Stehman, S.V.; Wickham, J.D.; Yang, L. Effects of landscape characteristics on land-cover class accuracy. *Remote Sens. Environ.* **2003**, *84*, 342–349.
71. Latifovic, R.; Olthof, I. Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data. *Remote Sens. Environ.* **2004**, *90*, 153–165.
72. Moody, A.; Woodcock, C.E. The influence of scale and the spatial characteristics of landscapes on land-cover mapping using remote sensing. *Landsc. Ecol.* **1995**, *10*, 363–379.

73. Roumenina, E.; Atzberger, C.; Vassilev, V.; Dimitrov, P.; Kamenova, I.; Banov, M.; Filchev, L.; Jelev, G. Single- and multi-date crop identification using PROBA-V 100 and 300 m S1 products on Zlatia test site, Bulgaria. *Remote Sens.* **2015**, *7*, 13843–13862.
74. Sterckx, S.; Benhadj, I.; Duhoux, G.; Livens, S.; Dierckx, W.; Goor, E.; Adriaensen, S.; Heyns, W.; Van Hoof, K.; Strackx, G.; *et al.* The PROBA-V mission: Image processing and calibration. *Int. J. Remote Sens.* **2014**, *35*, 2565–2588.
75. Erwin, W.; Wouter, D.; Jan, D.; Else, S. *PROBA-V Products User Manual v1.1*; VITO: Mol, Belgium, 2014.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).