Mathematical analysis of Sentinel-2 spectral signal evolution for mapping agriculture area in Senegal: case of millet, maize and peanuts

Gayane Faye (1,∗), Mamadou Mbaye (1,2), Modou Mbaye (1) and Abdou Kâ Diongue (2)

1 Laboratoire de Télédétection Appliquée-LTA (Applied Remote Sensing Laboratory) /Université Cheikh Anta DIOP de Dakar, Sénégal.
2 Laboratoire d’Etudes et de Recherche en Statistique et Développement, UFR de Sciences Appliquées et de Technologie, Département de Mathématiques Appliquées, Université Gaston Berger de Saint-Louis, Sénégal

Received on June 10, 2020. Accepted on December 01, 2020

Copyright © 2020. The African Journal of Applied Statistics (AJAS) and the Statistics and Probability African Society (SPAS). All rights reserved

Abstract. Agricultural monitoring has become an absolute necessity in the Sahel countries, especially with climate change which constitutes a real threat for this sector. The aim of this work is to develop a methodology for identifying crops and mapping agricultural areas using Sentinel-2 data from the Copernicus program. The purpose of this work consisted in discriminating the crops of millet, maize and peanuts. This is to analyse the scientific and technical obstacles related to this problem. For this, we have made a mathematical analysis of optical satellite images. High temporal and spatial resolution images (10m to 60m) of Sentinel 2 sensors were used in this work. This unique set of data coupled with field data, has permitted to carry out a diagnosis of land cover and cultivated land surfaces, and evaluating the contribution of this type of data for crop forecast.

Key words: remote sensing; sentinel-2, agricultural mapping; Senegal.
AMS 2020 Mathematics Subject Classification Objects : 97K80

∗ Corresponding author: Gayane Faye (gayane.faye@gmail.com)
Mamadou Mbaye : mbaye.mamadou@outlook.com
Modou Mbaye: my.gandhy@gmail.com
Abdou Kâ Diongue: abdou.diongue@ugb.edu.sn
Résumé (French Abstract) Le suivi agricole est devenu une nécessité absolue dans les pays du Sahel surtout avec les changements climatiques qui constituent une véritable menace pour ce secteur. Ce travail a pour but de développer une méthodologie d'identification des cultures et cartographie superficies agricoles à partir des données Sentinel-2 du programme Copernicus. L'objectif de ce travail consistait à discriminer les cultures de mil, maïs et arachide. Il s'agit d'analyser les verrous scientifiques et techniques liés à cette problématique. Pour cela, nous avons utilisé un modèle mathématique couplé à l'imagerie satellitaire optique. De nombreuses images à haute résolution spatiale (10m à 60m) et temporelle provenant des capteurs Sentinel 2A ont été utilisées durant ce travail. Ce jeu de données unique couplé avec les données de terrain a permis d'effectuer un diagnostic des occupations du sol et des surfaces cultivées, et d'évaluer ainsi l'apport de ce type de données pour la prévision des cultures.

The authors.

Gayane Faye, Ph.D., is an assistant professor at Earth Science Institute, Cheikh Anta DIOP University.

Mamadou Mbaye, M.Sc., is preparing a Ph.D. dissertation under the supervision of the fourth authors.

Modou Mbaye, Ph.D., a researcher at : Laboratoire de Télédétectiion Appliquée, Cheikh Anta Diop University, Dakar, (SENEGAL).

Abdou Kà Diongue, Ph.D., is full professor at Applied Mathematics Department, Gaston Berger University, Saint-Louis, SENEGAL.

1. Introduction

In a context of sustainable development, good management of agricultural resources is a fundamental issue not only from an environmental but also a socio-economic point of view. This, especially since agriculture is the main asset of the economy of many countries including Senegal because it alone brings together the majority of the working population. However, its true contribution to gross domestic product (GDP) is not well controlled, in particular due to the lack of reliable and up-to-date information on crop production (areas and yields, etc.). Indeed, in our regions, the collection of agricultural statistical data is still carried out with traditional methods very often causing a delay in the provision of data but also and above all their reliability. This limits the possibilities of governments to establish a good policy, monitoring, planning and development strategy for the agricultural sector. Therefore, it is urgent to set up methods allowing a better knowledge of the areas sown for different crops.

To remedy this, remote sensing tool is essential. Indeed, satellite remote sensing appeared as a tool making it possible to make regular and repetitive observations of almost the entire planet, at various steps of space and time, and in several wavelength. Thus, the remote sensing tool makes it possible to monitor land use and land cover, in particular crops, using vegetation indices (Fensholt and Proud (2013); Rakotoniaina and Rakotomandrindra (2014); Karkauskaite et al. (2017)). In addition, agricultural areas and their characteristics are constantly changing over time, which makes satellite images a preferred source of information of the environment (Bappel (2005)).

This work intends to study the capacity of remote sensing in the assessment of agricultural areas by developing mathematical methods of discrimination of three crops (millet, maize, and peanuts) in the departments of Nioro in Senegal.

Millet occupies the most important place among cereal crops in Senegal and is cultivated all over the country (Chevalier (1929); Tostain (1998); FAOSTAT (2014)). The leaves are alternate, simple and the blade is linear to linear-lanceolate, reaching 80cm to 1.5m (Brink and Belay (2006)). The rate of leaf surface development is slow at the start of the season due to the small size of the embryonic leaves, but increases rapidly around 15 to 20 days after emergence. The maximum leaf area is reached at around 50% flowering, a period in which the majority of tillers have enlarged all their leaves (Maiti and Bidinger (1981)).

Maize is a cereal whose tillering are generally low. It presents a wide morphological diversity according to the strains (Hamon (1999); Mémento de l’agronome (2002)). The whole duration of cycle varies from 75 to 150 days depending on the strain (Durovray (1976); Marchand et al. (1997)).

Peanuts are an oilseed cultivated in Senegal for more than a century (Freud et al. (1997)), and constitute the drivers of Senegal’s economic development. It has a C3 metabolism from 30 to 70 centimeters high, erect or creeping depending on the variety (Gillier and Silvestre (1969); Van der Vossen et al. (2007)).

In Senegal, there is little work on the estimation of agricultural areas by remote sensing. This is mainly due to the size of the plots crops which most often is less than 2 hectares. As a result, the satellite resolutions were not very suitable for mapping on this scale. In addition, there is the strong resemblance of certain crops such as millet and maize and also to the mixture of crops. However, some attempts have been made. Therefore, the most prominent was the AGRICAB project funded by the European Union through the FP7. The appearance of the new temporal series of Sentinel-2 satellite since 2015 is a major asset for agricultural monitoring. this data has the advantage of being free. The article is structured in four parts. The first section describes the study area. In the following sections, the satellite data used and the mathematical methods of processing are presented (part II), the results obtained are detailed in the third part and finally an analysis and discussion of the results (part IV).

2. Study area

The study perimeter is 20 kilometers square located in the department of Nioro du Rip in central western of Senegal (Fig. 1, page 1012). Its climate is that of Sudano-Saharan type with two distinct seasons: a rainy season which lasts approximately four months (mid-June to mid-October) with peak temperature of 40°C and a dry season of approximately eight months (mid-October to mid-June) with temperature varying between 20 to 30°C. The area is characterized by abundant rainfall, thus leading strong hydric erosion. This is illustrated through the numerous ravines which constitute a real brake on development. This situation is accentuated by a weakness of forest resources for the benefit of agriculture.

![Fig. 1. Study area](image)

The economy of the zone is essentially based on agriculture that constitutes on the one hand, the main source of income of the populations and on the other hand, the great activity consuming of manpower. However, agriculture is essentially rained, making it random with rainfall variability. It is shared between cereal crops (millet and maize) and cash crops (peanuts).

3. Material and methodology

3.1. The data

(a) Satellite imageries.
The Sentinel-2A sensor data used in this study is composed of 13 spectral bands with a spatial resolution of 10 m (4 visible and near infrared bands), 20 m (6 near and mid infrared bands), and 60 m (3 atmospheric correction bands), with a 10-day temporal repeatability (5 days with the arrival of Sentinel-2B since August 2017) facilitates the monitoring of a new range of reduced-size ground objects and with very strong temporal dynamics. This level 1C data is available on the ESA (European Space Agency) download platform, https://sentinel.esa.int/web/sentinel. In order to have good quality data so as not to bias the results of the study, we arbitrarily set the acceptable cloud cover rate at maximum 10%. In total, during the entire rainy season (June 15 to November 15), only three images meet this threshold (Table 1, page 1013).

<table>
<thead>
<tr>
<th>Satellite</th>
<th>Tuile</th>
<th>Images</th>
<th>Cloud coverage rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentinel-2A</td>
<td>T28PDA</td>
<td>08 August</td>
<td>0.22%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18 August</td>
<td>7.21%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>07 October</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Table 1. list of images of the study area taken between June 15 and November 15 and having cloud cover less than 10%.

(b) In situ measurements.

Field data are the contours of land digitized from GPS. The latter are 168 fields (Table 2, page 1013) distributed over the entire study area in order to have a good spatial representativeness of the sample (Fig 2, page 1014). These data are used as a basis for training and validation in the development of the discrimination method. To avoid edge effects and contiguous pixels, the contours of the land plots have been adjusted with google earth. This made it possible to correct the errors on the boundaries and to circumscribe them within the plots.

<table>
<thead>
<tr>
<th>Crops</th>
<th>Number of plot</th>
<th>Area (ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peanut</td>
<td>59</td>
<td>115</td>
</tr>
<tr>
<td>Maize</td>
<td>55</td>
<td>87</td>
</tr>
<tr>
<td>Millet</td>
<td>54</td>
<td>99</td>
</tr>
</tbody>
</table>

Table 2. summary of in situ data

3.2. Methodology

(a) Extraction of reflectances from satellite images: In the first time, level 1C images which are expressed in digital value (DN: digital number) and coded in 16 bits (0 to 65535) are corrected using the Sen2Cor module of SNAP software.
This eliminates atmospheric effects and converts numerical values to reflectance (Level-2A). After correction, all the bands are resampled at 10 meters and the NDVI and NDWI spectral indices are calculated. Finally, the pixel values of the control plots are extracted except for bands 1, 9 and 10 which do not provide information on the ground parameters.

Two types of variables are used here:
- **Explain variables**: "Peanut", "Maize", "Millet"

(b) **Analysis of variance (ANOVA)**: Its purpose is to identify the most suitable bands to discriminate the three crops. First, an analysis of the dispersion of the values of each culture is made according to the twelve variables and for each of the three images (August 8, August 18 and October 07). This in order to better appreciate the influence of extreme values which can induce a bias on the mean or the median.

In a second step, an analysis of the variance is made by applying the Turkey test. The purpose is to see for each image, the capacity of each of the twelve variables (ten bands and two indexes) to differentiate the crops taken two by two. A check is made with the confidence interval around the difference. If this contains the value zero (maximum positive and minimum negative), the data (band or index) is not be able to discriminate the two crops, thus it is excluded for the rest of the work.

(c) **Principal Component Analysis (PCA)**: It is carried out for each image from the variables selected by ANOVA. The goal is to look at the combinations allowing to differentiate crops (main axes) and to further reduce the number of variables.

(d) **Classification and Regression Trees (CART)**: The aim is to predict the crops based on the variables (spectral bands and indexes). For this, it partitions recursively the space in a binary way to finally optimize the prediction of the type of crop at the end of each chain (Breiman et al. (1984)). For each image, two CART classi-
Mathematical analysis of Sentinel-2 spectral signal evolution for mapping agriculture area in Senegal: case of millet, maize and peanuts.

4. Results

(a) Analysis of the data dispersion: From 3 images acquired by the Sentinel-2 satellite on August 8, August 18 and October 7, we studied the recording of the reflectance of each crop (millet, maize and peanut) according to the 10 spectral bands and the indices (Fig 3, page 1015).

![Image of data dispersion](image_url)

**Fig. 3.** mustache box of the values of the three crops according to all the twelve variables (ten spectral bands and two indices) and for the three images (August 08, August 18 and October 07)

(b) Analyse of the variance (ANOVA): The comparison of the values of the three crops taken two by two made it possible to identify and exclude the variables (10 bands and 2 indices) that are not capable to differentiating the crops (Fig 4, page 1017). We will analyse the test results.

**Image of 08/08/2016:** the reflectances of Millet and peanuts are identical in bands B2, B6, B7, B8 and B8A as well as those of Millet and maize in band B3 for this date. The bands B4, B5, B11, B12 and the indices NDVI and NDWI are used for the rest of the analysis of this image (Table 3, page 1016).
<table>
<thead>
<tr>
<th>Date</th>
<th>Excluded Bands</th>
<th>Millet-Peanut</th>
<th>Maize-Peanut</th>
<th>Millet-Maize</th>
</tr>
</thead>
<tbody>
<tr>
<td>08 August 2016</td>
<td>B6, B7, B8 and B8A</td>
<td>B4, B5, B11, B12, NDVI and NDWI</td>
<td>any</td>
<td>B3</td>
</tr>
<tr>
<td>18 August 2016</td>
<td>B3, B4, B5, B6, B7, B8 and B8A</td>
<td>B2, B3, B4, B5, B7, B8, B12</td>
<td>B4, B5, B6</td>
<td>B11, NDVI and NDWI</td>
</tr>
<tr>
<td>07 October 2016</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
</tr>
<tr>
<td>Retained Bands</td>
<td>B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12, NDVI and NDWI</td>
<td>B11, NDVI and NDWI</td>
<td>any</td>
<td>B3</td>
</tr>
</tbody>
</table>

Table 3. summarizes the results of the analysis of variance (ANOVA) which highlights the similarity of the signals of the different crops for the different bands.
Image of 18/08/2016: at this date, maize and peanuts have the same spectral behavior in bands B3, B4, B5, B6, B7, B8, B8A, and B12. Millet and peanut are identical in bands B3, B4, B5, B6, B7, B8 and B8A whereas millet and maize have the same reflectance in bands B4, B5 and B6. As a result, for this image the band B11 and the NDVI and the NDWI indices are kept for the rest of the study (Table 3, page 1016).

Image of 07/10/2016: ANOVA shows that all the bands have a discriminating power between the different crops taken two by two. For this image, all the bands and indices are kept for the rest of the study (Table 3, page 1016).

4.1. Principal Component Analysis (PCA)

The bands selected by ANOVA as having the capacity to discriminate between different cultures made it possible to perform PCA (Fig 5, page 1018, Fig 6, page 1019 and Fig 7, page 1019).

Image of August 8th, 2016: A strong dispersion of millet is noted around the two main axes, which explain almost 95% of the inertia (Axis1 76.2% and Axis2 18.18%). Maize is very homogeneous but mingled with the two other crops. On axis1, the spectral bands (B4, B5, B11 and B12) contribute positively while the
spectral indices have negative contributions. The contribution of the different bands is fairly balanced (19% for B4, 16% for B5, 20% for B11, 20% for B12 and 16% for NDVI) except for NDWI which contributes only 8%. For axis2, only the B12 has negative contribution, the other variables contribute positively. On this axis, NDWI contributes more than 54%, followed by band B5 (20%) and NDVI (15%). These results show that overall, all the variables (bands and spectral indices) contribute in a fairly balanced way (between 16 and 20%).

**Image of August 18**, 2016: Three axes explain 100% of the inertia (62% by axis1, 21% by axis2 and 17% by axis3); fig 6, page 1019. This situation is explained by the fact that only three variables were chosen by ANOVA. This means that PCA is not necessary because ANOVA already reduced the number of variables.

**Image of October 7**, 2016: the first two axes explain more than 95% of total inertia (67% by axis1, 29% by axis2). These two axes discriminate very well between peanuts and millet and to a lesser extent, maize, which present some confusion. On the first axis, the indices have negative contributions and bands B2, B3, B4, B5, B11 and B12 contribute positively. The NDVI and the NDWI have the most important contributions 12% each followed by the band B11 and B4 with each 10%. On the second axis, the contribution of the spectral indices is almost zero (0.1% for the NDVI and 1.1% for the NDWI). Bands B6, B7, B8 and B8A contribute more than 60% to the inertia of this axis.

![Fig. 5. Results of the PCA applied on the image of 08 August](image_url)

**4.2. CART classification**

For each image two classifications are made: a first made from the variables retained by the ANOVA (method C1) and a second which uses the first three axes of the PCA which explain at least 95% of the inertia (method C2).
Fig. 6. Results of the PCA applied on the image of 18 August

(a) Image of August 8th, 2016: The overall error of C1 classification is 47.1% compared to 45.5% for C2, which constitutes a near-zero improvement of 1.6 (Fig 8, page 1021). This means that the reduction of variables by PCA does not improve classification. In detail, we also note that for all crops, the use of PCA very slightly improves the accuracy of the classification with a maximum of +3.8% noted for maize. On the two methods, peanut shows the greatest error with more than 55%. For this date, more than 33% of the peanut fields are classified as millet, while only less than 17% of the millet fields are classified as peanut. The same situation is observed for maize, where more than 33% of the fields are seen as millet when about 21% of the millet fields are classified as maize. In short, the
model which is based exclusively on NDWI and band 11 tends to better classify millet.

(b) Image of August 18th, 2016: with less than 39% error, the image of 18 of August gives an accuracy of almost 7% compared to that of august 8th (Fig 9, page 1022). It is above all the detection of peanuts which is improved by passing from 44 to 56% accuracy, which is worth to a jump of more than + 12%. Likewise, the accuracy of millet is improved by almost 10%. This clear-cut improvement may result from the contribution of NDVI which, unlike the image of august 8th, is taken into account by the CART tree.

(c) Image of October 7th, 2016: unlike August, the use of PCA (method C2) on the image of October 07th degrades the overall accuracy of the classification, whose error goes from 17 to 21% (Fig 10, page 1023). This difference is especially visible for millet whose classification error goes from 19.5 (method C1) to 33% (method C2). For this image, the model manages to identify the peanut with 91.4% accuracy for both methods. This is the period when it is easier to differentiate crops. However, it is not necessary to reduce the variables because this affects the quality of the classification, especially that of millet.
Fig. 8. Decision tree of August 08th, 2016 with the two methods (C1 and C2).
Fig. 9. decision tree of August 18th, 2016 with the two methods (C1).

<table>
<thead>
<tr>
<th></th>
<th>Peanut</th>
<th>Maize</th>
<th>Millet</th>
<th>Error %</th>
</tr>
</thead>
<tbody>
<tr>
<td>With variables selected by ANOVA (C1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peanut</td>
<td>56.7</td>
<td>27.3</td>
<td>16.0</td>
<td>43.3</td>
</tr>
<tr>
<td>Maize</td>
<td>18.2</td>
<td>53.1</td>
<td>28.7</td>
<td>46.9</td>
</tr>
<tr>
<td>Millet</td>
<td>17.1</td>
<td>11.5</td>
<td>71.4</td>
<td>28.6</td>
</tr>
<tr>
<td><strong>Global error</strong></td>
<td><strong>38.8</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 10. decision tree of October 7th, 2016 with the two methods (C1 and C2).

5. Discussions

In August, most of the spectral bands of the sentinel-2 sensor were unable to differentiate the three crops, mostly peanut and millet, and to a lesser extent with peanut and maize. During this month a large part (about 1/3) of the peanut and maize fields are classified as millet. This situation is explained by the fact that during this period the rate of plant cover is not yet significant, which means that the contribution of the soil is much greater, thus inducing confusion. On the other hand, during this period wild grasses are often important in the fields and the latter are very similar to the millet at the start of the rainy season. So, a poorly weeded peanut or maize field can easily be mistaken for a millet field. The use of PCA with CART almost does not improve the results obtained without PCA. So, it is not necessary to reduce the variables because the CART selects the most significant variables, that is to say, the most discriminating among different crops.

In October the results are clearly better with an overall accuracy of over 82%. This period coincides with the end of the cereal crops harvest (millet and maize) whereas the peanut still has fairly marked chlorophyll activity. It is for this reason that the classification of peanuts is by far the best with only 8% error. For cereals, the precision hovers around 80% with still a significant part (1/5) of the maize fields classified millet against 10% of millet fields identified as maize. This situation is
justified, indeed, after the harvest, the stalks of millet and maize remain and dry slowly. So, a field of maize and millet can easily be mixed up. The PCA has no use for the data of the month of October as it alters the accuracy of the classification, in particular that of millet, which drops from 80.5 to 66.9% in precision, which is worth to a drop. 13%.

CONCLUSION

This study whose aims is to analyse the capabilities of Sentinel-2 optical data to map crops at plot level in the Senegalese context marked by small fields often less than 2 hectares. The analysis of variance has shown that, unlike October, when all the spectral bands and the NDVI and NDWI spectral indices are sensitive to the differences among crops. In August, it is not very easy to differentiate the three crops studied here. During this period, most spectral bands fail to discriminate the crops. Ultimately, to properly map crops at the plot level using Sentinel-2 optical data, it is better to use the end-of-season data in October corresponding to the harvest period. This, especially since during the whole rainy season the availability of optical images is quite reduced due to the high cloud rate. On the other hand, using principal component analysis to reduce the variables is ineffective. The ANOVA followed by the classification by CART is enough to exploit all the discriminating potential of images.

References


https://sentinel.esa.int/web/sentinel

https://step.esa.int/main/toolboxes/snap