Identifying cropped areas in small growers agricultural regions using data mining for food security

Elodie Vintrou, Valentine Lebourgeois, Agnès Bégué, Dino Ienco, Maguelonne Teisseire, Stéphane Dupuy, Fidiniaina Ramahandry Andriandrahona

IDENTIFYING CROPPED AREAS IN SMALL GROWERS AGRICULTURAL REGIONS USING DATA MINING FOR FOOD SECURITY

Vintrou Elodie(1), Lebourgeois Valentine(2), Bégué Agnès(2), Ienco Dino(3), Teisseire Maguelonne(3), Dupuy Stéphane(2), Ramahandry Andriandrahona Fidiniaina (4)

(1)CIRAD UR AIDA, Station Ligne Paradis, 7 chemin de l'Irat, 97410 Saint Pierre, La Réunion, France
Email : elodie.vintrou@cirad.fr
(2) CIRAD UMR TETIS, Maison de la Télédétection, 500 rue J.F. Breton, Montpellier, France
Email : valentine.lebourgeois@cirad.fr; agnes.begue@cirad.fr
(3) IRSTEA UMR TETIS, Maison de la Télédétection, 500 rue J.F. Breton, Montpellier, France
Email : teisseire@irstea.fr ; ienco@irstea.fr
(4) FOFIFA, Station Régionale de Recherche FOFIFA Tsivatrinikamo ANTSIRABE 110, Madagascar

ABSTRACT

The present study aimed at testing the potential of the future mission SENTINEL-2 (European Copernicus program) to map croplands in a region of Madagascar characterized by small size fields and frequent cloud covering. Two approaches were tested and compared : i) a classical remote sensing method (RS) using image object-based analysis, expert rules and supervised classification, and ii) a data mining (DM) approach consisting of the extraction of frequent patterns from the database and the use of these patterns in different algorithms (Naive Bayes, Random Forest, Decision Tree and Support Vector Machine) to build classification rules. Both methods used SPOT images and a ground data set of 324 GPS waypoints collected during the 2012-2013 cropping season.

The remote sensing and data mining approaches showed equivalent overall accuracies (82% vs 84% for RS and DM methods respectively). However, the DM approach showed its ability to handle a large volume of data and to do so in a timely manner. This approach has also the advantage to extract all the information at its disposal, even temporal behaviors, unlike the object-based RS approach which requires significant participation of the expert.

Data mining tools are thus recommended for their considerable potential for the classification without a priori of remotely sensed data, mixing multisource information and consequent time series, especially for the upcoming Sentinel-2 images that are expected to generate a large volume of data to store and process.

1. INTRODUCTION

Food insecurity is particularly worrying in Africa, where nearly one country in four suffers from hunger [1,2]. In front of the growing number of natural disasters, the increasing population, the emergence of dedicated crops to production of biofuels, land grabbing by foreign investors etc., there is a need for systematic and accurate monitoring of agricultural systems and their adaptation to changing environment, to assess the impacts on food security.

Early warning systems are designed to provide reliable information on preventive potential risk of a food crisis: estimating production is capital to compensate for the lack of food per food aid or imports in developing countries. For that purpose, data on cultivated surfaces and yields are an essential prerequisite for a good agricultural production forecast [3]. So far, low and moderate resolution satellite remote-sensing images have been extensively used for crop mapping and monitoring [4-6]. Their high temporal frequency and their extended geographical coverage associated with free or low costs per area unit makes them a particularly appropriate information source at both national and regional scales. However, with these data, the estimation of cultivated surfaces and the discrimination between different crops is still challenging in countries with fragmented farmland, small size fields or with specific weather conditions resulting in high regional variability in terms of agricultural systems and practices [7,8].

The upcoming availability of SENTINEL-2 data opens up new prospects for research, including methodological developments related to agriculture monitoring. By mid-2015, this future Earth Observation System will propose images of i) higher resolution (10–

60 m depending on spectral bands) and frequency (10 days in 2015 and 5 days in 2016) allowing a fine agricultural monitoring adapted to the study of crops in areas where agriculture is fragmented, ii) with significant swath (290 km) appropriate for regional and global studies, and iii) with a large number of spectral bands (13) allowing the characterization of patterns of land use, and quantitative estimation of biophysical variables related to the crop conditions. The pair of Sentinel-2 satellites will thus soon represent the best compromise between spatial and temporal resolution and will replace moderate-resolution images such as MODIS for agricultural monitoring.

To deal with such high spatial resolution images and complex landscape, object-based image analysis is an interesting approach. An overview of the development of object based methods can be found in [9]. These methods are used in the expectation that it will divide the image into i) relatively homogeneous, and ii) semantically significant groups of pixels. Whereas in classical pixel approaches spatial concepts are not used, in object-based methods a pixel is not studied alone, but together with its neighborhood, and this adds spatial information to the objects [10]. Dealing with agriculture, the object-based image analysis (OBIA) can help to delineate field boundaries, and thus to reach classification results at field scale. It is thus particularly suitable for high spatial resolution images. However, OBIA requires an expert intervention through supervised methods, which can be difficult when dealing with important volume of data.

As the number of sensors of higher spatial and temporal resolutions and possibilities of data sharing are increasing, a generation of new tools is being developed, which is able to handle large volumes of data but also to automatically extract knowledge from databases of multiple sources. Since recently, these data mining tools are beginning to be used in the field of remote sensing [11,12]. An example is the prediction of land use from time series of remote sensing images. In this case, data mining patterns' extraction and classification algorithms can be employed to perfectly solve this task and they are able to scale up over big dataset [13].

In this context, the present study aimed at testing the potential of the future mission SENTINEL-2 to map croplands in a region of Madagascar characterized by small size fields, a large heterogeneity of the cropping practices, and frequent cloud covering. The overall objective of this proposal was to provide new products from the future satellite mission, based on existing (SPOT satellite time series) or recent (PLEIADES) missions to support early warning systems for food security in fragmented agriculture.

For this, we developed two different approaches to map a cropland mask in fragmented landscapes: i) a classical remote sensing method using image object-based analysis, expert rules and supervised classification, and ii) an original method based on data mining techniques. This paper deals mainly with the data mining approach as an alternative to conventional methods for defining a learning mechanism based on multi-source data. The object-based Remote-Sensing (RS) and Data Mining (DM) classification results are compared using error matrices based on ground sample points. The same methodology is currently being developed for cropping systems mapping.

2. STUDY ZONE AND MATERIAL

2.1 Study area

Madagascar is an island country in the Indian Ocean, off the coast of Southeast Africa. Our 60*60 km study zone is located near Antsirabe, the capital of the Vakinankaratra region, in the central highlands. This region has the second highest population density of the country and is characterized by terraced, rice-growing valleys lying between grassy hills. Despite its small size (60*60 km), this study area is characterized by heterogeneous landscapes. The irrigation systems are well developed, and use all available water, which flows through narrow canals for considerable distances. Only the areas that cannot be irrigated are planted in dryland crops. Narrow terraces ascending the sides of steep valleys are mainly settled/planted with rainfed maize, cassava, and beans. The main crops such as maize and rice are sown at the beginning of the rainy season (between October and December) and harvested at the end (from March to May). Some of the plots cover no more than a few square meters. The mean size of an agricultural field is very small (about 0.03 ha) but contiguous fields with the same crop type can often result in larger crop patches.

2.2 Ground data

Fields surveys were conducted in Madagascar during the growing peak (end of February) of the 2012-2013 cropping season in order to characterize the main cropping systems. A total of 324 GPS waypoints (247 cropped and 90 non-cropped) were registered in the study area, chosen according to their accessibility and to be as well representative of the existing cropping systems as possible. The data gathered during the field survey concerned farmers' practices (type of crop, use of fertilizers and irrigation). GPS waypoints were also registered on different types of natural vegetation to obtain data on the non-crop class.

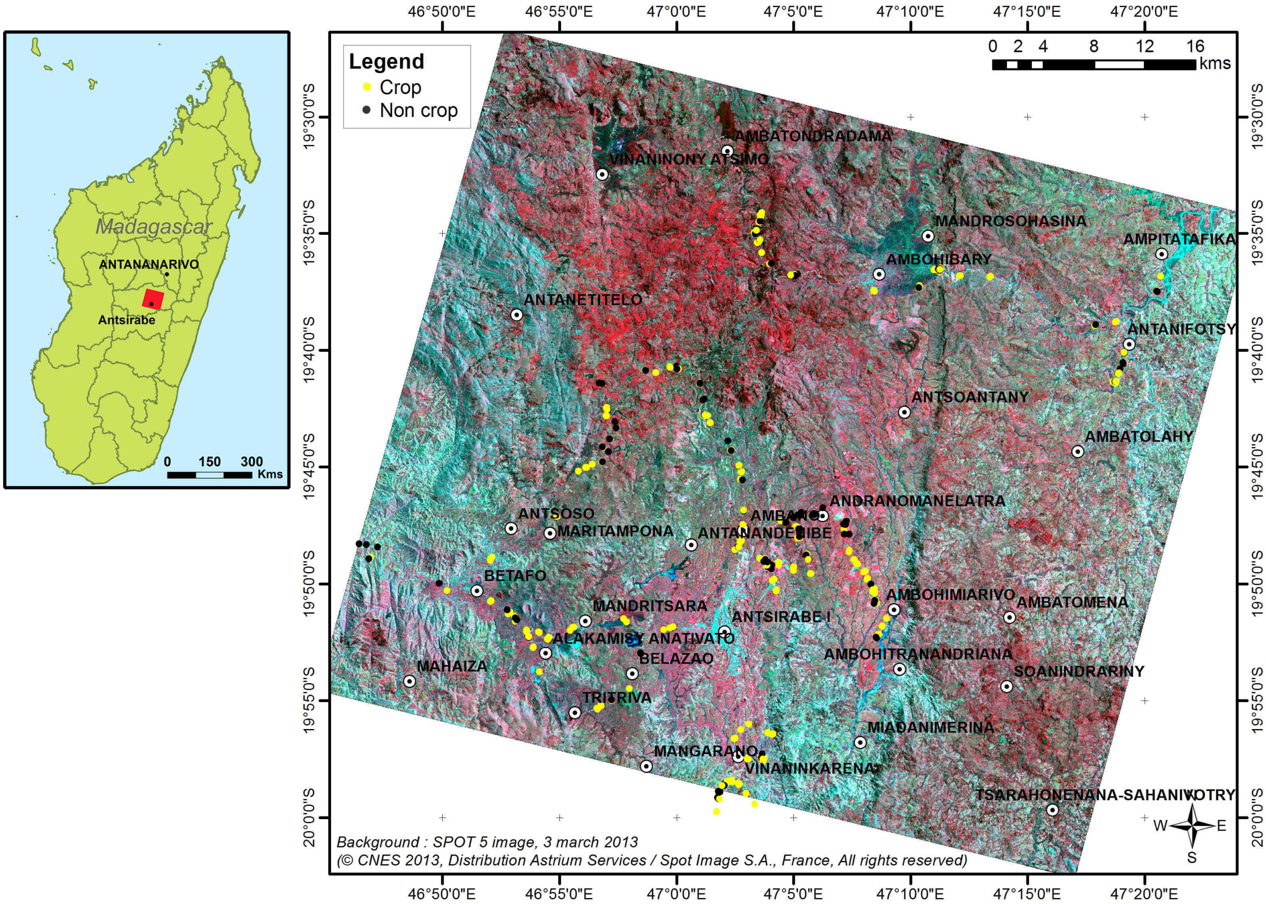

Figure 1. Map of the study area around Antsirabe showing the geographic location of the 324 fields sampled on the 2013/03/03 SPOT5 image.

2.3 Satellite Images

During the growing season extending from October 2012 to May 2013, combined acquisitions of SPOT4 and SPOT5 acquired from SEAS-OI satellite receiving station in Reunion Island or from CNES (Centre National d'Etudes Spatiales) SPOT4Take5 experiment were conducted and allowed us to obtain a time series of 25 decametric images with an average time repeatability of 12 days. These images were corrected geometrically and radiometrically (TOA reflectance). A digital elevation model (SPOT DEM at 20 m spatial resolution) was also acquired in order to extract the location of agricultural fields in the toposequence, thus giving information on the irrigated or rainfed regime of each concerned field. Finally, very high resolution PLEIADES images were acquired at the maximum of the growing season on our study zone to assist in the identification of the fields' outlines of the ground database.

3. METHODS

3.1 The object-based Remote-sensing (RS) method

This classical remote-sensing method consisted in classifying SPOT images in three steps: : i) a segmentation of SPOT images in objects, ii) a mask of urban/artificial areas and a classification of the toposequences based on expert rules, and iii) in each toposequence, a crop-non crop supervised classification of the objects based on ground data.

Segmentation

The study area was first segmented so that objects represent plots or groups of plots in the cultivated area, using SPOT images. Considering that field characteristics include a temporal pattern (crop seasonality) and a specific structure (human print), we assumed that the spatio-temporal variability of NDVI and textural indices of SPOT images can be used to segment and classify the study area. We used eCognition software, and for processing time reasons, we chose only two contrasted SPOT images free of clouds in the available time series (one during the vegetation peak on the 3rd of March, one during the dry season on the 21th of May) to derive these indicators (mean NDVI, mean variance and Euclidian textures). We tested varying combinations of segmentation parameters (shape and compactness) for optimized boundary separation and obtained about 1 300 000 objects. The average object size is about 0.24 ha, which can be considered for the cultivated domain as a “group of fields”.

Mask of urban and artificial areas and classification of toposequence

Urban and artificial areas were first isolated thanks to expert rules based on thresholds of brightness, euclidian texture, and NDVI.

In Madagascar, the location of the fields in the toposequence is an important driver of the cropping system as it has a direct impact on the soil quality and water conditions. Basically, irrigated rice is grown in shallows, whereas rainfed rice or maize are found on hills or uplands. To assign a class of toposequence to each field or goup of fields thus helps in reducing the variability of spectral responses of cropping systems used in the following supervised classification step.

The toposequence was classified in 4 classes (basins or shallows, lowlands and lower parts of the hills, hills, and uplands) thanks to the use of information extracted from the SPOT DEM processing (slopes, hydrological network) and thresholds (e.g. uplands objects have a slope lower than 3 degrees) or neighborhood rules (e.g. shallows objects intersect with water system, objects belonging to lower parts of the hills touch objects classified as shallows…).

Supervised classification

Inside each class of toposequence, the objects were classified in “crop” or “non crop” using a supervised maximum likelihood classification as proposed in eCognition software.

The object-based classification works in the same way as a pixel-based classification with the difference that we do not classify each pixel but combine all pixels of each object and classify them together [14]. In this study, we used the mean value of all pixels of an object in the four SPOT spectral bands and the mean NDVI value of the two selected images, and textural variance calculated on the SPOT NDVI image of March image (maximum of the season). The maximum likelihood classification was conducted thanks to 80% of the ground database and the afore-mentioned attributes associated.

Accuracy calculation

The accuracy of the resulting map, referred hereafter as RS classification, was assessed thanks to a five-fold cross-validation of the classification. A random sample of one fifth of the training set was used for validation purposes. Thereby, for each fold, 20% of the total GPS waypoints were left out, and the supervised classification was recalculated using a new training set consisting of the remaining 80% of the database. The overall accuracies for each fold were recorded and averaged to obtain the overall accuracy of the RS method.

3.2 Data mining (DM) method

A classification model was developed to discriminate between cropped and non cropped area using data mining process. Our proposal involved three main steps: i) satellite-derived metrics were calculated for the 324 plots corresponding to the ground samples, ii) the PrefixSpan data mining algorithm found the frequent sequential patterns of cropped and non-cropped plots, iii) these frequent patterns were used by data mining classification algorithms (Naive Bayes, Random Forest, Decision Tree and Support Vector Machine) and thanks to its attributes, each plot was affected to “crop” or “non-crop” class.

These three steps are detailed further.

Attributes extraction

Crops spatial and temporal behaviors can be captured through a set of attributes such as vegetation indices, intra-plot organization and plot layout in the landscape, and vegetation seasonality. These attributes are known to be accessible using time series of multispectral images. Several metrics were thus calculated and associated to each of the 324 plots of the database, whose outlines where digitalized thanks to PLEIADES 50 cm spatial resolution imagery:

- Static information, such as localization in the toposequence, plot size and distance to the river but also SPOT Haralick textural indices describing the plot “organization” and its place in the landscape and calculated at two acquisition dates (March and May).
- Temporal information from the entire SPOT time series (reflectance in the four SPOT spectral bands, mean and max NDVI per plot for the 25 images). A linear interpolation of these temporal variables was conducted on the cloudy values.

Extraction of frequent patterns

To establish the link between the crop or non-crop land cover and these various indicators, we used the PrefixSpan algorithm [15] to find discriminating sequential patterns of the cropped or non cropped plots using 80% of the dataset. This algorithm extracts all frequent sequential patterns that have a “support” greater than a given threshold. A support of a sequential pattern is the number of objects in which the considered pattern appears.

Classification process

The extracted frequent patterns are used to establish "classifiers" defining if a plot is more likely to be cropped or not according to its static and temporal attributes. For this point, several machine learning algorithms are available in the WEKA tool; more precisely, the following classification techniques were applied on the data set: Naive Bayes, Random Forest, Decision Tree and Support Vector Machine. We choose these data mining models as they span over different assumptions and they well cover different families of classification algorithms:

- Naive Bayes [13] is a probabilistic-based algorithm that employs the "Bayes principle" in order to make its prediction. It assigns a new object to the class that maximizes its likelihood.
- Random Forest [16] is an ensemble learning schema that builds a multitude of random decision trees. It makes its prediction on a new example as the mode of the classes predicted by each individual tree.
- Decision tree [13] model the training data using a tree-shape structure. Internal nodes of the tree represent test over description variables while the leaves represent class assignments. A new instance is classified following a root-leaf path induced considering the test in the inner nodes.
- Support Vector Machine (SVM) is one of the most effective and recent classification techniques proposed in the machine learning community [17]. This approach searches for a set of support vectors inducing an hyperplane (in the space in which data are represented) able to well separate instances of different classes and obtaining, at the same time, as much generalization as possible. The decision on a test instance is done considering its distance with regards to the support vectors.

Accuracy calculation

These different classification processes were then applied to the 20% of the waypoints left, using a five-fold cross-validation as for the Remote-Sensing method.

3.3 RS and DM approaches comparison

The RS and DM classifications obtained after the 5 fold cross validation were evaluated through the error matrices based on the ground data set (324 GPS waypoints). The classification accuracy criteria were: i) the fraction of correctly classified pixels, ii) the commission and omission errors. In addition, for DM classifications, the four DM algorithms were compared through their overall accuracies.

4. RESULTS

The four data-mining classification algorithms gave good accuracies (from 79 to 84%) but the most accurate classification was derived from the Support Vector Machine (SVM) approach (Table 1).

Algorithm	Overall accuracy
NaiveBayes	78%
Random Forest	82%
Decision tree	79%
SVM	**84%**

Table 1. Overall accuracies for the four classification algorithms used in the DM methodology.

The remote sensing and data mining (SVM) approaches showed equivalent overall accuracies (82% vs. 84% for RS and DM methods respectively). Both methods provided stable results for the crop class, that is to say a commission of pixels equivalent to the omission (Table 2). Omission and commission errors were for both methods limited for the crop class (between 7 and 16%), whereas they were quite important for the non-crop class (from 20 to 38%).

		RS_method	DM_method
CROP	Omission error	11%	7%
	Producer accuracy	89%	93%
	Commission error	14%	16%
	User accuracy	86%	84%
NON CROP	Omission error	34%	38%
	Producer accuracy	66%	62%
	Commission error	29%	20%
	User accuracy	71%	80%
Overall accuracy		**82%**	**84%**

Table 2. Accuracy assessment of the Remote Sensing (RS) and Data Mining (DM) classification approaches based on the 324 GPS waypoints.

Figure 2 shows the spatial distribution of cropland in our study area (in this example with the RS approach). As expected, the majority of shallows are cultivated, and rainfed crops are colonizing gradually the hills

.

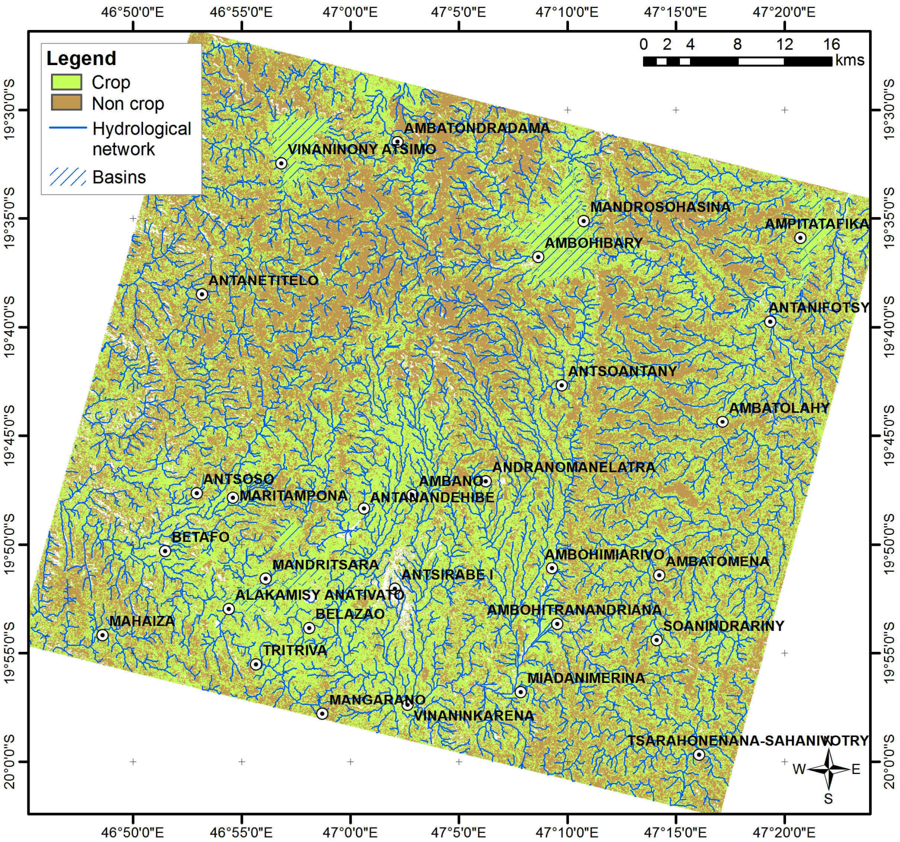

Figure 2. Crop classification of the study area around Antsirabe using the object-based remote-sensing method.

5. DISCUSSIONS AND PERSPECTIVES

Two approaches were developed and compared for mapping cropped areas small growers agricultural regions such as in Madagascar. The first method (Remote Sensing or RS) involved an object-based classification, and the second method was a data mining approach (DM). The RS method showed a fraction of pixels correctly classified of 82%, against 84% for the DM approach. The crop class was for both methods better classified than the non-crop class, which is certainly due to the number of crop samples higher than the non-crop ones (247 cropped vs. 90 non-cropped samples).

Misclassifications per class (crop and non-crop) were between 11 and 34% for the RS approach and between 7 and 38% for the DM approach, which can be explained by two main factors:

- cropped patches/objects too small: i) to be detected by SPOT at 10 m resolution, ii) to be separated from the surrounding natural vegetation, iii) to extract a pure signal of the sole cropped field when mean radiometric values are extracted for DM process for example. In this region of Madagascar, the mean plot size is 0.03 ha, which represents less than 2*2 SPOT pixels.
- insufficient reference data collection leading to small size training problem. According to [18], training samples should be exhaustive and made up of samples that completely describe the intra-class variability, to encompass all the possible spectral signatures in the classes (cropped and non-cropped domains in our case). However, our study zone is suffering from specific weather conditions resulting in high regional variability in terms of natural landscapes, agricultural systems and practices. Our ground database may not contain samples describing all the land-cover classes present in the investigated area.

In both methods, the results could be improved with a bigger training dataset, which would also allow the use of an independent validation dataset for more rigorous statistical results [19]. About the object-based remote-sensing techniques, using more than two SPOT images for the generation of the cropland mask would be one solution to further reduce errors in both segmentation and field-masking. But this makes the RS method even more cumbersome and time-consuming, and the use of a calculation server is recommended.

Whereas they obtained similar classification accuracies, these two methods differ widely. Thus, for fair comparison of the classifications it is important to note

some specific aspects. First, from a same data set, the RS method used less data than the DM method. The object-based RS method being "user-dependent", we chose indicators and SPOT images acquired at given interest dates that seemed essential to distinguish crop class of non-crop class, so as not to saturate the eCognition software with too many images. That is to say this method used only 2 SPOT images and 2 textural indices, whereas the DM method was based on the whole data set (25 SPOT images and their derived metrics and more than 60 textural images as an example).
The extraction of frequent patterns step is the highlight of the DM method. Unlike conventional method with classifiers such as Random Forest [7] or SVM, this pretreatment step is used to extract temporal behavior that could not be detected otherwise. The advantages of this DM approach are also its ability to handle quantitative and qualitative data but also dynamic and static data and to do so without a priori, in a timely manner [20]. In contrast, the RS method is heavy in its implementation: it mobilizes a full-time expert for the eCognition experts rules to define, and is thus very time consuming. Consequently, as confirmed by [21], the data mining procedure is generally recommended, primarily in studies that mobilize wide datasets. Furthermore, to the best of our knowledge, few authors have addressed the problem of classifying multidimensional temporal data [22,23] and sequential pattern mining of remote sensing images has only been applied at the pixel level without taking into account texture information in the mining process.
Many of the problems in mapping land cover noted in the literature relate to the methods used to extract the land cover information from the imagery. This has driven a considerable amount of research into classification methods and supervised classifications in particular [24]. Seeing that their classification accuracy is generally the same order of magnitude as that obtained with classical classifiers [11], researchers are trying to automate methods and to find methods that maximize the information extraction from the dataset without a priori. Support vector machines (SVM) and Random Forest have recently attracted the attention of the remote sensing community [7,25-30], as they have considerable potential for the classification of remotely sensed data.
The same RS and DM methods are currently being developed for cropping systems mapping. Further analysis will consist in testing the potential of the future mission SENTINEL-2 for crop monitoring and estimation of agricultural production.

6. CONCLUSION

This study showed the relevance of the use of data mining tools for crop mapping in regions with fragmented agriculture, using decametric images. This should be even clearer in the future, as the new satellites such as Sentinel-2 are expected to generate a large volume of data to store and process. This data mining techniques appear to be robust enough to be applied to a diverse variety of data sets and to be able to integrate information extracted at multiple spatial and temporal scales.

Acknowledgments

This work was supported by CNES funds (E. Vintrou's fellowship and V.Lebourgeois' TOSCA project). The authors would like to thank the SEAS-OI reception station in Reunion Island for the SPOT data access, and the JECAM network for integrating the Malagasy site in the initiative.

REFERENCES

1. FAO, W., Ifad (2012) the state of food insecurity in the world 2012: Economic growth is necessary but not sufficient to accelerate reduction of hunger and malnutrition. *FAO, Rome* **2012**.
2. FAO, Statistical yearbook 2013: World food and agriculture. *FAO (Food and Agriculture Organziation of the United Nations), Rome* **2012**.
3. Brown, M.E., *Famine early warning systems and remote sensing data*. Springer Verlag: 2008.
4. Doraiswamy, P.C.; Sinclair, T.R.; Hollinger, S.; Akhmedov, B.; Stern, A., Application of modis derived parameters for regional crop yield assessment. *Remote Sensing of Environment* **2005**, *97*, 192-202.
5. Herold, M.; Mayaux, P.; Woodcock, C.E.; Baccini, A.; Schmullius, C., Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment* **2008**, *112*, 2538-2556.
6. Vintrou, E.; Desbrosse, A.; Bégué, A.; Traoré, S.B.; Baron, C., Crop area mapping in west africa using landscape stratification of modis time series and comparison with existing global land products. *International Journal of Applied Earth Observation and Geoinformation* **2012**, *14,* 83-93.
7. Vintrou, E.; Soumaré, M.; Bernard, S.; Bégué, A.; Baron, C., Mapping fragmented agricultural systems in the sudano-sahelian environments of africa using random forest and ensemble metrics of coarse resolution modis imagery. *Photogrammetric engineering and remote sensing* **2012**, *78*, 839-848.

8. Lobell, D.B.; Asner, G.P., Cropland distributions from temporal unmixing of modis data. *Remote Sensing of Environment* **2004**, *93*, 412-422.
9. Blaschke, T., Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing* **2010**, *65*, 2-16.
10. Blaschke, T.; Strobl, J., What's wrong with pixels? Some recent developments interfacing remote sensing and gis. *GeoBIT/GIS* **2001**, *6*, 12-17.
11. Pitarch, Y.; Vintrou, E.; Badra, F.; Bégué, A.; Teisseire, M., Mining sequential patterns from modis time series for cultivated area mapping. *Advancing Geoinformation Science for a Changing World* **2011**, 45-62.
12. Julea, A.; Méger, N.; Rigotti, C.; Trouvé, E.; Jolivet, R., Efficient spatiotemporal mining of satellite image time series for agricultural monitoring. *Transactions on Machine Learning and Data Mining* **2012**, *5*, 23-44.
13. Witten, I.H.; Frank, E., *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann Series in Data Management Systems: 2005.
14. Walter, V., Object-based classification of remote sensing data for change detection. *ISPRS Journal of photogrammetry and remote sensing* **2004**, *58*, 225-238.
15. Pei, J.; Han, J.; Mortazavi-Asl, B.; Wang, J.; Pinto, H., Mining sequential patterns by pattern-growth: The prefixspan approach. *Knowledge and Data Engineering, IEEE Transactions on* **2004**, *16*, 1424-1440.
16. Breiman, L., Random forests. *Machine learning* **2001**, *45*, 5-32.
17. Cristianini, N.; Shawe-Taylor, J., *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press: 2000.
18. Muñoz-Marí, J.; Bruzzone, L.; Camps-Valls, G., A support vector domain description approach to supervised classification of remote sensing images. *Geoscience and Remote Sensing, IEEE Transactions on* **2007**, *45*, 2683-2692.
19. Foody, G.M., Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric engineering and remote sensing* **2004**, *70*, 627-634.
20. Vintrou, E.; Ienco, D.; Begue, A.; Teisseire, M., Data mining, a promising tool for large-area cropland mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.* **2013**, *6*, 2132-2138.
21. Vintrou, E.; Ienco, D.; Bégué, A.; Teisseire, M., Data mining, a promising tool for large-area cropland mapping. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing. IEEE Early Access Article on line (DOI : 10.1109/JSTARS.2013.2238507). Under publication.* **2013**.
22. Qin, Y.; Obradovic, Z., Efficient learning from massive spatial-temporal data through selective support vector propagation. *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS* **2006**, *141*, 526.
23. Tseng, V.S.; Lee, C.-H., Effective temporal data classification by integrating sequential pattern mining and probabilistic induction. *Expert Systems with Applications* **2009**, *36*, 9524-9532.
24. Foody, G.M.; Mathur, A., A relative evaluation of multiclass image classification by support vector machines. *Geoscience and Remote Sensing, IEEE Transactions on* **2004**, *42*, 1335-1343.
25. Pasolli, E.; Melgani, F.; Tuia, D.; Pacifici, F.; Emery, W.J., Svm active learning approach for image classification using spatial information. **2014**.
26. Maulik, U.; Chakraborty, D., A self-trained ensemble with semisupervised svm: An application to pixel classification of remote sensing imagery. *Pattern Recognition* **2011**, *44*, 615-623.
27. Du, Q.; Xia, J.; Chanussot, J.; He, X. In *Hyperspectral remote sensing image classification based on the integration of support vector machine and random forest*, Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International, 2012; IEEE: pp 174-177.
28. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R., Random forests for land cover classification. *Pattern Recognition Letters* **2006**, *27*, 294-300.
29. Pal, M., Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* **2005**, *26*, 217-222.
30. Stumpf, A.; Kerle, N., Object-oriented mapping of landslides using random forests. *Remote Sensing of Environment* **2011**, *115*, 2564–2577.