ADVANCED SAMPLING OF REMOTE SENSING IMAGES FOR AREA ESTIMATION

JACQUES Damien[1], DEFOURNY Pierre[1]

damien.jacques@uclouvain.be, pierre.defourny@uclouvian.be

funded by the *Belgian National Fund for Scientific Research* through a FRIA grant

[1] UCL, Louvain-la-Neuve, Belgique

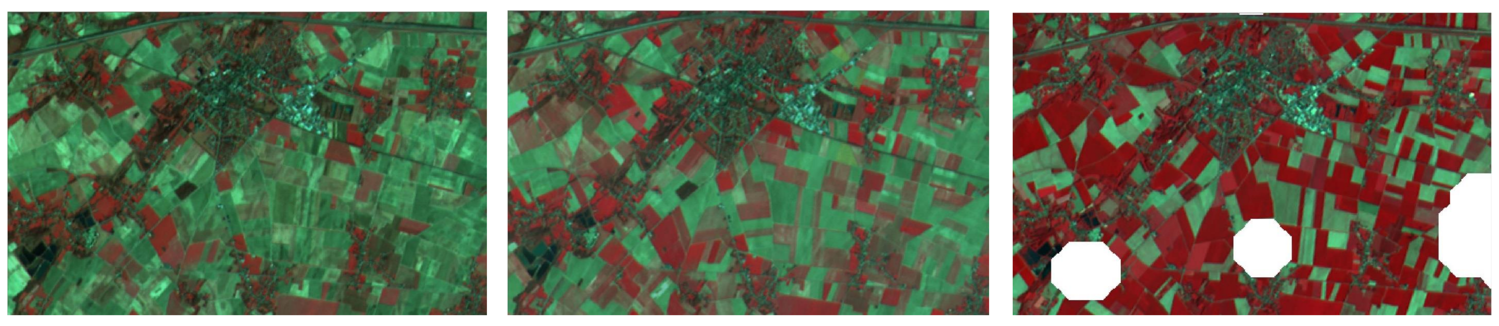
ABSTRACT

Classification methods of Earth observation data obtained by remote sensing allow to produce land cover map of every places of the world, each day, a major asset to compute estimations of area of a particular land cover class (e.g. crop area). However, the commission and omission errors occurring during the classification process (partly due to mixed pixels) are not counterbalanced and therefore, bias the results. Because of this constraint, operational systems have to set up a sampling scheme of in-situ data and/or survey to assess area rather than making use of Earth observation solely. For example, the last version of LUCAS, the crop acreage European monitoring system, is a two-stage systematic design scheme of unclustered points which are identified in the field. Remote sensing data are then used to produce a crop mask and a stratification of the territory to improve the efficiency of the sampling. This research explores ways for improving the accuracy of the information provided by Earth observation data classification by applying sampling directly to the satellite image in order to get a subset classifiable more accurately. The selection criteria used to build the sampling has to be as independent as possible to the land cover type, so ideally derivable a priori i.e. before the classification process. Using a series of 3 SPOT5 images acquired during 2013 growing season in Belgium, this research seeks to develop an object-based method for the estimation of a priori misclassification probabilities of (i) the crop land and (ii) the crop type. The validation is done using the Integrated Administration and Control System (IACS), a vectorial and annually updated GIS which contains information on most of the agricultural parcels in the Walloon region (field limits and crop type). First results show that a set of spectral and spatial object features, completely and automatically extractable from one image, are useful to produce a misclassification probabilities map of the crop land. From this output, a sampling of the image object can be performed based on a trade-off between the sampling rate (to maximize) and misclassification rate (to minimize). The influence of the sampling on the classification bias is studied via the evolution of commission and omission errors with the sampling rate. Those results could be used as a significant input to operational agriculture monitoring systems. In the case of unbiased sampling, it could be directly used as an alternative estimator to field sampling for crop area. In any case, it could improve the efficient of field sampling, given information on area where crop type has been identified with a high confidence level and help to prioritize field visit through the identification of areas hard to classify with remote sensing. That information could also be used for crop yield monitoring because it enables to follow parcels with low misclassification probabilities.

1. Data

Earth Observation data

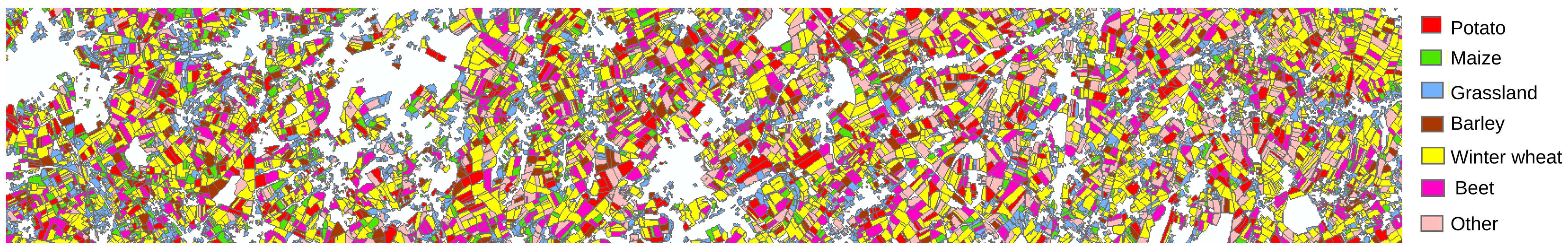
3 images from the series of SPOT4 take 5 acquired in 2013 for the Belgium site are used.



Subsets of SPOT 4 take 5 images for the dates, April 2, April 22, and June 6 for the 2013 growing season (white areas represent cloud and shadow)

Validation data

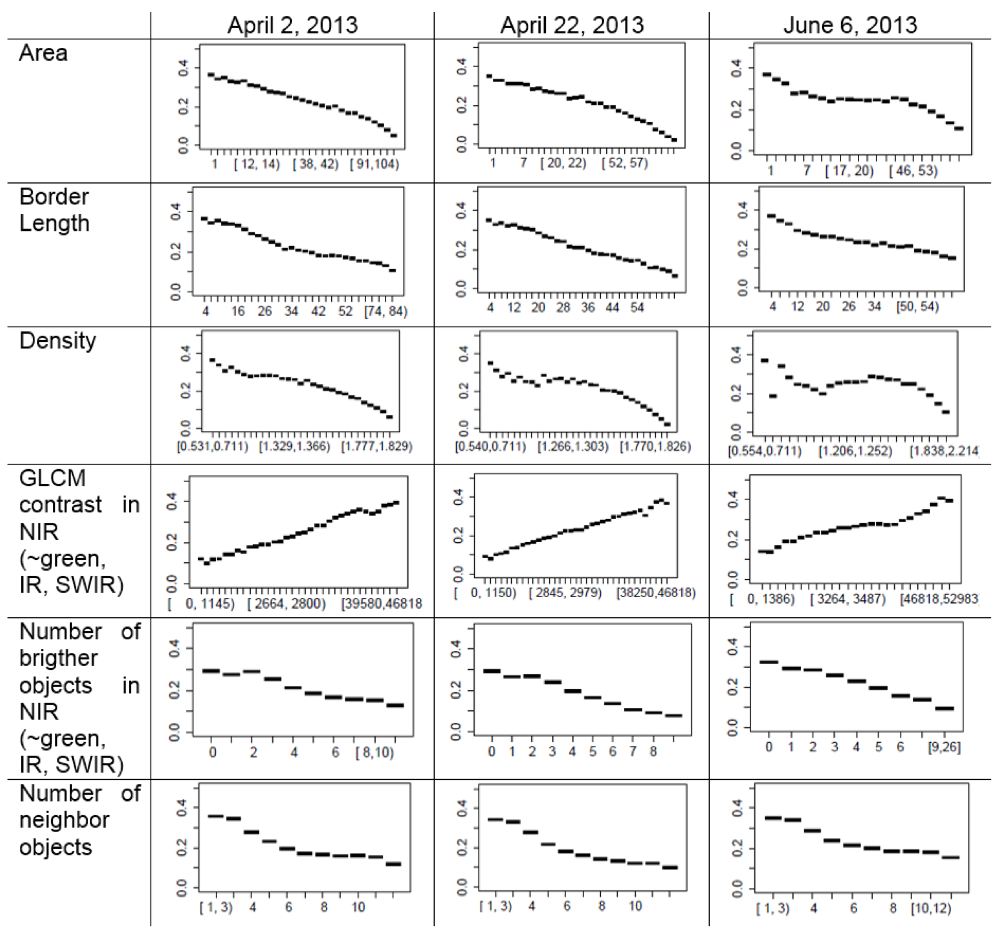
The validation is done using the Integrated Administration and Control System (IACS), a vectorial and annually updated GIS which contains information on most of the agricultural parcels in the Walloon region (field limits and crop type). The average field area is around 3 ha in this region.



2. Improvement of classification accuracy by sampling the image

Relationship between object features and misclassification

A multiresolution segmentation is applied on each images with the same parameters. 56 object features (4 categories: shape, spectral, neighborhood and localisation) have been computed for each objects. 6 of them show a strong relationship, for each date, between the classification (k-means) error and the feature values.



Relationship between classification error (y-axis) and different object features values (x-axis)

GLCM contrast: The gray level co-occurrence matrix (GLCM) is a tabulation of how often different combinations of pixel gray levels occur in a scene. GLCM contrast is a measure of the amount of local variation in the image.

Density: The Density feature describes the distribution in space of the pixels of an image object. The most "dense" shape is a square; the more an object is shaped like a filament, the lower its density. The density is calculated by the number of pixels forming the image object divided by its approximated radius, based on the covariance matrix.

Toward a mapping method of misclassification probability

Area, *GLCM* and the *Number of brighter objects*, 3 features less correlate than others and falling into three different categories of features: shape, spectral (texture) and neighborhood. Their combination could be bring further information.

April 2, 2013

	Area	GLCM Con NIR	Num Bright NIR
Area	1	-0.4	0.48
GLCM Con NIR	-0.4	1	-0.26
Num Bright NIR	0.48	-0.26	1

April 22, 2013

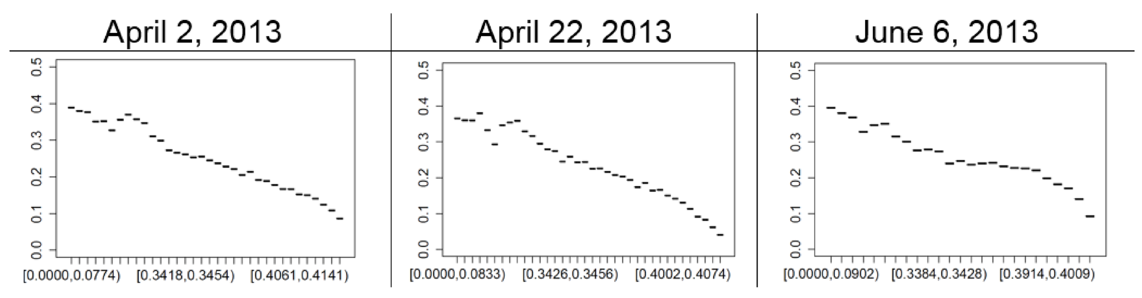
	Area	GLCM Con NIR	Num Bright NIR
Area	1	-0.38	0.41
GLCM Con NIR	-0.38	1	-0.26
Num Bright NIR	0.41	-0.26	1

June 6, 2013

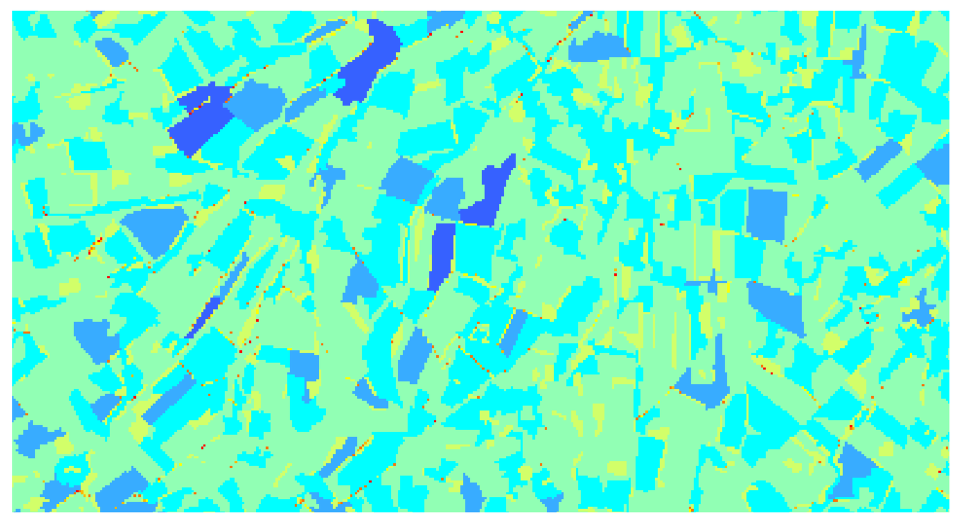
	Area	GLCM Con NIR	Num Bright NIR
Area	1	-0.35	0.48
GLCM Con NIR	-0.35	1	-0.27
Num Bright NIR	0.48	-0.27	1

Correlation matrix between Area, GLCM and the Number of brighter objects features for each date

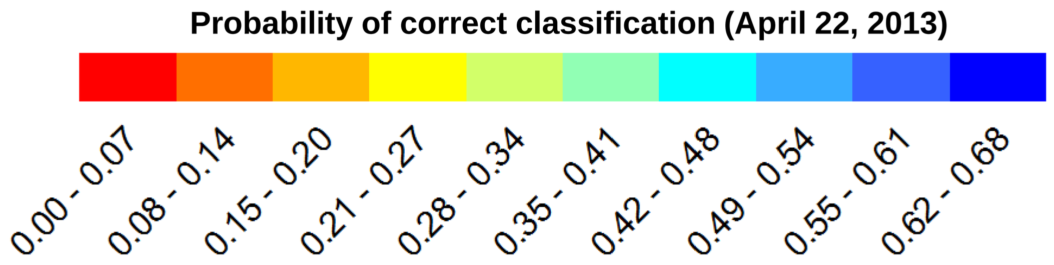
The combination of the 3 features can be used to predict the probability of an object to be correctly classified (same weight for the 3 features). The relationship with the classification error is slightly improved than for a single variable. This result allows to produce a map of misclassification probability based only on those 3 features.



Relationship between classification error (y-axis) and the probability of correct classification (x-axis)

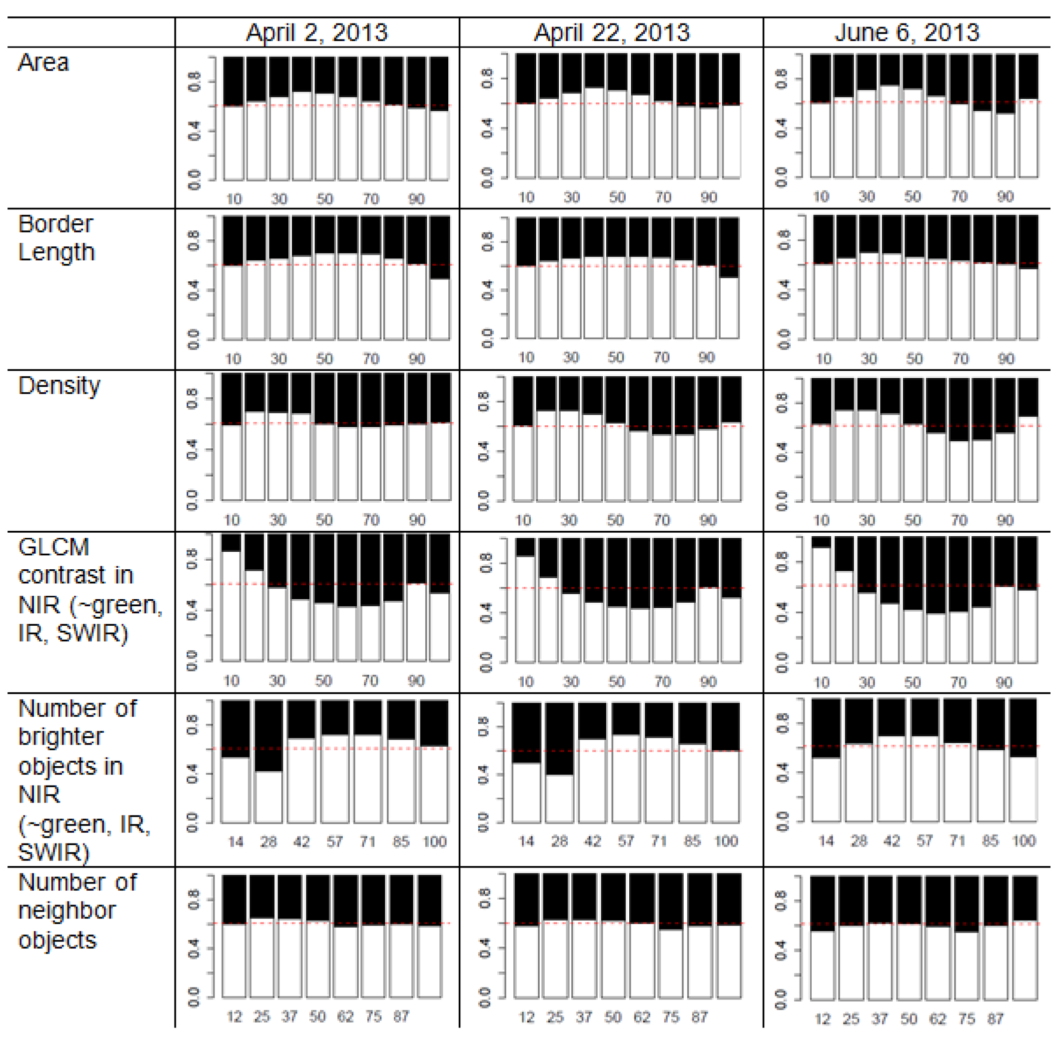


Probability of correct classification (April 22, 2013)



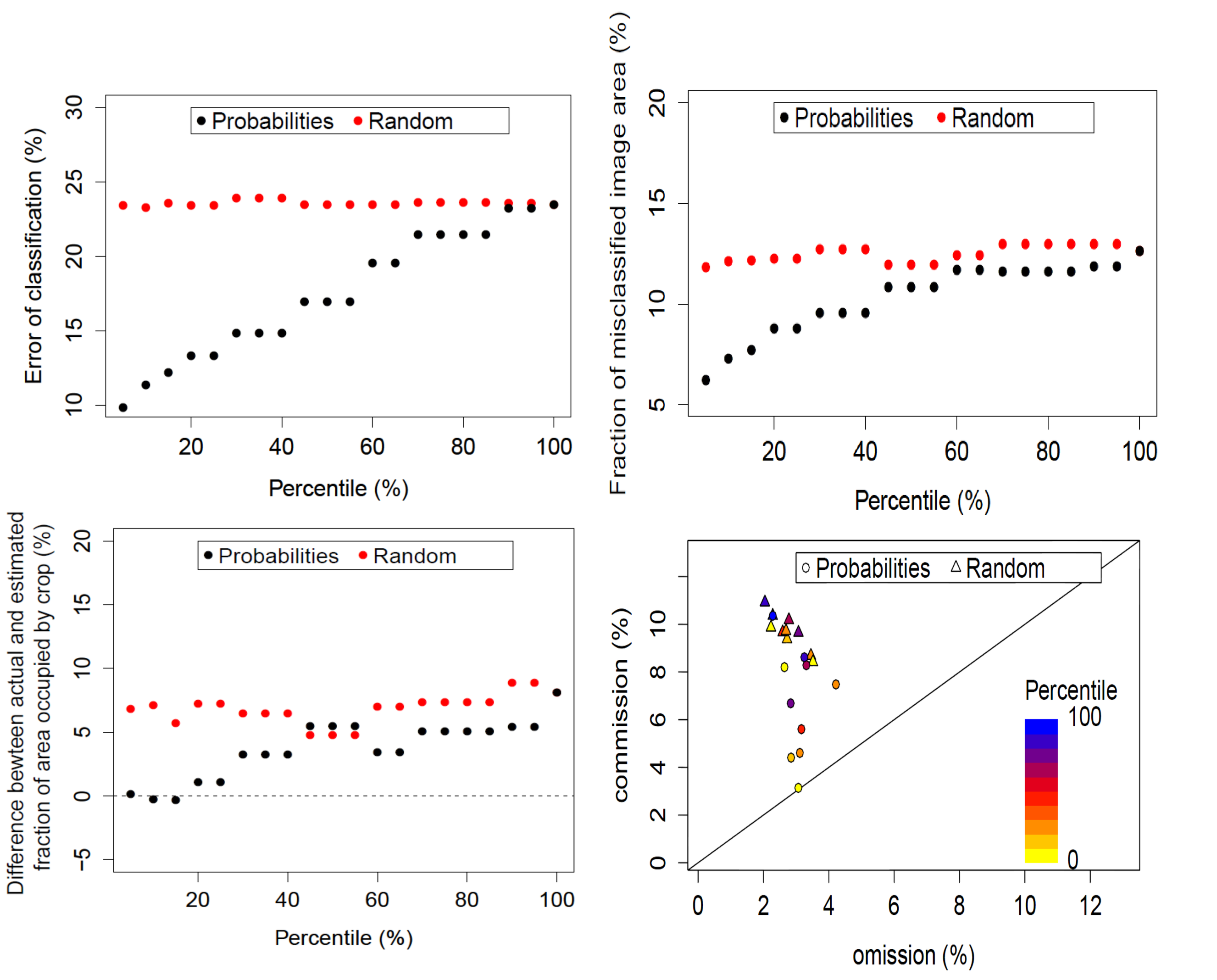
5. What about the crop area estimation ?

In order to improve the crop area estimation, the sampling of the objects that leads to the improvement of the classification accuracy should not have an effect on the ratio between cropland and non-cropland area. Among the identified features, only the *Number of neighbor objects* meets this objective.



Relationship between the fraction of class area (cropland in white and non-cropland in black) and the percentile of each feature distribution. The dotted red line shows the actual fraction of class area.

The figures below show an example of the use of the *Number of neighbor objects* feature to sampling the April 22 image. The method is compared with a same-size random sampling. The classification accuracy is improved, the relative number of objects (top left) and total area (top right) that are correctly classified increases with the sampling rate. The crop area estimation is also improved (bottom left) partly due to a balance between commission and omission error.



CONCLUSION

Results

A statistical approach to spatialize the misclassification probability based on object features easily retrieved (shape, spectral and neighborhood) has been developed. This approach is consistent through the growing season. Furthermore, it has been shown that an unbiased method to improve the crop area estimation *a priori* can be implemented. In the presented case, it is based on the *number of neighbor objects* feature.

This method can be fully automated except for the labelling of the unsupervised classification.

Those results could improve the efficient of field sampling, given information on area where crop type has been identified with a high confidence level and help to prioritize field visit through the identification of areas hard to classify with remote sensing. That information could also be used for crop yield monitoring because it enables to follow parcels with low misclassification probabilities.

Further research

The method has to be generalized under different landscapes and through different years (first tests for 2009 show similar results).

Different algorithms of segmentation and classification methods have also to be tested.