Parcel-Based Active Learning for Large Extent Cultivated Area Mapping

Ines Ben Slimene Ben Amor, Nesrine Chehata, Jean-Stéphane Bailly, Imed Riadh Farah, Philippe Lagacherie

Parcel-based active learning for large extent cultivated area mapping

Ines BEN SLIMENE BEN AMOR, Nesrine CHEHATA,*IEEE Member*, Jean-Stéphane BAILLY, Imed Riadh FARAH and Philippe LAGACHERIE

***Abstract*—This paper focuses on agricultural land cover mapping at a high resolution scale and over large areas from an operational point of view and from a high-resolution monodate image. In this context, training data are assumed to be collected by successive journeys of field surveys and thus, are very limited. Supervised learning techniques are generally used, assuming that the classes distribution is constant over the whole image. However in practice, a data shift often occurs on large areas due to various acquisition conditions. To alleviate these issues, active Learning (AL) techniques define an efficient training set by iteratively adapting it through adding the most informative unlabeled instances. They can improve the classification process efficiency while keeping a limited training dataset. The novelty in this paper is the application of AL techniques on multispectral images for agricultural land cover mapping, using field sampling instead of pixel sampling which is rarely done in the literature. Besides, we proposed a parcel-based active learning scheme which is suitable for an operational land cover mapping in cultivated areas since the parcel is an agricultural unit and field observations are processed at parcel scale. Random Forests classifier was used. Results were processed on a 6m multispectral Spot6 image over a 35 km^2 Mediterranean cultivated area, in Lebna Catchment, north eastern Tunisia.**

The contribution of AL techniques was assessed with comparison to a random and stratified random strategies for sampling new instances. For iterative sample selection, two criteria are used and often coupled: uncertainty and diversity. For diversity metric, a new clustering-based metric was proposed based on a mean-shift clustering which improved the classification accuracy. AL techniques showed to be efficient with complex data and fine land cover legend improving random-based selection up to 10%. Besides, the maximum of classification accuracy is reached using mean-shift-BT metric in just 5-day field survey, i.e. 30 days less compared to the random selection. Finally, results showed that the finer the definition of land cover classes, the more crucial is the choice of AL metrics.

***Index Terms*—Active Learning, object-based, Random Forest, Uncertainty, Diversity, Multispectral, mean-shift clustering, Agriculture**

I. Introduction

Agricultural practices and land cover are major drivers of water flows in cultivated landscapes. The spatial arrangements and connectivities of cultivated fields have a strong impact onto the run off and soil erosion at the landscape and watershed scales [1]. Land cover and use have also a great impact on water quality [2], [3]. In practice, there is a need for agricultural land cover mapping at a scale and extent that could be integrated into hydrological models to better simulate the considered hydrological behavior of catchment and thus improve water resources management. This impact on hydrological fluxes is mainly required at large extent which corresponds to water management scale. Besides, the context of this study is Mediterranean cultivated areas which present a high level of parcel fragmentation, with small and lineform parcels ($\sim$ 1 ha) which needs to be mapped with a high spatial resolution imagery.

Remote sensing data and supervised classification techniques even on monodate image are well appropriate for mapping agricultural land cover at a high resolution scale and over large areas [4]. In remote sensing classification problems, the collection of labeled samples can be derived according to the following: 1) field survey; which is associated to high cost and requires time, 2) photo-interpretation; which is cheaper and faster but less accurate, even not appropriate for some crop discrimination (e.g. barley and wheat), or 3) hybrid solutions; where both photo-interpretation and ground are used. The choice of labeling strategy depends on the considered problem and the image type [5].

For agricultural landscapes, the supervised classification task has proven to be challenging due to highly unbalanced data (i.e unbalanced crop surfaces), a high intra-variability of crop parcels and to data shift issues over large study areas. Indeed, the classes distribution is assumed to be constant over the whole image. The classifier model is built locally and then applied to the whole image. However in practice, a data shift often occurs on large areas due to various acquisition conditions for instance. Therefore, the classifier model should 1) exploit more unlabeled samples since labeled samples are very limited and 2) be adapted by adding the most informative ones from non-represented areas in the initial training set.

In machine learning literature, two kinds of learning methods can alleviate the problem of limited training sets. Semi-supervised Learning (SSL) and Active Learning (AL) methods allow to use unlabeled data in the classifier model [6]. SSL methods use both labeled and unlabeled data. No manual labeling is needed. Many SSL techniques assume a default principle where closely located data points are likely to share similar label information. Active Learning (AL) techniques define an efficient and parsimonious training set by iteratively adapting it through adding the most informative unlabeled instances [7]. These instances are then labeled manually before re-training the classifier.

In agricultural context, with a highly fragmented landscapes

This research was supported by the French National Research Agency (ANR) through the ALMIRA project (ANR-12-TMED-0003).

Ines Ben Slimene Ben Amor is PhD student between UMR LISAH (Supagro, IRD, INRA) Montpellier and RIADI Lab (ENSI) Tunisia.

Email: inesbenslimen@hotmail.com / ines.ben-slimane@supagro.fr

and a high class intra-variability, manual labeling is required and neighborhood homogeneity hypothesis can not be assumed. For these reasons, we propose to use active learning techniques to build iteratively the most efficient learning sampling that ensures the best classification accuracy.

AL has shown a great potential in remote sensing to improve the efficiency of the classification process while keeping a limited training dataset [8]. The AL process is conducted according to an iterative process. At each iteration, the most informative unlabeled samples are chosen for manual labeling, and the supervised model is processed with additional labeled samples. In this way, the unnecessary and redundant labeling of non-informative samples is avoided, greatly reducing the labeling cost and time [9]. This is particularly well suited to a context where the manual labeling is processed by field survey. Thus, the field sampling should be efficient and parsimonious.

AL techniques are often pixel-based and are processed by photo-interpretation [8]. In our context, we proposed a parcel-based active learning scheme which is suitable for an operational land cover mapping in cultivated areas, since the parcel is an agricultural unit and field surveys are processed at parcel scale. Indeed, such object-based approaches are less time consuming than the classical pixel-based approaches along in classification task and for field surveys. In this paper, our goal is to answer these questions : are AL techniques well suited in agricultural context and using multispectral imagery? How to propose an operational AL strategy based on parcel field survey? In which cases, AL techniques are more efficient (providing a gain in terms of time and quality) than random sampling strategies ? Is the choice of AL metrics critical ?

Results were processed on a 6m multispectral Spot6 image on a 35 km^2 Mediterranean cultivated area, in Lebna Catchment, north eastern Tunisia. An hierarchical land cover legend was used to assess the contribution of AL techniques depending on the needed definition of land cover classes. This paper is organized as follows: in section II, AL techniques and metrics are detailed, study site and data are presented. The sub section II-C details the proposed methodology, the random forest classifier, the chosen features and the experimental setup. Results are presented and discussed in section III and IV, respectively. Finally conclusions are drawn.

II. Material and methods

A. AL state-of-the-art methods

A general AL process can be modeled as an hexatruple of (L, U, C, F, S, N) where L is the set of labeled samples and U is the set of unlabeled data. C is the classifier, which is trained on the labeled training set L. F is a function used to select the most informative unlabeled samples from U. The AL process is an iterative process, where the user interacts with the system by iteratively labeling the most informative samples selected by the function F at each iteration. At the first iteration, an initial training set L of few labeled samples is required for the first training of the classifier C. After initialization, the function F is used to select a set of samples S from the U, and the user assigns them the true class label. Then, these new labeled samples are included into L, and the classifier C is re-trained using the updated training set. This loop continues for N iterations [9]. In the literature, most of works on active learning for remote sensing data are conducted over hyperspectral data and using the SVM classifier where samples correspond to pixels [8]. AL techniques are generally based on individual pixels, batches [8], or superpixels [10] which allows to decrease calculation time. Up to our knowledge, no work exists using active learning techniques on agricultural context using a parcel-based approach. The main question is "What are the selection methods to be applied in order to choose the most informative and representative samples?" This point is crucial for the success of an AL algorithm : the machine needs a strategy to rank data. Two criteria can be used and often coupled: uncertainty and diversity. The samples should be the most informative (i.e uncertain for the current classifier model) and diverse (i.e non redundant).

1) AL uncertainty metrics: Based on the literature in remote sensing and in other applications, uncertainty measure metrics can be grouped into three main families : 1) Sampling by uncertainty, 2) Sampling by committee and 3) Sampling by model re-estimation.

In the first family, *Sampling by uncertainty*, the most informative unlabeled samples are the most uncertain ones. Three uncertainty metrics are usually used in literature (Entropy, Breaking ties (BT) and Class boundaries). Entropy metric takes into account the posterior class probability and the most uncertain sample maximizing this entropy [11]. While the Breaking Ties metric is based on the difference between the two highest posterior class probabilities. The lower the difference between both highest probabilities, the lower the confidence of the model. Uncertain sample minimizes this confidence [8]. [12] introduced the metric based on class boundaries. Uncertain samples are located in class boundaries. The large margin classification is the most known approach based on this principle [8] as for instance the support vector machines (SVM) methods. The separation absolute distance of the hyperplane is a simple way to estimate the confidence of the model for an unlabeled sample. For large margin metrics, there are Margin sampling (MS) [13], MultiClass Level Uncertainty (MCLU) [9] and Significance Space Construction (SSC)[14]. However, the large margin heuristics are dependent on the SVM classifier.

In the second family *sampling by committee*, the uncertainty of a sample is measured by considering a learning committee [8], [11]. Committees can be built in different ways. Either from a single classifier that generates several models by varying the learning set (Boosting or bagging)[7] or by dividing the feature space into subsets or views (subsets of bands)[15] or also by combining several types of classifiers that generate different models. The algorithm then selects samples showing maximum disagreement between the different classification models in the committee. To measure this disagreement, there are two approaches (Entropy and Kullback-Leibler (KL) divergence). For entropy, the committee disagreement is estimated by an entropy measure on the posterior class probabilities. Maximal

entropy corresponds to a maximal disagreement between the committee models i.e. maximal uncertainty. This approach has two metrics, Normalized Entropy Query-by-Bagging (nEQB) where the bagging is used to build the committee [16] and Adaptive Maximum Disagreement (AMD) where Di & Crawford [17] have generated different views of the hyperspectral image based on the band correlation index. According to the divergence of Kullback-Leibler (KL) approach, the sample that maximizes the Kullback-Leibler divergence is selected. This approach takes into account the modification in classes distribution between models. The KL divergence is then calculated between the class distribution estimated by a committee model and the mean distribution estimated by all committee models. Maximal divergence corresponds to a maximal disagreement between the models of the committee [18].

Finally, the *sampling by model re-estimation* family corresponds to methods that quantify the uncertainty of a sample by estimating the impact of its inclusion in the learning set. In order to measure this impact, there are three approaches (minimization of estimated error, Kullback-Leibler-max and Breaking Ties). The first approach aims to select the sample that, when added to the learning set, minimizes the classification error [19]. While Kullback-Leibler-max tend to select samples that there inclusion in the training set maximizes changes of the posterior distribution. The KL divergence is then calculated between the posterior distributions of the models with and without the sample [20]. The sample that maximizes the divergence is selected. The Breaking Ties approach in this family is applied to the two highest estimated posterior probabilities in order to evaluate the confidence of the new model (with the new sample)[8].

2) AL diversity metrics: The diversity measure concerns the ability of the model to reject redundant samples in the training. Based on the literature in remote sensing, diversity metrics can be grouped into three main families : 1) Distance, 2) Angle and 3) Clustering.

The distance-based diversity is based on similarity distances between the sample spectral vectors [7]. Three metrics are usually used in literature: close Support Vector (cSV), Most Ambiguous and Orthogonal (MAO) and Mahalanobis Distance. In close Support Vector (cSV) measure, the samples diversity was ensured by choosing samples associated to the nearest support vectors (cSV). This method ensures diversity between new samples and the current training set, but does not guarantee the diversity of samples between them. This metric is dependent on the SVM classifier. Most Ambiguous and Orthogonal (MAO) metric is based on the Kernel function between two candidate samples which decreases with the increase of the distance between them. This function encourages the selection of unlabeled samples that are different between each other [21]. Mahalanobis Distance metric corresponds to the distance between a sample and the candidate sample distribution. The highest distance corresponds to the most diverse sample [22].

The angle-based diversity family is based on measuring a spectral angle to evaluate the similarity. For instance, Angle-Based Diversity (ABD) approach measures samples diversity by considering the angle cosinus between them [9]. Samples that minimize the cosinus (maximize the spectral angle) are selected as the most diverse.

The clustering-based diversity family measures sample diversity by spectral clustering. Clustering techniques evaluate the distribution of the samples in a feature space and group the similar samples into the same clusters. Since the samples within the same cluster are correlated and provide similar information, a representative sample is selected for each cluster. In literature, there are several metrics such as : Cluster Based Diversity (CBD), Enhanced Cluster Based Diversity (ECBD), informative hierarchical Margin Cluster Sampling (hMCS-i) and Fuzzy Kmeans (FKM). In CBD, cluster centers are selected as the most representative samples [9]. ECBD method is similar to CBD with an enhanced selection of representative samples. In [9], K-means [23] was used to cluster the samples. In [8], the hMCS-i method was explained. This heuristic aims to exclude from selected samples the most likely to become a Support Vector. Therefore, the redundancy affecting samples close to each other in the feature space among different iterations is controlled. This metric is the combination of the cSV and ECBD methods and it is also highly dependent on the SVM classifier. The Fuzzy Kmeans is a clustering method that calculates a similarity measure between a sample and a cluster and determines how much they are alike in multi-variable space [22].

B. Site & data

1) Study site: The study site (Fig.1) is located in the Cap Bon region in northern Tunisia, 60 km east of Tunis. The study area covers 35 km^2 and is included in the Lebna catchment (in red on Fig.1), which is characterized by a semi-arid climate. It is mainly rural and devoted to cereals in addition to legumes, olive trees, natural vegetation for breeding and vineyards. Besides the cultivated landscape shows a high fragmentation of the parcels, where up to 80% of parcels are less than 1 ha with a high intra-parcel variability leading to a high intra-class variability.

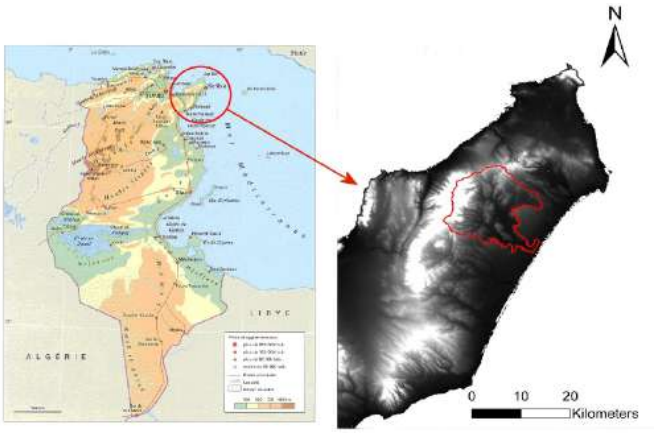

Fig. 1: Study area location - Lebna catchment in North Eastern Tunisia

2) Satellite image: A Spot-6 multispectral satellite was acquired on the 21^{st} of March 2016. It has a spatial resolution of 6 meters, dynamic range of 12 bits per pixel and 4 spectral bands (B, G,R and NIR). The acquisition conditions were optimal with 0% cloud cover (Fig.2).

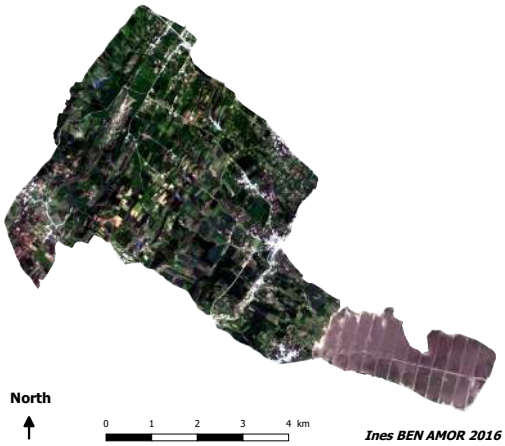

Fig. 2: Study area SPOT-6 satelitte image

3) Ground Truth data : First, the parcels boundaries were digitized using a Pleiades image (2m MS) from 2013. The ground truth was obtained by a 24-day field survey during the period of maximal growth vegetation (April 2016). The area has been divided to tiles (The tile corresponds to 600m * 600m). Adjacent tiles could be processed in one day. Observations were made either directly in the parcel (in case of high ambiguity on the crop type) or using pair of binoculars from high altitude point of view for non-ambiguous crops (such as coriander or beans). During our field campaign, almost 60 parcels could be labeled per day considering 4 adjacent tiles. The land cover observations were made at the parcel scale over 3198 parcels. The whole dataset is then used to find the best strategy for the active learning. An hierarchical legend was used, that is inspired from Corine land cover legend and adapted to the local context (TABLE I). In this work, we focused on the more semantically accurate levels, i.e. Level 3 (L3) which has 19 classes and Level 4 (L4) with 24 classes. The corresponding ground truth maps and legends are illustrated in figures Fig.3 and 4, respectively. One can observe that the distribution of classes is highly unbalanced which is challenging for the land cover classification task. The differences between both levels correspond to more L4 sub-classes of *arable land* class, which is the major class on our study site. In order to clarify some ambiguities between classes: for level 4, *Fodder* parcels were distinguished from *Cereals* since they are observed in field as parcels under pasture. In addition, the landscape is very fragmented with a high variability of parcel surfaces.

C. Methodology

In this section, the proposed methodology will be presented as well as the chosen classifier, features and the selected AL state-of the art metrics. The proposed mean-shift clustering diversity metric will be introduced. Finally the experimental setup will be presented.

TABLE I: L3 and L4 Land cover legend.

L3 land cover legend	**L4 land cover legend**
Continuous urban fabric	Continuous urban fabric
Discontinuous urban fabric	Discontinuous urban fabric
Industrial or commercial units	Industrial or commercial units
Road and rail networks	Road and rail networks
Arable land	Cereals
	Pulse crop
	Vegetable
	Fodder
	Fallow
	Plowed land
Vineyards	Vineyards
Fruit trees and berry plantations	Fruit trees and berry plantations
Olive groves	Olive groves
Pastures	Pastures
Heterogeneous agricultural areas	Heterogeneous agricultural areas
Forests	Forests
Shrub and/or herbaceous vegetation	Shrub and/or herbaceous vegetation
Natural grassland	Natural grassland
Moors and heathland	Moors and heathland
Sclerophyllous vegetation and garrigues	Sclerophyllous vegetation and garrigues
Other open spaces	Other open spaces
Bare rock	Bare rock
Inland marshes	Inland marshes
Water bodies	Water bodies

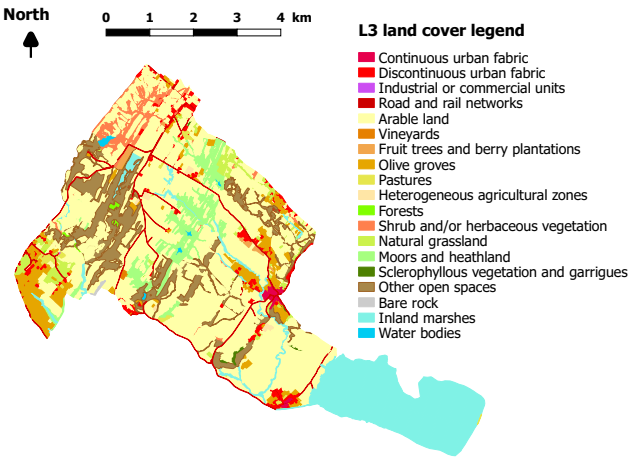

Fig. 3: Study area ground truth (Level 3)

Assuming that:

- L_i : Set of labeled training parcels in the iteration i
- U_i: Set of unlabeled training parcels (candidates)
- T: Fixed set of test parcels (validation)
- C: Classifier
- F_{Un} : Function to select the most uncertain parcels
- F_{Div}: Function to select the most diverse parcels
- S_{Un_i} : Set of uncertain selected parcels per iteration (uncertain Batch)
- S_i : Set of final selected parcels per iteration (Batch)
- N : Number of iterations

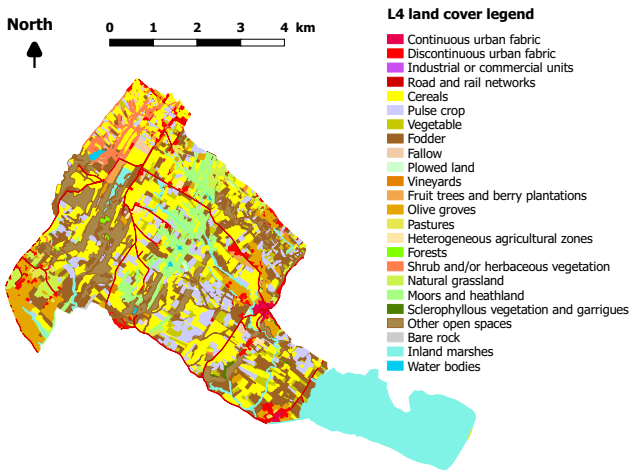

Fig. 4: Study area exhaustive ground truth (Level 4)

Algorithm 1 : Parcel-based AL algorithm

$i \leftarrow 1$
$L_i \leftarrow L_1$ (initial training set)
while $i < N$ **do**
  Train the classifier C with the training set L_i;
  Predict the unlabeled samples in U_i;
  Validate predictions in iteration i (accuracy measure OA and Kappa), using test set T;
  for each candidate (unlabeled parcel) x in U_i **do**
    Evaluate the user defined function F_{Un};
  end for
  Rank candidates in U_i according to the score of F_{Un};
  Select the most uncertain parcels S_{Un_i};
  for each uncertain selected parcel x in S_{Un_i} **do**
    Evaluate the user defined function F_{Div};
  end for
  Rank candidates in S_{Un_i} according to the score of F_{Div};
  Select the Batch S_i;
  Assign labels to the Batch S_i by a user;
  Remove the Batch from candidates : $U_{i+1} \leftarrow U_i \setminus S_i$
  Add the Batch to the training : $L_{i+1} \leftarrow L_i \cup S_i$
  $i \leftarrow i + 1$
end while

1) Classifier : Random Forest : The proposed methodology is generic and can be applied to different classifiers since they provide a posterior class probability. In [4], the Random Forests (RF) and Support vector machines (SVM) performances were compared on various datasets in a similar context: high spatial resolution and large scale agricultural environment. Authors showed that RF classifier is more efficient than SVM in this specific application case.

RF [24], is a modification of bagging applied with decision trees. It can achieve a classification accuracy comparable to boosting [24] or SVM [25]. This classifier is a combination of tree predictors built from multiple bootstrapped training samples. For each node of a tree, a subset of features is randomly selected. Then, the best test in the node is calculated with an evaluation measure over the subset of features. For classification, each tree gives a unit vote for the most popular class at each input instance. The final label is determined by a majority vote of all trees.

2) AL metrics: As presented in section II-A, there are several AL methods that allow to choose the most informative samples in terms of uncertainty and diversity. With respect to our thematic context and operational constraints, some AL state-of-the-art metrics were chosen to be tested. First, since the methodology should be independent from the classifier, all large margin based metrics were not considered since they are based on SVM classifier.

In addition, since the labeling is made by field surveys, our initial training set is very limited. Thus, for uncertainty metric, the two last families based on *sampling by committee* and *Sampling by model re-estimation* were not considered since they are not robust to a small initial training sets [8]. Therefore, we selected sampling by uncertainty family and we kept two metrics that are entropy and breaking ties (BT).

As for diversity metrics, hMCS-i and cSV metrics were not considered since they are dependent of SVM classifier. Angle-based metric ABD has an additional calculation cost. MAO is not effective with a multi-class problem. Thus the kept distance-based metric is Mahalanobis distance [26] although is it parametric. As for the clustering-based metrics three metrics could be selected in our context (CBD, EBCD and FKM) but they require the pre-definition of the number of clusters. The number of clusters may be set to the batch size. However, since in the case of field labeling, larger batches are preferred [8], this parameterization (according the number of clusters to the batch size) highly increases the clustering computing time. To resume, the selected reference AL metrics in our context are Entropy [11] and Breaking Ties [8] as uncertainty measures and the Mahalanobis distance [26] as a diversity measure.

The new proposed AL diversity metric is a clustering-based one. The mean-shift algorithm [27] was chosen since it is a non-parametric clustering technique which does not require a prior knowledge of the number of clusters.

Mean-shift clustering is a procedure for locating the maxima of a density function given discrete data sampled from that function[28]. It is useful for detecting the modes of this density. This is an iterative method, and it starts with an initial estimate x. Let a kernel function $K(x_i - x)$ be given. The first question, then, is how to estimate the density function given a sparse set of samples. One of the simplest approaches is to just smooth the data by convolving it with a fixed kernel of width h. Given n data points $x_i, i = 1, 2, ..., n$, the multivariate kernel density estimate obtained with kernel K and h the mean-shift parameter called the bandwidth [27]:

$$f(x) = \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad (1)$$

Once $f(x)$ has been computed from equation (1), the local maxima can be found using gradient ascent or any

optimization technique.

After mean-shift clustering of uncertain batch S_{Un}, each unlabeled sample is associated to a single cluster. Euclidean spectral distances between each sample and its corresponding centroid are calculated. All samples are ordered by increasing distances. The closest samples $(Card(S_i))$ to their centroid with regard to the spectral distance are selected as the most diverse.

3) Data description: In [4] the authors compared different input feature (spectral bands, spectral and textural index) and proved that the most important features for agricultural land cover mapping are the spectral bands and the Normalized Difference Vegetation Index (NDVI). In our parcel-based approach, the attributes were derived per parcel by calculating the mean and standard deviation of the spectral values (B, G, R, NIR) of all parcel pixels and similarly for the NDVI feature.Thus, ten features were used as input of the random forest classifier.

4) Experiments: For AL methods, the data was split into three independent datasets using a stratified random method: a train set (labeled data), a candidate set (unlabeled data) and a test set (validation data). These sets are kept the same for all AL approaches. Tables II and III resume the experimental setup for land cover legend at levels 3 and 4, respectively. One can observe that the class distribution is highly unbalanced. Besides, for some minor classes, only one training sample (parcel) is used. The most present class correspond to *arable land*.

TABLE II: Experimental setup for L3 land cover legend with 19 classes.

Classes	Train set (L)	Cand set (U)	Test set (T)	Total
Continuous urban fabric	1	5	1	7
Discontinuous urban fabric	4	156	16	176
Industrial or commercial units	1	0	0	1
Road and rail networks	1	8	1	10
Arable land	72	2320	247	2639
Vineyards	1	0	0	1
Fruit trees and berry plantations	1	9	1	11
Olive groves	4	123	13	140
Pastures	1	24	3	28
Heterogeneous agricultural areas	1	5	1	7
Forests	1	4	0	5
Shrub and/or herbaceous vegetation	1	20	2	23
Natural grassland	3	71	8	82
Moors and heathland	1	14	1	16
Sclerophyllous vegetation and garrigues	1	4	1	6
Other open spaces	1	14	2	17
Bare rock	1	0	0	1
Inland marshes	1	23	3	27
Water bodies	1	0	0	1
Total	**98**	**2800**	**300**	**3198**
Percentage %	**3,06**	**87,56**	**9,38**	**100**

The RF classifier has been applied with $T = 50$ trees , the number of considered features per node has been fixed to the mean square of total feature number [24], i.e. $M = 3$. Classification accuracies, overall accuracy (OA) and Kappa (K), were averaged over 50 runs of Random Forest.

TABLE III: Experimental setup for L4 land cover legend with 24 classes.

Classes	Train set (L)	Cand set (U)	Test set (T)	Total
Continuous urban fabric	1	5	1	7
Discontinuous urban fabric	4	156	16	176
Industrial or commercial units	1	0	0	1
Road and rail networks	1	8	1	10
Cereals	26	859	91	976
Pulse crop	17	474	51	542
Vegetable	7	202	22	231
Fodder	20	755	80	855
Fallow	1	23	2	26
Plowed land	1	7	1	9
Vineyards	1	0	0	1
Fruit trees and berry plantations	1	9	1	11
Olive groves	4	123	13	140
Pastures	1	24	3	28
Heterogeneous agricultural areas	1	5	1	7
Forests	1	4	0	5
Shrub and/or herbaceous vegetation	1	20	2	23
Natural grassland	3	71	8	82
Moors and heathland	1	14	1	16
Sclerophyllous vegetation and garrigues	1	4	1	6
Other open spaces	1	14	2	17
Bare rock	1	0	0	1
Inland marshes	1	23	3	27
Water bodies	1	0	0	1
Total	**98**	**2800**	**300**	**3198**
Percentage %	**3,06**	**87,56**	**9,38**	**100**

AL approaches were also applied with a random and a stratified random approaches for iterative unlabeled sample selection. Besides, different reference AL metrics for uncertainty and diversity were combined and tested so that six AL approaches were tested for selecting new samples (batch) as following: Random, Stratified Random, Mahalanobis-Entropy, Mahalanobis-BT, Mean-shift-Entropy and Mean-shift-BT.

The proposed mean-shift based diversity has been parameterized as follows: the width $h = 20$ and the used kernel function is "epanechnikovKernel", which is a discontinuous parabola kernel with a restricted support that is recommended with the blurring mean-shift version.

$$K(x) = \frac{3}{4}(1 - x^2)l_{(|x|\leq 1)} \tag{2}$$

Where $l_{(x)}$ is the indicator function that gives 1 when x is true and 0 if not. As for AL parameterization, for each iteration i, the added samples (batch) have to be labeled by field survey. Thus, we fixed the batch size to 65 parcels (S_i) which corresponds approximately to a one day field survey. First, the size of uncertainty batch S_{Un_i} is fixed to 200 and then the 65 more diverse unlabeled parcels S_i are selected The

maximal number of iterations N was fixed to 40 in order to use all the candidates.

III. Results

R library was used for experiments [29]. Package mean-shift was used to run the corresponding diversity metric.

A. Comparison of AL approaches

As mentioned above, six AL approaches were compared. Figures 5 and 6 illustrate the curves of overall accuracy for all approaches for both land cover levels: L3 and L4, respectively. As for X axis, bottom axis shows the number of training parcels at each AL iteration. The number of iteration appears on the top x-axis, which nearly equals the number of field days. Red and black transparent polygons show the 95% confidence bounds for random and stratified random approaches calculated over 500 random batches per iteration.

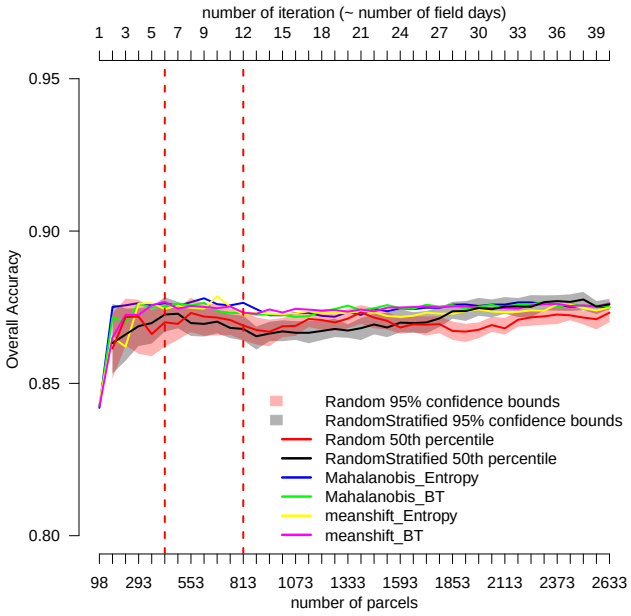

Fig. 5: Comparative curve of metrics OA for Level 3

First, on figure 6, one can observe that generally AL metrics do better with few iterations than random selection of samples for level 4. Among all AL metrics, only the proposed Mean-shift-BT AL metric exceeds the confidence interval of Stratified Random and improves overall accuracy up to 3.3% at 6^{th} iteration. Besides, the mean-shift-BT AL metric has the faster OA improvement curve since the first iterations in comparison with other AL metrics. However, when comparing classification accuracies for level 3 and level 4, one can observe that the initial OA is enhanced up to 9.7% for level 4, while it is slightly enhanced for level 3 reaching 3.6%. This means that the contribution of AL approaches is higher for finer land cover legends, i.e. for a finer definition of land cover classes. Moreover, for Level 3, all AL metrics lead to very similar results, while for level 4, the mean-shift-BT highly improves the overall accuracy with comparison to

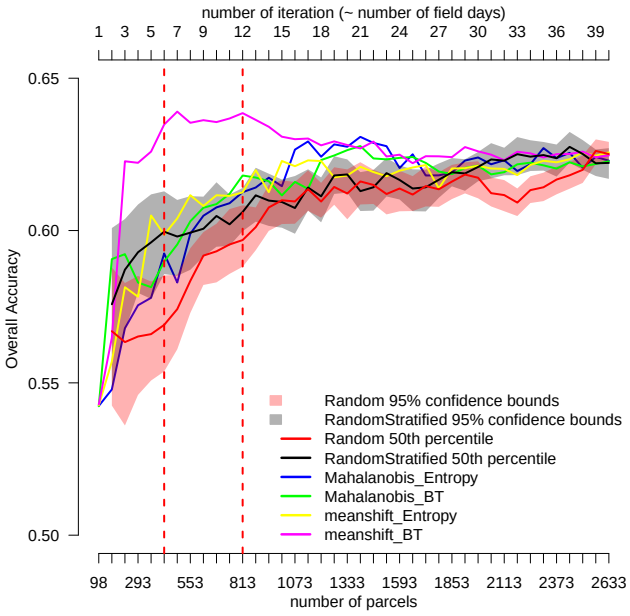

Fig. 6: Comparative curve of metrics OA for Level 4

other AL metrics. Thus, the choice of AL metrics is more crucial with a finer definition of land cover classes.
For the following illustrations, 6^{th} and 12^{nd} iterations were selected as a compromise between a good overall accuracy for both levels and an acceptable amount of field survey days, which correspond almost to 5-day and 11-day field survey, respectively.

B. Land cover maps

For the all metrics, a zoom of prediction maps in 12^{nd} iteration (11-day field survey) are displayed as well as the corresponding ground truth map for L4 land cover in Fig.7.

Fig.7 confirms that parcels are better classified using AL metrics than Random and Stratified Random strategies. Indeed, on corresponding prediction maps Fig.7-(a) and Fig.7-(b) some classes still do not appear even after the model improvement (12^{nd} iteration) such as *Moors and heathland* and *Inland marshes* classes for random strategies. While these classes are well predicted with other AL metrics and especially with the proposed mean-shift-BT metric (Fig.7-(f)).Besides, one comparing prediction maps in Fig.7, one can observe that mean-shift-BT result is the most similar visually to the ground truth map (Fig.7-(h)), which confirms first quantitative results of the comparative curve (Fig.6).

For the best proposed AL metric mean-shift-BT, global prediction maps in 6^{th} and 12^{nd} iterations are displayed as well as the global ground truth map for L4 land cover in Fig.8.

Fig.8 illustrates that the AL classification is improved between the 6^{th} and 12^{nd} iteration for the process, especially for the *Inland marshes* class. Note that, in the 6^{th} iteration,

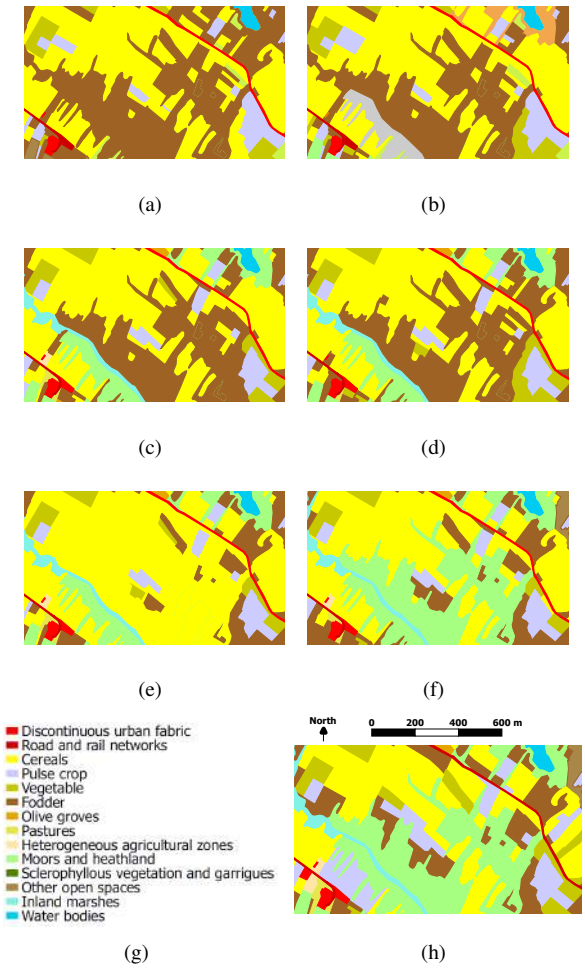

Fig. 7: Zoomed prediction maps for L4 in the 12^{nd} iteration using several methods to select informative parcels for AL. (a)Random, (b)Stratified Random, (c)Mahalanobis-Entropy, (d)Mahalanobis-BT, (e)mean-shift-Entropy, (f)mean-shift-BT. (g)L4 land cover legend and (h)L4 zoomed ground truth map.

samples are always carried out to the majority class *Fodder* which is not the case in the iteration 12.

C. Per-class classification accuracies (producer's accuracy)

For the best AL metric mean-shift-BT, the producer's accuracy in 6^{th} and 12^{nd} iterations are displayed in Table IV.

First, one can observe that some classes have no parcels in the test set (table III). This is due to the fact that our dataset is highly unbalanced and contains only one parcel for those classes. Contrary to other works that eliminate such minor classes [4], we decided to keep the integral of the dataset, since it is an operational study case. Thus, when a unique parcel is available per class, it is used in the training set. For the corresponding classes, no classification accuracy is measured and it was symbolized by "-" in table IV. Table IV shows the classification accuracy measures per class. AL improves the classification accuracy between both iterations for different representative (major) classes (In bold in table IV). The overall accuracy increases by 6%, 5% and 4% for

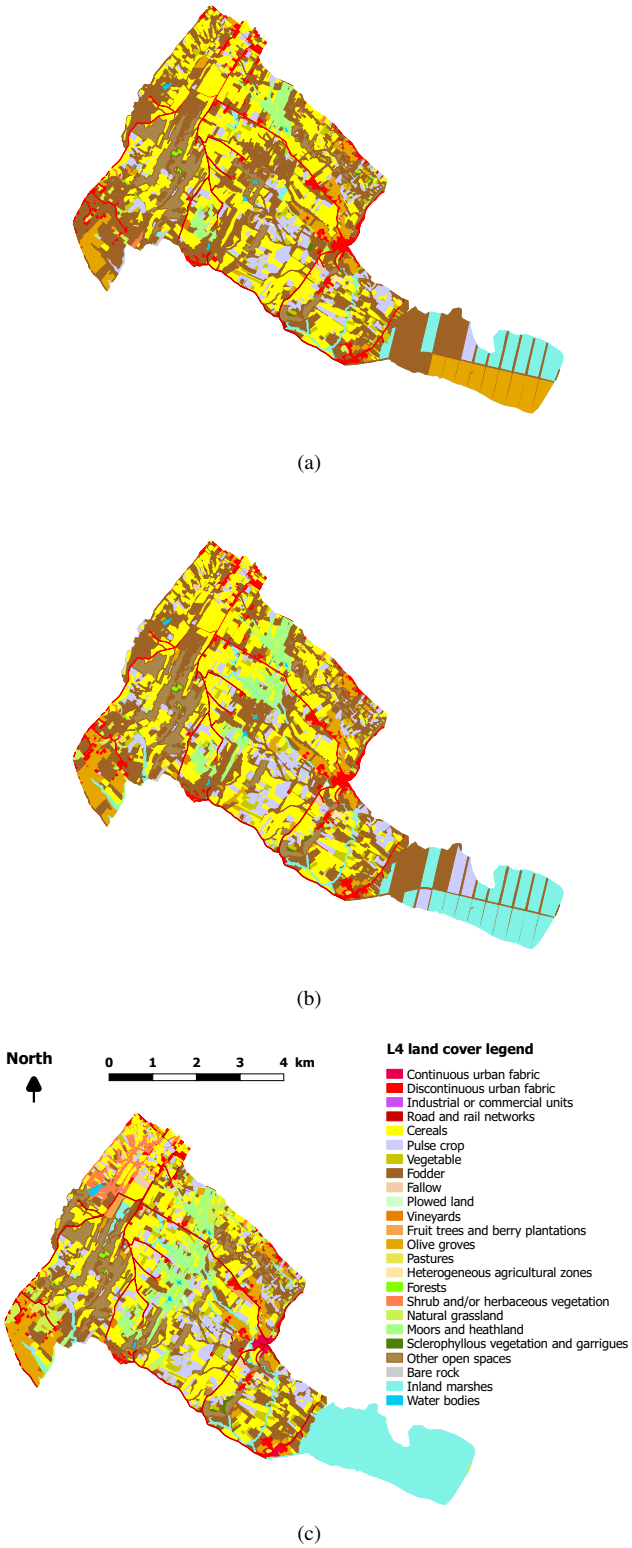

Fig. 8: Global maps using mean-shift-BT metric (a) Prediction map at 6^{th} iteration (5-day field survey), (b) Prediction map at 12^{nd} iteration (11-day field survey) and (c) L4 land cover ground truth map.

Discontinuous urban fabric, *Cereals* and *Vegetable* classes,

TABLE IV: Producer's accuracy for mean-shift-BT metric.

Classes	6^{th} iteration	12^{nd} iteration
Continuous urban fabric	0	0
Discontinuous urban fabric	**0.87**	**0.93**
Industrial or commercial units	-	-
Road and rail networks	0	0
Cereals	**0.86**	**0.91**
Pulse crop	**0.58**	**0.56**
Vegetable	**0.36**	**0.40**
Fodder	**0.62**	**0.62**
Fallow	0	0
Plowed land	0	0
Vineyards	-	-
Fruit trees and berry plantations	0	0
Olive groves	**0.30**	**0.30**
Pastures	0	0
Heterogeneous agricultural areas	0	0
Forests	-	-
Shrub and/or herbaceous vegetation	0	0
Natural grassland	0.12	0.12
Moors and heathland	0	0
Sclerophyllous vegetation and garrigues	0	0
Other open spaces	0	0
Bare rock	-	-
Inland marshes	0	0
Water bodies	-	-
Global OA	**0.634**	**0.638**
Global Kappa	**0.520**	**0.526**

respectively.

However, the table IV also presents some zeros for certain classes. They mainly correspond to minor classes having very few parcels in the test set (unrepresentative classes) such as *Sclerophyllous vegetation and garrigues* class that has only one parcel in the test set. In such cases, if the model does not properly predict the class, the corresponding per-class accuracy will be zero.

This may explain the differences between maps (Fig.8) and precision per class in table IV. Indeed, visually, one can see different prediction improvements for *Inland marshes*, *Road and rail networks* and *Moors and heathland* classes for instance while the corresponding per-class accuracies are equal to zero or do not evolve, respectively between iterations in table IV.

IV. Discussion

A. AL contribution

The proposed parcel-based AL methodology was applied on a Mediterranean cultivated area of 35 km^2 has been classified with an accuracy close to 65%. In a such context, these results are judged acceptable for a land cover map from a monodate image compared to other similar work such as in [4] where the accuracy is about 77% with a finer definition of land cover classes, identifying some crop types and using multi-temporal spot5 images. Indeed, using solely spectral information AL improved the results, in comparison with random sampling, up to 10% for a fine definition of land cover classes i.e. a fine land cover legend (level 4). This improvement is comparable to the one in [10], which integrates spatial information. Results confirm that in the finer definition of land cover classes, the more important is the contribution of AL techniques.

B. AL mean-shift-BT contribution

Besides, for land cover mapping at level 3, all AL metric gave very similar results while the proposed metric mean-shift-BT highly improved the results in comparison with other AL metrics (cf. figure 6). When comparing the results of different metrics (i.e. at level 4 on figure 6), one can note that the proposed metric mean-shift-BT allows to reach the maximum of classification accuracy after 5-day filed survey. More field observations will increase the cost without leading to significant classification improvement. In addition, one can observe that mean-shift diversity metric when combined to different uncertainty measures lead to different results (cf. Figure 6). Indeed, when Combined to BT uncertainty measure, mean-shift clustering give better results than using entropy uncertainty. This can be explained by the fact that BT is based on the difference between the two highest posterior class probabilities while Entropy takes into account all posterior class probabilities. Thus, the latter may be more disturbed by high intra-variability parcels that will lead to many non-zero posterior class probabilities. This results are similar to those in [8], where entropy and BT uncertainty metrics were tested with the SVM classifier on three dataset. Results showed that BT metric is more efficient than entropy one.

C. Operational Strategy

To resume, in the operational agricultural land cover mapping at a fine definition of land cover classes (level 4) and based on a labeling by field surveys i.e very limited training set, active learning techniques are recommended with the proposed mean-shift-BT metric as sampling metric. Our results also showed that 5-day survey may be sufficient to highly improve the results (up to 10%) and that more field observations may not be necessary. In practice, one limitation is the sub-representativeness of classes problem. In fact, our dataset is highly unbalanced and therefore some classes are not represented in the test set. However, this problem is standard in literature especially for agricultural areas and our choice was to keep the whole dataset for validation.

V. Conclusion

The novelties of this work is the application of AL techniques for agricultural land cover mapping, using in-field observations which is rarely done in the literature. We assess the contribution of AL techniques in agricultural context by suggesting an operational parcel-based active learning scheme which is suitable for agricultural land cover mapping at a high and large cultivated areas, using a VHR monodate image contrarily to previous works that generally used hyperspectral data. In addition, the proposed methodology is independent from the classifier, provided that posterior class probabilities can be calculated which also differs from existing works that are dependent of large margin classifier, i.e. SVM classifier. The method is operational since we assume that the labeling

is processed by field survey. Thus, the initial labeled data is very limited (98 parcels for 24 classes). Besides, field survey observations are made at the parcel scale which justifies our strategy choice. Finally the number of AL iterations was correlated to the number of field survey days. The contribution of AL techniques was assessed with comparison to a random and stratified random strategies for sampling new samples. State of the art AL metrics were used and compared. We proposed a new clustering-based methodology based on mean-shift algorithm that doesn't need a prior knowledge of cluster number. This metric is recommended, combined with Breaking ties uncertainty measure, since it improved the classification accuracy results when using finer definition of land cover classes (level 4). Otherwise, at level 3, all AL techniques gave similar results and has not improved random-based AL. Results confirm that in the finer definition of land cover classes, the more important is the contribution of AL techniques.

As future work, an operational in-field strategy will be investigated with regard to two major points: 1) the choice of the initial training samples and 2) the integration of spatial information by including the cost transportation [5] [30] and using in-field point of views.

ACKNOWLEDGMENT

The SPOT-6 image was funded by the French Research Agency in the framework of the program Investissements d'Avenir through the support of the GEOSUD project (ANR-10-EQPX-20)

REFERENCES

[1] F. Colin, R. Moussa, and X. Louchart, "Impact of the spatial arrangement of land management practices on surface runoff for small catchments," *Hydrological Processes*, vol. 26, no. 2, pp. 255–271, 2012.

[2] Bahman Jabbarian Amiri and Kaneyuki Nakane, "Modeling the Linkage Between River Water Quality and Landscape Metrics in the Chugoku District of Japan," *Water Resources Management*, vol. 23, no. 5, pp. 931956, 2009.

[3] Bethany Pratt and Heejun Chang, "Effects of land cover, topography, and built structure on seasonal water quality at multiple spatial scales," *Journal of Hazardous Materials*, vol. 209–210, pp. 48–58, 2012.

[4] Charlotte Pelletier, Silvia Valero, Jordi Inglada, Nicolas Champion, and Grard Dedieu, "Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas," *Remote Sensing of Environment*, vol. 187, pp. 156 – 168, 2016.

[5] B. Demir, L. Minello, and L. Bruzzone, "An effective strategy to reduce the labeling cost in the definition of training sets by active learning," *Geoscience and Remote Sensing Letters, IEEE*, vol. 11, no. 1, pp. 79–83, Jan 2014.

[6] Claudio Persello and Lorenzo Bruzzone, "Active and semisupervised learning for the classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 6937–6956, 2014.

[7] D. Tuia, F. Ratle, F. Pacifici, M.F. Kanevski, and W.J. Emery, "Active learning methods for remote sensing image classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 7, pp. 2218–2232, July 2009.

[8] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 3, pp. 606–617, June 2011.

[9] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 3, pp. 1014–1031, March 2011.

[10] Jielian Guo, Xiong Zhou, Jun Li, Antonio Plaza, and Saurabh Prasad, "Superpixel-based active learning and online feature importance learning for hyperspectral image analysis," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 347–359, 2017.

[11] Ido Dagan and Sean P Engelson, "Committee-based sampling for training probabilistic classifiers," *Proceedings of the Twelfth International Conference on Machine Learning*, pp. 150–157, 1995.

[12] Simon Tong and Daphne Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, pp. 45–66, Mar. 2002.

[13] Pabitra Mitra, B. Uma Shankar, and Sankar K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recognition Letters*, vol. 25, no. 9, pp. 1067 – 1074, 2004.

[14] E. Pasolli and F. Melgani, "Model-based active learning for svm classification of remote sensing images," in *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, July 2010, pp. 820–823.

[15] Ion Muslea, Steven Minton, and Craig A. Knoblock, "Active learning with multiple views," *J. Artif. Int. Res.*, vol. 27, no. 1, pp. 203–233, Oct. 2006.

[16] Leo Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[17] Wei Di and M.M. Crawford, "Multi-view adaptive disagreement based active learning for hyperspectral image classification," in *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, July 2010, pp. 1374–1377.

[18] Andrew McCallum and Kamal Nigam, "Employing em and pool-based active learning for text classification," in *Proceedings of the Fifteenth International Conference on Machine Learning*, San Francisco, CA, USA, 1998, ICML '98, pp. 350–358, Morgan Kaufmann Publishers Inc.

[19] David Cohn, Les Atlas, and Richard Ladner, "Improving generalization with active learning," *Machine learning*, vol. 15, no. 2, pp. 201–221, 1994.

[20] S. Rajan, J. Ghosh, and M.M. Crawford, "An active learning approach to hyperspectral data classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 4, pp. 1231–1242, April 2008.

[21] M. Ferecatu and Nozha Boujemaa, "Interactive remote-sensing image retrieval using active relevance feedback," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 4, pp. 818–826, April 2007.

[22] P Lagacherie, DR Cazemier, P FM van Gaans, and PA Burrough, "Fuzzy k-means clustering of fields in an elementary catchment and extrapolation to a larger area," *Geoderma*, vol. 77, no. 2, pp. 197–216, 1997.

[23] M. Girolami, "Mercer kernel-based clustering in feature space," *Neural Networks, IEEE Transactions on*, vol. 13, no. 3, pp. 780–784, May 2002.

[24] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[25] M. Pal, "Random Forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.

[26] PC Mahalanobis, "Mahalanobis distance," in *Proceedings National Institute of Science of India*, 1936, vol. 49(2), pp. 234–256.

[27] Dorin Comaniciu and Peter Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[28] Yizong Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, Aug 1995.

[29] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008, ISBN 3-900051-07-0.

[30] B. Demir, L. Minello, and L. Bruzzone, "Definition of effective training sets for supervised classification of remote sensing images by a novel cost-sensitive active learning method," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 2, pp. 1272–1284, Feb 2014.

Ines BEN SLIMENE BEN AMOR Ines BEN SLIMENE BEN AMOR is PhD student between UMR LISAH (Supagro, IRD, INRA) Montpellier and RIADI Lab (ENSI) Tunisia in computer science and soil science. She received the Engineering degree in computer science from the National Institute of Implicated Sciences and Technologies (INSAT),Tunis, Tunisia, in 2013. Her project focuses on active learning for remote sensing classification applied in a cultivated environment.

Nesrine CHEHATA Nesrine Chehata received the Diplom. Eng. in image processing from the Ecole Nationale Suprieure de Physique de Strasbourg, 2001 and the M.S degree in Image processing from the university of Strasbourg I, 2001. She defended her PhD Thesis in computer vision and image processing in 2005 at the university of Paris 5 in collaboration with the IGN (French Cartographic Institute) in MATIS Lab and the CNES (French Space Agency). Since 2006, she has been helding a teaching position as associate professor in the EGID institute (Environment and geosciences institute) at the university of Bordeaux. She worked as research fellow at the IRD -UMR LISAH in Tunisia from 2012 to 2014. Since 2014, she is associate professor at ENSEGID -Bordeaux INP (Engineering School of Environment and Geosciences). Her research interests are feature selection, pattern recognition and machine learning applied to remote sensing data such as airborne lidar data and VHR optical satellite images for environmental applications. She is member of the IEEE and the IEEE Remote Sensing Society.

Jean Stephane BAILLY Jean-Stphane Bailly received the Engineering degree in agronomy, the M.Sc. degree in biostatistics, and the Ph.D. degree in hydrology from the University of Montpellier, Montpellier, France, in 1990, 2003, and 2007. He is a Senior Lecturer in physical geography and geostatistics in AgroParisTech, France. His research interests include spatial observations and modeling for hydrological issues. He is a research fellow at the UM LISAH lab.

Imed Riadh FARAH Imed Riadh FARAH is currently Professor at ISAMM Institute, University of Manouba,Tunisia. He received the Engineer degree in 1991 from ENIG School Engineering, MD degree in 1995 from ISG Institute, and the PhD degree in computer science from National School of Computer Science (ENSI), Tunis, Tunisia in 2003. Since November 2009, he has been Associate Professor and then Professor of computer science. He is an Associate Researcher in the Department ITI-Telecom Bretagne, Brest, France, since January 2009. Since 2011, he is head of the Higher Institute of Multimedia Art of Manouba (ISAMM). His research interests are artificial intelligence, data mining, image processing, recognition and interpretation, especially for remote sensing applications, and earth observation.

Philippe LAGACHERIE Philippe Lagacherie trained as an Agronomist and completed his PhD in soil science in 1992. He obtained his senior scientist degree from Montpellier 2 University in 2002 for his researches on digital soil mapping methods. He is currently senior researcher at INRA Montpellier (France). Dr Philippe Lagacherie has been involved in research dealing with Digital Soil Mapping, remote sensing and spatial modeling of cultivated landscapes. He developed spatial approaches that associated field knowledge, GIS, remote sensing, geostatistics, and fuzzy logic.