AN ASSESSMENT OF IMAGE FEATURES AND RANDOM FOREST FOR LAND COVER MAPPING OVER LARGE AREAS USING HIGH RESOLUTION SATELLITE IMAGE TIME SERIES

C. Pelletier, S. Valero, J. Inglada, G. Dedieu

CESBIO - UMR 5126
18 avenue Edouard Belin
31401 Toulouse CEDEX 9 - FRANCE

N. Champion

IGN Espace - MATIS / Université Paris-Est
6 avenue de l'Europe
31521 Ramonville CEDEX - FRANCE

ABSTRACT

New high resolution Satellite Image Time Series (SITS) are becoming crucial to land cover mapping over large areas. Their high temporal resolution will allow to better depict scene dynamics. However, it will also increase the amount of data to process. The classification of these data involves therefore new challenges such as: (1) selecting the best feature set to use as input data, (2) dealing with data variability coming from landscape diversity, and (3) establishing the robustness of existing classifiers over large areas. This work aims at addressing these questions through three different studies. Experimental results are obtained by using SPOT-4 and Landsat-8 SITS.

***Index Terms*—** Land cover mapping, Satellite Image Time Series, High resolution, Classification, Random Forest

1. INTRODUCTION

New Satellite Image Time Series (SITS) will become essential to produce accurate land cover maps over large areas thanks to their high temporal, spectral and spatial resolutions given by satellite missions such as Sentinel-2, Venμs, or Landsat Data Continuity Mission (LDCM).

These time series contain a large amount of information involving new challenges for classification systems. More precisely, standard supervised classification processing chains need to be adapted in order to: (1) provide the classifier with the best input data, which fully exploit the quantity of information given by SITS; (2) deal with the data variability arising from the landscape diversity over large areas; and (3) achieve a good trade-off between classification performances, the classifier robustness and computing time.

Many different features (also named variables) can be used as input data in a classification system. Besides classical spectral and spatial features, temporal features can help to depict the scene dynamics thanks to the high temporal resolution of the new SITS. In a classification scheme, the number and quality of input features are linked to resulting accuracies, but also to computing time.

Robust classification methods, explored in the literature such as Support Vector Machines (SVM) or Random Forest, have shown that their performances are likely to remain unchanged even by adding insignificant features. Therefore, recent classification studies have proposed a huge number of features extracted from original data [1]. In these previous works, features were introduced to enrich data because a single image or very few images were available due to the lack of acquisitions. When working with SITS acquired at high temporal resolution, the contribution of several input features is uncertain. Indeed, the spectral signatures of temporal profiles can be enough to characterize land cover categories.

However, features can help to deal with large variability of the landscapes when working on large areas. Therefore, it becomes interesting to determine the best subset of smaller size among all available features in order to achieve equal accuracies and reduce computational cost.

Regarding the choice of the classifier, the stability of its optimal parameters is an important requirement especially for operational and automated production. Several studies recently demonstrated the ability of Random Forest (RF) [2] to yield accurate land cover maps. RF accomplished performances comparable to traditional classifiers such as decision trees or SVM, with a lower computing time [3, 4, 5].

Hence, this work aims at evaluating the performances and the robustness of the RF classifier on large areas by using different feature sets as input data. For this purpose, three studies are carried out. Firstly, the sensitivity of the RF parameters is analysed. Then, the use of different sets of features as input data in the classification system is studied. Finally, the classifier stability is tested on a larger area.

The paper is organised as follows: the experimental set-up is detailed in Section 2, Section 3 is devoted to experimental results, and conclusions are drawn in Section 4.

This work is funded by the French spatial agency (CNES) and the French mapping agency (IGN).

2. EXPERIMENTAL SET-UP

2.1. Available data

The study area is located in the southwest of France. Two data sets are used.

The first data set is a SITS of 15 SPOT-4 (Take-5 experiment [6]) (February - June 2013) and 8 Landsat-8 (April - December 2013) (resampled at 20 m.) images. It covers an area of around 17,000 km^2. Reference data contain 18 thematic categories. They were collected during fieldwork campaigns, and they also come from the French National Land Cover database produced by the French mapping agency (IGN).

The second data set is a SITS made of 15 Landsat-8 images at 30 m. spatial resolution. It covers around 20,000 km^2. Reference data are composed of 12 categories, which come from the French Land Parcel Information System (*Registre Parcellaire Graphique* in French) and the French National Land Cover database.

Note that SPOT-4 images have four spectral bands (green, red, near infra-red and shortwave infra-red) at 20 m. spatial resolution. Landsat-8 images have seven spectral bands (coastal aerosol, blue, green, red, near infra-red and two shortwave infra-red bands) at 30 m. spatial resolution.

2.2. Input feature data sets

The contribution of features will be evaluated by incorporating spectral and temporal features in addition to spectral bands (SB) in the classification system. Spectral features (SF) correspond to vegetation indices such as NDVI (Normalized Difference Vegetation Index), built-up indices, water indices and tasseled cap features [7]. Temporal features (TF) correspond to statistical values (mean, variance, maximum, *etc.*), and phenological features computed by approximating the NDVI profile with a double logistic function as [8]. A total of five feature sets are studied:

1. spectral bands only (SB),
2. spectral bands and spectral features (SB-SF),
3. spectral bands and NDVI (SB-NDVI),
4. spectral bands and temporal features (SB-TF),
5. spectral bands, NDVI and temporal features (SB-NDVI-TF).

2.3. Random Forest classifier

The RF classifier is an ensemble learning method that consists in learning several weak classifiers (decision trees) to generate a classifier with a strong decision rule. The use of RF requires tuning four parameters: (1) K, the number of trees; (2) m, the number of features randomly selected at each node; (3) max_depth, the maximal depth of each tree; and (4) $min\_samples$, the minimal number of samples per node.

These parameters depend on the input data (number and quality) to classify, and they can affect the classifier performances [4, 5, 9]. Recent studies have established that K and m are the two main parameters affecting classifier performances [10]. Accordingly, only the influence of K and m is studied here.

3. EXPERIMENTAL RESULTS

To attest the effectiveness of RF with high resolution SITS over large areas, three tests have been carried out. Firstly, the sensitivity of RF parameters is evaluated for two different input data configurations. Secondly, the five feature sets, described in Section 2.2, are tested as input data on the classification system. Finally, the assessment of the previous results is performed on a larger area. Overall Accuracy (OA) and Kappa coefficient metrics are computed from confusion matrices to evaluate the classification performances [11].

3.1. Random Forest parameter sensitivity

Two subsets containing half of the reference data polygons are generated from the first data set (described in Section 2.1). The first subset is used for validation: all samples contained in polygons are kept, it represents a total of 450,000 samples. The second subset is used for training. It contains 5,000 samples per category randomly selected. Five random trials are carried out to evaluate the results statistically (see Section 3.2).

The influence of RF parameters on classification performances is evaluated on the five feature sets presented in Section 2.2. For the sake of brevity, only the results obtained for the simplest case (spectral bands only) and the best set of features (spectral bands with spectral features) are shown. The classification of both input data sets is performed by using different K and m values. max_depth and $min\_samples$ are set to 25. Tables 1 and 2 display the results of this evaluation.

In both tables, the difference between the minimum and the maximum Kappa is around 3%. General trends emerge from the two tables: increasing the number of trees leads to a slight rise of Kappa, whereas a too high m value causes a decline.

Tables 1 and 2 also show that it is preferable to select a K value equal to or higher than 100. As the computational cost increases linearly with the number of trees, K is set to 100. $m = 10$ and $m = 17$ provide the best results for SB and SB-SF respectively. These values are entirely consistent with the literature since they are close to the square root of the total number of features composing the input data [12].

K \ m	2	10	38	58	116
50	0.7243	0.7275	0.7265	0.7219	0.7006
100	0.7243	0.7324	0.7289	0.7262	0.7058
150	0.7283	0.7312	0.7316	0.7287	0.7095
200	0.7290	0.7327	0.7314	0.7286	0.7074
400	0.7286	**0.7348**	0.7316	0.7329	0.7128

Table 1. Values obtained for Kappa by using different Random Forest parameter configurations. The input data set is only composed of spectral bands (SB), with a total of 116 input features. K is the number of trees, and m the number of features randomly selected at each node.

K \ m	2	17	100	151	302
50	0.7376	0.7470	0.7461	0.7379	0.7219
100	0.7454	0.7533	0.7482	0.7442	0.7253
150	0.7460	0.7532	0.7482	0.7465	0.7255
200	0.7457	**0.7536**	0.7485	0.7467	0.7256
400	0.7478	0.7535	0.7492	0.7490	0.7291

Table 2. Values obtained for Kappa by using different Random Forest parameter configurations. The input data set is composed of spectral bands and spectral features (SB-SF), with a total of 302 input features. K is the number of trees, and m the number of features randomly selected at each node.

3.2. Comparison of input feature sets

The impact of the five features sets on classification accuracies is also evaluated with the first data set. Table 3 displays OA with 95% confidence intervals for the five feature sets (described in Section 2.2). The values of RF parameters are optimized for each feature data set.

Although SB-SF obtain the best score, the resulting OA is only higher of about 1%. In addition, confidence intervals obtained by the other feature data sets overlap SB-SF confidence interval.

Results also show that the proposed temporal features fail to bring additional relevant information to spectral bands. Related temporal information may be contained in the spectral signature of SITS. In addition, these features are complex and increase the computing time.

These results suggest that the examined spectral and temporal features are redundant with high resolution SITS. However, feature contribution may be more important over large areas where landscapes change.

3.3. Random Forest stability over large areas

The goal here is to assess the classifier robustness over large areas. More precisely, the classifier performances will be evaluated when validation areas are moving away from the training area. For this purpose, the second data set is used (Section 2.1). The training area is spatially localised on a circular area of 15 km. radius. This area contains 5,000 training samples per category randomly selected. 35,000 validation samples per category are randomly selected outside the training area on all the study area (in grey in Figure 1). Nineteen sub-areas are defined for local validations (in blue in Figure 1). The SB, SB-SF and SB-NDVI data sets are studied here. In the following, $K = 100$, $m = \sqrt{p}$, $max\_depth = 25$, and $min\_samples = 25$.

Feature set	OA	K, m, max_depth, min_samples
SB	82.10 ± 2.29 %	400,10,50,10
SB-SF	**83.27 ± 2.47 %**	400,2,25,10
SB-NDVI	82.21 ± 2.53 %	400,11,50,10
SB-TF	82.05 ± 2.52 %	400,11,25,10
SB-NDVI-TF	82.10 ± 2.63 %	400,12,50,10

Table 3. Overall Accuracies (OA) with 95% confidence intervals for different feature sets with optimized Random Forest parameters. (SB: spectral bands, SF: spectral features, NDVI: normalized difference vegetation index, TF: temporal features. K: number of trees, m: number of features randomly selected at each node, max_depth: maximal depth of each tree, $min\_samples$: minimal size of terminal nodes.)

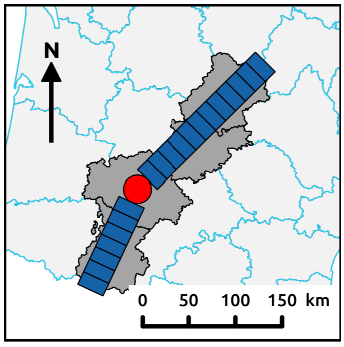

Fig. 1. Study area in grey with training area in red and validation areas in blue.

Figure 2 displays the Kappa coefficient evaluated in all the sub-areas for the three feature sets (SB, SB-SF, and SB-NDVI). The horizontal axis represents the distance (in km.) between the evaluation area and the training area. Positive (resp. negative) distances represent sub-areas located at the northeast (resp. southwest) of the training area.

The Kappa values are not significantly different between the different feature sets. Adding spectral features does not increase classification accuracies even on regions far from the training area. For each feature set, the Kappa coefficient remains mainly stable on regions closer to the training area because of the similarity between landscapes. In contrast, Kappa values fall dramatically for the uppermost negative dis-

tances. It is due to the presence of the Pyrénées mountains whose land cover categories differ from those of the training area. The use of only spectral bands lead to good results in terms of accuracy and computing time.

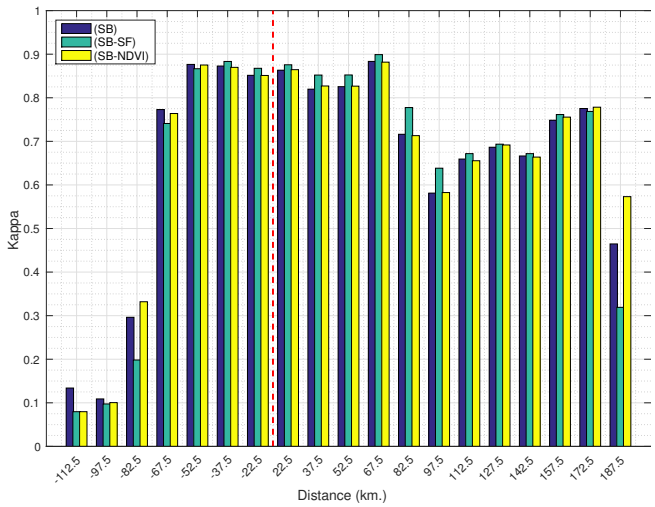

Fig. 2. Values obtained for Kappa for three feature sets with respect to the training area distance. The vertical dashed red line shows training area location. (SB: spectral bands, SF: spectral features, NDVI: normalized difference vegetation index)

4. CONCLUSION

Some of the main challenges of land cover mapping over large areas using high resolution SITS have been discussed. Our results show that RF can handle classification with only the information contained on new high resolution SITS.

Spectral and temporal features have been proposed in order to characterize scene dynamics. Although they improve the performances of classification, the gain in accuracy is weak compared with the increase in the computational cost. Therefore, the use of only spectral bands is a good trade-off between accuracy and computing time. Furthermore, it has been shown that the setting of RF parameters causes little influence on the classification accuracy.

Spectral features have also been analysed when the training area is spatially localised. The study reveals that the classifier can achieve good performances, regardless of the input feature data sets, as long as the landscapes remain similar to those of the training area.

Future works will be carried out with Venμs and Sentinel-2 images (10 m.). They will analyse the contribution of spatial features and new spectral indices based for instance on the red edge channel.

5. REFERENCES

[1] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 257–272, 2013.

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

[3] J. Inglada, M. Arias, B. Tardy, O. Hagolle, S. Valero, D. Morin, G. Dedieu, G. Sepulcre, S. Bontemps, P. Defourny, et al., "Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery," *Remote Sensing*, vol. 7, no. 9, pp. 12356–12379, 2015.

[4] V.F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J.P. Rigol-Sanchez, "An assessment of the effectiveness of a Random Forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93 – 104, 2012.

[5] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random Forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006.

[6] O. Hagolle, S. Sylvander, M. Huc, M. Claverie, D. Clesse, C. Dechoz, V. Lonjou, and V. Poulain, "SPOT-4 (Take 5): simulation of Sentinel-2 time series on 45 large sites," *Remote Sensing*, vol. 7, no. 9, pp. 12242–12264, 2015.

[7] R. J. Kauth and G. S. Thomas, "The tasselled cap - a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat," in *LARS Symposia*, 1976, p. 159.

[8] P. SA. Beck, C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore, "Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI," *Remote Sensing of Environment*, vol. 100, no. 3, pp. 321–334, 2006.

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

[10] A-L. Boulesteix, S. Janitza, J. Kruppa, and I. R. König, "Overview of Random Forest methodology and practical guidance with emphasis on computational biology and bioinformatics," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 6, pp. 493–507, 2012.

[11] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sensing of Environment*, vol. 37, no. 1, pp. 35–46, 1991.

[12] A. Liaw and M. Wiener, "Classification and regression by Random Forest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.