

CLASSIFICATION OF VEGETATION CLASSES BY USING TIME SERIES OF SENTINEL-2 IMAGES FOR LARGE SCALE MAPPING IN CAMEROON

Hermann Tagne^{1,2,3,*}, Arnaud Le Bris¹, David Monkam², Clément Mallet¹

¹ LASTIG, Univ Gustave Eiffel, IGN, ENSG, 73 avenue de Paris, F-94160 Saint-Mandé, FRANCE - firstname.lastname@ign.fr

² University of Douala, Faculty of Science, Department of Physics, P.O. Box 24157 Douala, CAMEROON - monkam1@yahoo.fr

³ National Institute of Cartography, P.O. Box 157 Yaoundé, CAMEROON

KEY WORDS: Cameroon; large-scale; classification; fusion; land-cover; vegetation; mapping; Satellite; time series; Sentinel-2.

ABSTRACT:

Sentinel-2 satellites provide dense image time series exhibiting high spectral, spatial and temporal resolutions. These images are in particular of utter interest for Land-Cover (LC) mapping at large scales. LC maps can now be computed on a yearly basis at the scale of a country with efficient supervised classifiers, assuming suitable training data are available. However, the efficient exploitation of large amount of Sentinel-2 imagery still remain challenging on unexplored areas where state-of-the-art classifiers are prone to fail. This paper focuses on Land-Cover mapping over Cameroon for the purpose of updating the Very High Resolution national topographic geodatabase. The ι^2 framework is adopted and tested for the specificity of the country. Here, experiments focus on generic vegetation classes (five) which enables providing robust focusing masks for higher resolution classifications. Two strategies are compared: (i) a LC map is calculated out of a year long time series and (ii) monthly LC maps are generated and merged into a single yearly map. Satisfactory accuracy scores are obtained (>94% in Overall Accuracy), allowing to provide a first step towards finer-grained map retrieval.

1. INTRODUCTION

The increasing needs in food supply in the near future will require higher agricultural yields (Foley et al., 2011). Other factors increasing the pressure on agricultural lands are urban sprawl and densification as well as the production of bio-fuels (Godfray et al., 2010, Rounsevell et al., 2005, Searchinger et al., 2008). These pressures will also have negative consequences on natural ecosystems (Green et al., 2005, Tilman, 2001). Indeed, agricultural activities are a major cause of ecosystem degradation at the global scale (Benayas, Bullock, 2012), and therefore, land use change monitoring related to farming is crucial for sustainable land management (Schwilch et al., 2010). Vegetation classes extent estimates and vegetation classes mapping provide crucial information for vegetation classes monitoring and management. Remote sensing imagery in general and, in particular, high spectral, spatial and temporal resolution data as the ones which are available with spaceborne optical systems such as Sentinel-2, are suitable for this kind of application.

The current Sentinel-2 for Vegetation Class Mapping project at the National Cameroonian Mapping Agency aims at fully exploiting the Sentinel-2 observational capabilities and its offered possibilities for vegetation monitoring through the development of open source processing chains such as the ι^2 chain, capable of large-scale land-cover map production (Inglada et al., 2017). Three Earth Observation products – dynamic vegetation class masks, vegetation type, and vegetation class status – have been identified for this processing chain and will be demonstrated at the national scale.

The national Cameroonian mapping agency aims at implementing such a fully operational automatic Vegetation Class map production system using both time series of Sentinel-2 optical images and available Very High Resolution (VHR) aerial images. So far, the vegetation classes are established on the

basis of the visual interpretation of the latter images complemented with field surveys. The nomenclature of the vegetation classes is not yet exhaustive (5 classes). It is targeted at investigating how Sentinel-2 and aerial images can be jointly beneficial for this task since they exhibit complementary strengths. This paper focuses on the first step of this exploratory workflow: the automatic analysis of time series of Sentinel-2 images for yearly LC map generation. It aims at assessing the relevance of a state-of-the-art operational classifier to this specific context (Mallet, Le Bris, 2020). Indeed, a first difficulty lies in the lack of reference data at the spatial resolution of Sentinel images for the training of a supervised classifier. Most datasets exhibit higher spatial resolution which leads often to spatial discrepancies between both sources, detrimental for the learning step. The proposed solution consists in a direct use of existing Very High Resolution topographic database within the existing ι^2 LC classification framework. Another challenge consists in working on areas with numerous cloudy days and significant intra-annual variability (Mertens, Lambin, 2000). In this paper, we assess various solutions which consists in performing either direct yearly classification of time series of images or first classifying monthly datasets followed by a fusion decision step. Section 2 introduces the area of interest and the datasets. Section 3 presents the adopted workflow and the experimental setup, while Section 4 provides both a quantitative and qualitative assessment of the results. Conclusions are drawn in Section 5.

2. DATA AND STUDY AREA

In order to design a reliable fully operational automatic Cameroon LC map system out of Sentinel-2 time series, experiments are conducted over a study area in the Extreme-North (EN) of Cameroon.

2.1 Study area

Current experiments and analyses are limited to one specific site, located over Maroua³, Mora and Petté in the EN of

* Corresponding author

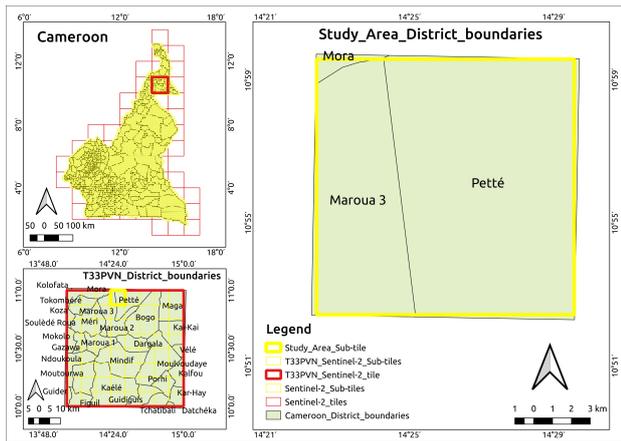


Figure 1. Location of the study area in Cameroon.

Cameroon, (see Figure 1). Covering 189.072 km², it was selected both for its challenging conditions and the availability of reference and Copernicus data. It has a short rainy season from June to September and a long dry season from October to May. The slope in the whole scene greatly varies with some hilly areas and rocks. The soil is sandy in the dunes with sandy clay at the median part levels, and essentially clayey at the shallow levels. The drainage is almost nonexistent. The essential waterways, Mayel Tchilé and Mayel Petté, are loosing their activities in September. The principal LC categories are vegetation (principally accacias, balanites, and goods of thorns), water areas, buildings and roads.

2.2 Data

2.2.1 Remote sensing data Sentinel-2 is the only image data source used in this work: 10 spectral bands from 10 to 60 m, (see Table 1). All Levels 2A and 2B data available through the Copernicus Open Access Hub¹ over the study area for the year 2018 were used. For each month, 4-7 images were available (Table 1). More specifically, a focus was made on a sub-area of 10 km × 10 km.

2.2.2 Reference data The proposed approach relies on existing geodatabases available over the study area to build the reference datasets, needed for the supervised classification of the LC maps. The topographic LC database of the Cameroonian National Cartographic Institute (BDINC, simplified for our purpose) was the only one data source used in order to build a unique reference dataset, without a tedious harmonization step. In further experiments, five generic vegetation classes are considered for the classification: *Tree Savannah* (TS), *Shrub Savannah* (SS), *Crop Area without Trees* (CA), *Crop Area with Trees* (CAT), and *Floodable Crop Area* (FCA).

3. PROPOSED WORKFLOW AND EXPERIMENTS

3.1 ι^2 framework

Land-cover map generation is cast as a supervised classification task as routinely adopted in the literature. The ι^2 or ι^2 versatile framework (Inglada et al., 2017) is adopted for generating land-cover maps from generic remote sensing and reference data source. Indeed, it is specifically dedicated to process Landsat or Sentinel time series and it has proven to be an efficient and highly parametrizable open-source solution. ι^2 has

¹ <https://scihub.copernicus.eu>

already demonstrated its high efficiency in performing country wide LC classification out of multi-temporal high resolution Landsat and Sentinel imagery at pixel level (David et al., 2021). It can run either on standard desktop computers or High Performance Computing clusters.

A per-pixel machine learning classification scheme is retained, on which simple post-processing steps can be plugged. Contextual classifiers or object-based approaches (Derksen et al., 2018) would have been interesting but have been discarded yet.

3.2 Random Forest

A Random Forest (RF) classifier is used (Breiman, 2001). This classifier is an aggregation of a set of decision CART trees. Indeed, to improve the performance of single CART classifier, RF classifier is based on two main principles: *bagging* and *randomsubspace*. A fusion of the different basic classifiers is carried out by majority vote: each decision tree predicts a label for a new sample to be classified and the label finally assigned to it is the one which receives the greatest number of votes.

This study uses the RF algorithm as base classifier as it has proven to be efficient for land-cover discrimination at large scales under noisy labels, both in terms of computation times and classification quality (Rodriguez-Galiano et al., 2012, Pelletier et al., 2016). Several parameters come into play during the construction of a RF classifier: the number K of trees (arbitrarily set to a large value), the number m of attributes drawn during each cutting of the node of one of the trees (usually set to the value of the square root of the number of attributes), the maximal depth *max_depth* of each tree, and the minimal number of samples *min_samples* per node.

3.3 Process workflow

The proposed workflow can be decomposed into the next steps, detailed afterwards:

1. Data preparation: convert all images bands and create masks of all images;
2. Feature computation: compute standard features on images;
3. Training: select training samples from all images and reference ground truth maps and build classification models;
4. Prediction: LC map production;
5. Classification fusion: fuse all monthly classifications.

3.3.1 Data preparation This step (Figure 2) consists in adapting the Copernicus Sentinel-2 data to fit the input format requirements of ι^2 . First, images bands are converted from *.jpg2* to *.tif* format using GDAL *translate* functionalities. Then, for each Sentinel-2 image, three different masks are computed in a manual way using the *r.mask.rast* command of the *Raster(r.*)* library of GRASS in QGIS. They correspond to unusable image areas because of clouds (*CLM_R1**), saturation (*SAT_R1**) or diverse reasons (*EDG_R1**).

3.3.2 Feature computation After data preparation, the feature vector used for the classification of each pixel has to be built. For each Sentinel-2 image, four standard features are computed out of the original bands:

- Normalized Difference Vegetation Index (NDVI) (Tucker, 1979) to highlight vegetation cover;
- Normalized Difference Water Index (NDWI) (Gao, 1996) to enhance water and wet areas;

Table 1. *Iota2* (i^2) configuration for monthly and yearly classifications over the study area.

Sensor	Months	I	IS	TR	B	SpeR	SpaR	F	M	
S2 L2A	January	6	1368×1385	5 days	B2 (Blue)	0.46-0.52	10	SB	60	
	February	4			B3 (Green)	0.54-0.58	10		40	
	March	6			B4 (Red)	0.65-0.68	10	SB-NDVI	66	
	April	6			B5 (Red-edge 1)	0.70-0.71	20			
	May	6			B6 (Red-edge 2)	0.73-0.74	20	SB-NDWI	66	
	June	6			B7 (Red-edge 3)	0.77-0.79	20			
	July	6			B8 (Near infra-red 1)	0.78-0.90	10	SB-SAVI	56	
	August	5			B8A (Near infra-red 2)	0.85-0.87	20			
	September	6			B11 (Water vapor)	0.93-0.95	60	SB-NDRI	66	
	October	7			B12 (Shortwave infra-red 1)	1.57-1.66	20			76
	November	6								66
	December	6								

S2 L2A: Sentinel-2 Level-2A; **I:** total number of images per month; **IS:** images size (in pixel); **TR:** temporal resolution; **B:** spectral bands; **SpeR:** spectral resolutions (in μm); **SpaR:** spatial resolutions (in m); **F:** features variables; **SB:** spectral band; **NDVI:** normalized difference vegetation index; **NDWI:** normalized difference water index; **SAVI:** soil-adjusted vegetation index; **NDRI:** normalized difference ratio index; **M:** total number of features depending on the month (for January here).

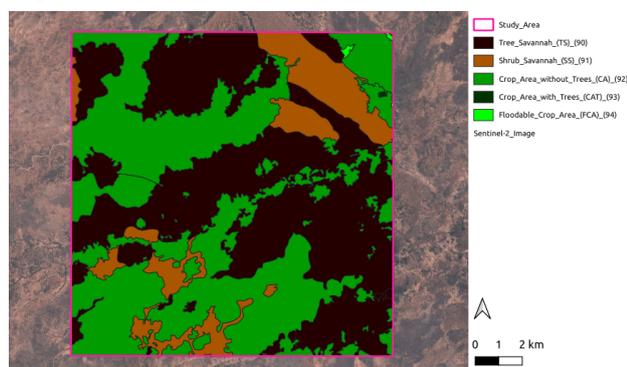


Figure 2. Reference data on the training study area.

- Soil Adjusted Vegetation Index (SAVI);
- Normalised Difference Red Edge (NDRE).

Selecting *batchProcessing* parameter in i^2 configuration file improves feature computation times. At the end, each pixel is characterised by the original spectral bands (limited to the 10 bands exhibiting a native 10 or 20 m GSD) and the spectral indices corresponding to the different dates of the time series. For a yearly time series, it amounts to 760 features: 10 spectral bands plus 4 indices on 70 dates (Table 1). For monthly and yearly classifications, all features are stacked into a single set.

3.3.3 Training the classifier As the proposed approach relies on a supervised RF classifier, training is an important step of the workflow. It consists in (i) sampling data to obtain a training set and (ii) building a classification model per image that will be applied in the next step (Section 3.3.4).

Training set design. Five generic vegetation classes *Tree Savanna* (TS), *Shrub Savanna* (SS), *Crop Area without Trees* (CA), *Crop Area with Trees* (CAT) and *Floodable Crop Area* (FCA) are considered in these experiments.

Training samples are extracted for these classes from the reference dataset (Section 2.2.2), *i.e.*, the existing Very High Resolution national topographic database. Using existing databases as training data has proved to be a suitable solution (Gressin et al., 2014, Postadjian et al., 2017), even if noisy labels may appear due to misregistration, changes and geodatabase specifications. Pixels are randomly taken as a 20% subset of each class, keeping the same ratio between initial reference and training sets. In practice, a maximum sampling of 1,051,919 pixels is

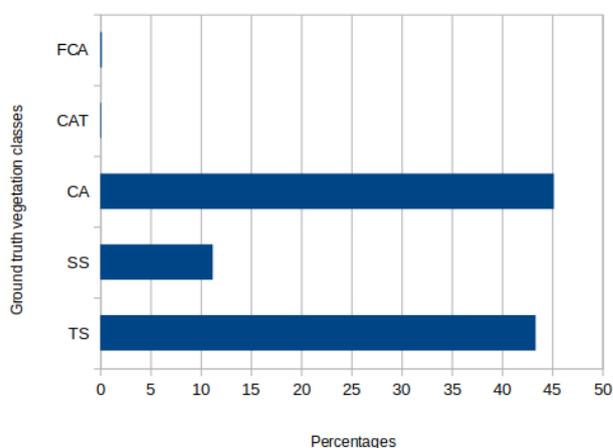


Figure 3. Percentages of vegetation classes in the reference data over the study area.

performed over the area covered by the (month or year long) Sentinel-2 time series. The number of training samples selected for each class is proportional to the area covered by this class over the useful parts of the image. These areas are different (see Figure 4) depending on classes and so the class samples are unbalanced.

Building the RF model. In supervised classification, the balance between class samples is important. Data augmentation is used here, generating synthetic samples with *jitter* (strategy jitter standard factor of 10) and *smote* (strategy smote neighbors of 5) methods by using the *minNumber* samples strategy to set the minimum number of samples by class required. RF is not very sensitive to the parameter choice (Pelletier et al., 2016). Thus, it is here applied with the standard i^2 configuration (see Table 1): a number K of 1,000 trees, a number m of features randomly selected at each node equals to the square root of the total number of features, a maximum depth *max_depth* for each tree of 25 levels and a minimum number *min_samples* of 5 samples per node have been used. RF was training with the polygons of classes in the study area sub-tile and testing on the entire T33PVN tile.

3.3.4 Prediction This step simply consists in classifying Sentinel-2 data according to model obtained at the previous step. Here, a LC map was generated for year 2018 at a spatial resolution of 10 m over the areas of interest for the five generic

vegetation classes listed above.

3.3.5 Classification fusion This last step is specific to the case when several LC maps are calculated for different time series configurations. For instance, in the next experiments, the year is split into epochs (months), and one map is computed for each epoch (ranging from four to seven Sentinel-2 images). For each per epoch classification, each pixel contains a value corresponding to its class. To synthesize this information to a unique map, one can rely on the Fusion by Voting strategy (Le Bris et al., 2019). It consists in merging classifications by letting each of them vote for its label and choosing the final label as the winner of this majority vote. Such vote can be applied to hard label classification results, but it can also exploit confidence information associated to these maps. This strategy is here applied to the twelve monthly hard label classifications (assuming a perfect coregistration).

3.4 Experiment set up

This work aims at demonstrating the ability of state-of-the-art RF classifier to classify time series of Sentinel-2 for large scale land cover mapping in the specific Cameroonian context. Several strategies are considered and compared. On one hand, a LC map is calculated out of the complete year long time series for 2018. On the other hand, the year is split into epochs (months), and one map is computed for each epoch. The quality of these twelve monthly classifications is then assessed to identify whether some periods are more prone to deliver good results. At the end, these monthly classifications can be merged into a unique map. One interest of such per month strategy compared to year-long time series analysis states in its ability to handle smaller amount of data for each classification.

3.5 Quality assessment

Results are all assessed quantitatively and qualitatively.

3.5.1 Qualitative assessment A qualitative and visual assessment of obtained results is done. Indeed, it may sometimes be useful to assess the quality of the classifications locally (here in a given yellow squared area of the study area (Figure 5)). For this aim, a classification indicating the confidence of the RF classifier on the decision for each pixel of the classification can be provided. Finally, a visual evaluation in order to highlight anomalous characteristics will also be presented.

3.5.2 Quantitative evaluation Classification results are compared to the reference data (section 2.2.2). Then, several classic metrics are derived from the confusion matrix : Overall Accuracy (OA), Kappa coefficient (K) and F-Score averaged (Cohen, 1960), (Fleiss, Cohen, 1973) for global assessment, as well as F-Score, producer's and user's accuracies for per class analysis.

4. RESULTS AND DISCUSSION

The different classifications described in Section 3 were calculated over the study area for the year 2018 and evaluated as explained in Section 3.5. A visual assessment of a yellow squared particular area displayed was also proposed to highlight some phenomena which can not be detected with the metrics.

4.1 Quantitative evaluation

Figure 4 presents the mean F-Score and Kappa coefficients as global metrics as well as the per-class F-Score, recall and precision quality metrics for the two classification modes (monthly

and yearly) and for the fusion scheme. These metrics were calculated over one run using the same training and validation sets (of 1,051,919 pixels each). Kappa coefficients are very good with values above 0.81 (0.90-0.99 for monthly classifications, 0.99 for the yearly one and 0.97 for the fusion). Table 15 shows the confusion matrix for the fusion of monthly classifications. It can be highlighted that the Overall Accuracy and the Kappa coefficient values are 98.10% and 0.97, respectively. One can observe that the May and yearly classifications yield the same better results, followed by the September and November ones, than the other monthly classifications and the fusion of the twelve monthly classifications for both metrics. The fusion classification results are also better than the January, July, October and December ones, which validates the necessity of the yearly synoptic view of Sentinel time series.

When considering per-class metrics, the F-Score values are very good for the classes *Tree Savannah* (TS), *Shrub Savannah* (SS) and *Crop Area without Trees* (CA) for all the classifications modes. The results are just good for the class *Crop Area with Trees* (CAT) for January, July, October and December classifications. The results are moderate with the *Floodable Crop Area* (FCA) for October and December, respectively, mainly due to the limited training samples of these two last vegetation classes (see Figure 3). It can also be shown from Figure 4 that the improvement for these minority classes (CAT and FCA) mostly states in the precision value, *i.e.*, underdetection is reduced.

Tables 2, 5, 7, 11, and 13 show the confusion matrices for the January, April, June, October and December classifications, respectively. They allow a detailed analysis of the errors for monthly classifications. It can be highlighted that:

- the 24.98% and 26.37% of the class *Crop Area without Trees* (CA) classified as *Crop Area with Trees* (CAT) and *Floodable Crop Area* (FCA) classes in January, are reduced to 13.11% and 16.91% respectively in April;
- in June, the confusion of *Crop Area without Trees* (CA) in *Floodable Crop Area* (FCA) is very much reduced (divided by 5). However, *Crop Area without Trees* (CA) in *Crop Area with Trees* (CAT) is reduced from 24.98% to 13.81%, but *Tree Savannah* (TS) in *Floodable Crop Area* (FCA) increases from 8.53% to 12.30%;
- in July, *Crop Area without Trees* (CA) increases from 24.98% to 27.84% in *Crop Area with Trees* (CAT), but decreases very much from 26.37% to 14.25% in *Floodable Crop Area* (FCA);
- similarly, in October, *Crop Area without Trees* (CA) increases much more from 24.98% to 30.77% in *Crop Area with Trees* (CAT), but also decreases much more from 26.37% to 10.45% in *Floodable Crop Area* (FCA); however the two other main confusions of *Tree Savannah* (TS) with *Shrub Savannah* (SS) and *Floodable Crop Area* (FCA) respectively, are multiplied by 5;
- the confusion between *Crop Area with Trees* (CAT) and *Crop Area without Trees* (CA) is reduced from 30.77% to 20.79%; however the confusion of FCA classified as *Crop Area without Trees* (CA) increases from 10.45% to 13.59% in December;
- always in December, the confusion between *Shrub Savannah* (SS) and *Tree Savannah* (TS) is reduced to less than 4%; however the confusion between *Floodable Crop Area* (FCA) and *Tree Savannah* (TS) is very much reduced from 44.62% to 18.61%, but the confusion between CAT and TS increases from 7.60% to 11.02%.

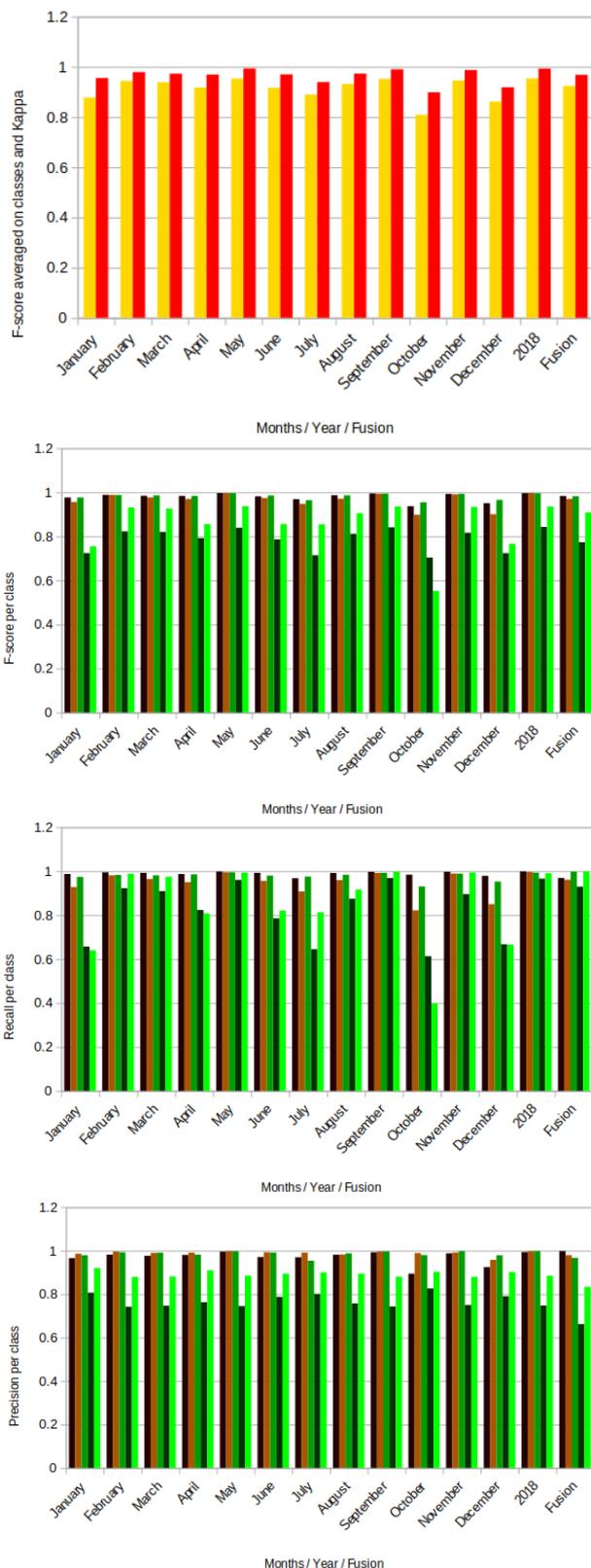


Figure 4. Global (F-Score averaged on vegetation classes ■ and Kappa ■) and per vegetation class (F-Score, recall and precision) quality metrics for the study area for months and year and for fusion of monthly classifications. ■: TS; ■: SS; ■: CA; ■: CAT; ■: FCA.

Table 14 shows the confusion matrix for the full year 2018 classification. Several aspects of this improvement can be highlighted:

- the confusion between *Crop Area with Trees* (CAT) and *Crop Area without Trees* (CA) is considerably reduced to only 1.18%;
- the confusion between *Floodable Crop Area* (FCA) and *Crop Area without Trees* (CA) decreases drastically to 0.18%;
- the confusion of *Shrub Savannah* (SS), *Crop Area with Trees* (CAT) and *Floodable Crop Area* (FCA) with *Tree Savannah* (TS) decreases to 0.20%, 1.39% and 0.40% respectively.

In parallel, Table 15 shows the confusion matrix for the fusion of the twelve monthly classifications. One can note that:

- the confusion of *Crop Area with Trees* (CAT) and *Floodable Crop Area* (FCA) with *Crop Area without Trees* (CA) decreases to 0.42% and 0.20% respectively;
- the confusion of *Shrub Savannah* (SS) and *Crop Area with Trees* (CAT) with *Tree Savannah* (TS) decreases to 0.25% and 0.42% respectively;
- no confusion between *Floodable Crop Area* (FCA) and *Tree Savannah* (TS) in both directions.

4.2 Classification confidence

The RF classifier associates a confidence (unsupervised margin) calculated out of the distribution of the labels predicted by the trees (Frenay, Verleysen, 2014) of the forest. Figures 5 show the predicted yearly LC map and the yellow squared areas with its corresponding confidence map. One can observe that the confidence is influenced by the proportion occupied by the vegetation class samples in the reference data (see Figure 3). For example, the *Crop Area without Trees* (CA) class in that specific yellow zone is easily recognised, with high confidence. Besides, the *Tree Savannah* (TS) class generally exhibits high confidence, which is encouraging for a use of such information. It is useful to study how confidence values are related to correct and incorrect classifications. For that, the distribution of confidence values was calculated for only the learning sample categories for yearly, (January-April 2018), (May-August 2018) and (September-December 2018) classifications, respectively. One observes that all the learning samples were correctly classified so there was not incorrectly classified learning samples for each of these classifications. However, these correctly classified samples have different behaviours for confidence values lower than 40% and higher than 60% and the same behaviour between 50% and 60%. We have more pixels (<500) which have confidence values lower than 40% in the January, July, October and December classifications compare to February, March, April, June, August, September and November ones, where we have less and less pixels (<200) having these confidence values respectively. May and yearly classifications do not have pixels having confidence values lower or equal to 40%. The number of pixels with confidence values between 60% and 100% is decreasing for all classifications. We have much more pixels with low confidences values in January, July, October and December classifications compare to other months with the exception of May with the yearly classifications where most of the pixels have high confidence values. These results are consistent with the quantitative evaluation demonstrating that the May and yearly classifications exhibit better results than the other monthly ones. October also exhibits the worst accuracy among the per month results. This difficulty to classify the October

month could be due to the presence of clouds and cloud shadows associated with bad landscape in our study area. Thus, introducing confidence to weight the majority vote would probably improve fusion results.

4.3 Visual analysis

Quantitative assessment is not able to reveal alone several kinds of errors present in the classifications. Conversely, a visual analysis makes it possible to identify other phenomena which are analysed in this section. Indeed, important parts of the study area are not labeled in the reference databases, and thus not included in the quantitative evaluation.

For the considered nomenclature, the study area is globally not difficult to classify using 10 m Sentinel-2 images even if the small size of *Crop Area with Trees* (CAT) and *Floodable Crop Area* (FCA) vegetation classes make it more difficult to classify. Figures 5 illustrate the classification of the yellow squared area over Maroua3 in the South-West of the study area. One can observe an important over-detection of the *Shrub Savannah* (SS) class, not necessarily revealed by the metrics, as SS are mostly over-detected over *Tree Savannah* (TS).

Table 2. Confusion matrix **January 2018**
 OA = 97.40% ; K = 0.96

	TS	SS	CA	CAT	FCA
TS	98.70	0.14	1.14	0.00	0.00
SS	3.39	92.76	3.83	0.00	0.00
CA	2.39	0.17	97.39	0.02	0.01
CAT	5.79	3.55	24.98	65.66	0.00
FCA	8.53	1.07	26.37	0.07	63.94

Table 3. Confusion matrix **February 2018**
 OA = 98.80% ; K = 0.98

	TS	SS	CA	CAT	FCA
TS	99.46	0.02	0.51	0.00	0.00
SS	0.79	98.09	1.11	0.00	0.00
CA	1.53	0.07	98.28	0.05	0.04
CAT	3.69	0.55	3.48	92.25	0.00
FCA	0.18	0.00	0.73	0.22	98.85

Table 4. Confusion matrix **March 2018**
 OA = 98.40% ; K = 0.97

	TS	SS	CA	CAT	FCA
TS	99.21	0.12	0.65	0.00	0.00
SS	2.23	96.42	1.34	0.00	0.00
CA	1.68	0.12	98.09	0.04	0.04
CAT	1.67	2.72	4.67	90.92	0.00
FCA	0.40	0.84	0.88	0.36	97.48

5. CONCLUSION

This paper presented multiple experiments for the fully automatic production of vegetation maps at regional scale out of Sentinel-2 images time series using the ν^2 supervised classification chain. This approach uses all available Sentinel-2 image

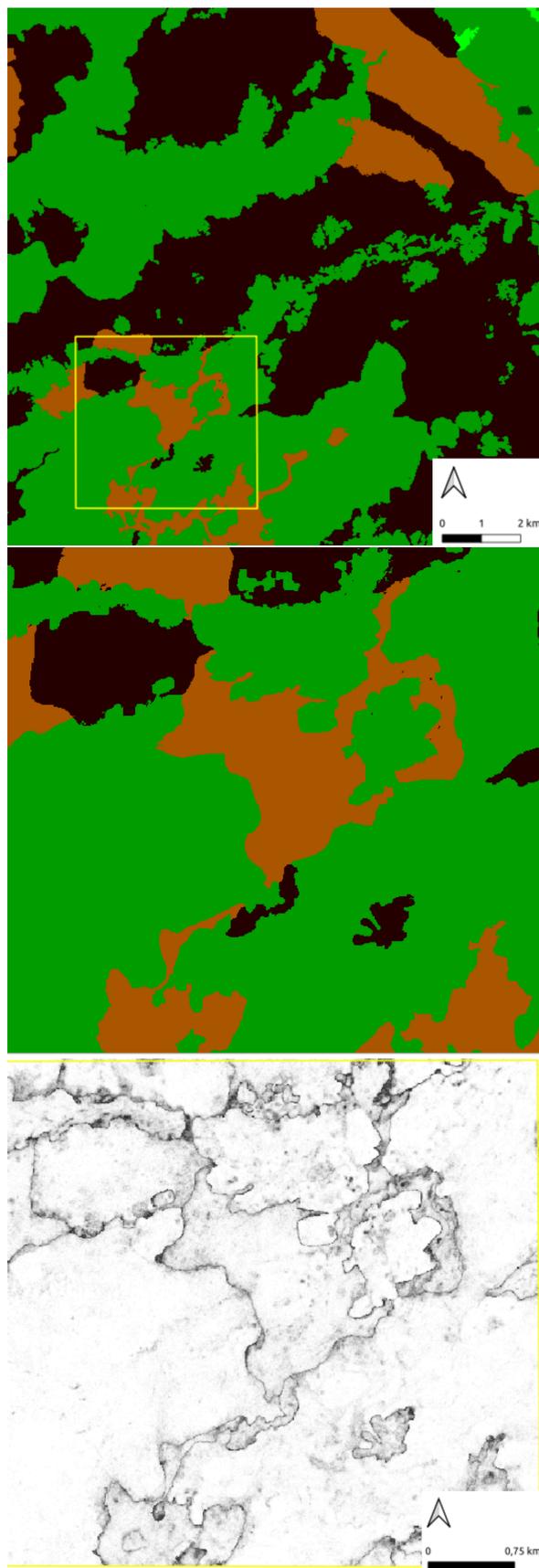


Figure 5. Vegetation and confidence maps. ■: TS; ■: SS; ■: CA; ■: CAT; ■: FCA. *Top*: full area. *Middle*: zoom of the study yellow squared area (full year). *Bottom*: Confidence map; ■: low → ■: high values.

Table 5. Confusion matrix **April 2018**
OA = 98.20% ; K = 0.97

	TS	SS	CA	CAT	FCA
TS	98.69	0.08	1.22	0.00	0.00
SS	2.37	94.98	2.64	0.00	0.00
CA	1.26	0.13	98.53	0.04	0.02
CAT	3.41	1.11	13.11	82.34	0.00
FCA	1.36	0.73	16.91	0.22	80.75

Table 6. Confusion matrix **May 2018**
OA = 99.60% ; K = 0.99

	TS	SS	CA	CAT	FCA
TS	99.82	0.01	0.16	0.00	0.00
SS	0.07	99.49	0.42	0.00	0.00
CA	0.41	0.02	99.46	0.05	0.03
CAT	1.11	0.69	2.16	96.02	0.00
FCA	0.03	0.03	0.36	0.22	99.33

Table 7. Confusion matrix **June 2018**
OA = 98.20% ; K = 0.97

	TS	SS	CA	CAT	FCA
TS	99.21	0.08	0.69	0.00	0.00
SS	3.63	95.58	0.78	0.00	0.00
CA	1.88	0.08	97.97	0.03	0.03
CAT	6.62	1.04	13.81	78.50	0.00
FCA	12.30	0.33	5.20	0.14	82.00

Table 8. Confusion matrix **July 2018**
OA = 96.40% ; K = 0.94

	TS	SS	CA	CAT	FCA
TS	96.76	0.06	3.17	0.00	0.00
SS	2.14	90.76	7.09	0.00	0.00
CA	2.27	0.14	97.52	0.02	0.02
CAT	6.14	1.53	27.84	64.48	0.00
FCA	4.35	0.00	14.25	0.07	81.30

Table 9. Confusion matrix **August 2018**
OA = 98.40% ; K = 0.97

	TS	SS	CA	CAT	FCA
TS	99.14	0.24	0.60	0.00	0.00
SS	1.41	95.92	2.65	0.00	0.00
CA	1.39	0.17	98.34	0.04	0.03
CAT	5.30	1.81	5.44	87.43	0.00
FCA	2.40	1.92	3.80	0.25	91.61

Table 10. Confusion matrix **September 2018**
OA = 99.40% ; K = 0.99

	TS	SS	CA	CAT	FCA
TS	99.68	0.06	0.24	0.00	0.00
SS	0.29	99.14	0.55	0.00	0.00
CA	0.63	0.03	99.22	0.05	0.04
CAT	1.18	0.76	1.18	96.85	0.00
FCA	0.00	0.00	0.00	0.29	99.70

Table 11. Confusion matrix **October 2018**
OA = 94.00% ; K = 0.90

	TS	SS	CA	CAT	FCA
TS	98.42	0.08	1.48	0.00	0.00
SS	15.90	82.17	1.92	0.00	0.00
CA	6.83	0.12	93.01	0.02	0.01
CAT	7.60	0.34	30.77	61.27	0.00
FCA	44.62	4.95	10.45	0.22	39.74

Table 12. Confusion matrix **November 2018**
OA = 99.20% ; K = 0.99

	TS	SS	CA	CAT	FCA
TS	99.71	0.10	0.17	0.00	0.00
SS	0.79	98.91	0.29	0.00	0.00
CA	0.94	1.00	98.85	0.04	0.04
CAT	5.30	0.48	4.67	89.53	0.00
FCA	0.18	0.00	0.07	0.29	99.44

Table 13. Confusion matrix **December 2018**
OA = 95.20% ; K = 0.92

	TS	SS	CA	CAT	FCA
TS	97.87	0.68	1.44	0.00	0.00
SS	12.22	84.97	2.79	0.00	0.00
CA	4.42	0.26	95.26	0.02	0.02
CAT	11.02	1.46	20.79	66.71	0.00
FCA	18.61	1.07	13.59	0.11	66.60

Table 14. Confusion matrix **Year 2018**
OA = 99.60% ; K = 0.99

	TS	SS	CA	CAT	FCA
TS	99.86	0.02	0.10	0.00	0.00
SS	0.20	99.58	0.21	0.00	0.00
CA	0.57	0.02	99.30	0.05	0.03
CAT	1.39	0.83	1.18	96.58	0.00
FCA	0.40	0.00	0.18	0.33	99.07

data (no scene selection in terms of appropriate dates or cloud cover) and uses existing geodatabases as reference training and validation data for a supervised classification process robust to

errors in the reference data (e.g., out of date databases). The process is efficient (enabling a fast delivery of the classifications after the acquisition of the satellite images) and not

Table 15. Confusion matrix **Fusion**
OA = 98.10% ; **K** = 0.97

	TS	SS	CA	CAT	FCA
TS	96.91	0.51	2.55	0.00	0.01
SS	0.25	96.10	3.63	0.00	0.00
CA	0.14	0.00	99.77	0.03	0.04
CAT	0.42	0.00	0.42	92.93	6.21
FCA	0.00	0.00	0.20	0.00	99.79

expensive (requiring neither field surveys for model calibration, nor human operators for decision making, and using open and freely available imagery). To conclude, the use of fusion of monthly classifications is a good trade-off between accuracy et computational time. This work was conducted in the context of defining a joint use of airborne and free satellite images to update and enrich existing LC data. The advantage of the proposed approach resides in the regular availability of new images provided by Sentinel-2 to reduce the uncertainty in the detection of study areas where airborne acquisitions and advanced analysis are mandatory. The considered legend was rather simple, including *Tree Savannah*, *Shrub Savannah*, *Crop Area without Trees*, *Crop Area with Trees*, and *Floodable Crop Area* vegetation classes. Despite the obtained accuracy is higher than most comparable (in terms of areas covered and number of classes) state-of-the-art approaches, *Crop Area with Trees* and *Floodable Crop Area* (corresponding to thin objects) suffered from poor recognition and advocates for the integration of VHR imagery. On the opposite, better results (90%) were obtained with the *Crop Area without Trees* class.

Acknowledgments: This work has been partly supported by the scholarships in support of joint supervision theses of the University Gustave Eiffel for the year 2020 and the national scholarship for the public servants of the Cameroon state as well as by the PARCELLE project funded by the French CNES TOSCA program.

REFERENCES

Benayas, J. M. R., Bullock, J. M., 2012. Restoration of Biodiversity and Ecosystem Services on Agricultural Land. *Ecosystems*, 15, 883–899.

Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), 5-32.

Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 27-46.

David, N., Giordano, S., Mallet, C., 2021. Investigating operational country-level crop monitoring with Sentinel 1 and 2 imagery. *Remote Sensing Letters*, 12(10), 970-982.

Derksen, D., Inglada, J., Michel, J., 2018. Spatially precise contextual features based on superpixel neighborhoods for land cover mapping with high resolution satellite image time series. *IGARSS*, 200–203.

Fleiss, J., Cohen, J., 1973. The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613-619.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O’Connell, C., Ray, D. K., et al., P. C. W., 2011. Solutions for a Cultivated Planet. *Nature*, 478, 337–342.

Frenay, B., Verleysen, M., 2014. Classification in the presence of label noise: A survey. *IEEE Trans. Neural Netw. Learn. Syst.*, 25, 845-869.

Gao, B., 1996. NDWI. A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58, 257-266.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., Toulmin, C., 2010. Food Security: The Challenge of Feeding 9 Billion People. *Science*, 327, 812–818.

Green, R. E., Cornell, S. J., Scharlemann, J. P. W., Balmford, A., 2005. Farming and the Fate of Wild Nature. *Science*, 307, 550–555.

Gressin, A., Vincent, N., Mallet, C., Paparoditis, N., 2014. A unified framework for land-cover database update and enrichment using satellite imagery. *IEEE International Conference on Image Processing*, 5057–5061.

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., Rodes, I., 2017. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sensing*, 9(9), 19-35.

Le Bris, A., Chehata, N., Ouerghemmi, W., Wendl, C., Postadjian, T., Puissant, A., Mallet, C., 2019. Chapter 11 - decision fusion of remote-sensing data for land cover classification. M. Y. Yang, B. Rosenhahn, V. Murino (eds), *Multimodal Scene Understanding*, Academic Press, 341–382.

Mallet, C., Le Bris, A., 2020. Current challenges in operational very high resolution land-cover mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020, 703–710.

Mertens, B., Lambin, E., 2000. Land-Cover-Change Trajectories in Southern Cameroon. *Annals of the Association of American Geographers*, 90(3), 467-494.

Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187, 156-168.

Postadjian, T., Le Bris, A., Sahbi, H., Mallet, C., 2017. Investigating the potential of deep neural networks for large-scale classification of very high resolution satellite images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1/W1, 183–190.

Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104.

Rounsevell, M., Ewert, F., Reginster, I., Leemans, R., Carter, T., 2005. Future scenarios of European Agricultural Land Use. *Agric. Ecosyst. Environ.*, 107, 117–135.

Schwilch, G., Bestelmeyer, B., Bunning, S., Critchley, W., Herrick, J., Kellner, K., Liniger, H., Nachtergaele, F., Ritsema, C., et al., B. S., 2010. Experiences in Monitoring and Assessment of Sustainable Land Management. *Land Degrad. Dev.*, 22, 214–225.

Searchinger, T., Heimlich, R., Houghton, R. A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D., Yu, T. H., 2008. Use of US Croplands for Biofuels Increases Greenhouse Gases through Emissions from Land-Use Change. *Science*, 319, 1238–1240.

Tilman, D., 2001. Forecasting Agriculturally Driven Global Environmental Change. *Science*, 292, 281–284.

Tucker, C., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127-150.