HAL open science

Classification of time series of Sentinel-2 images for large scale mapping in Cameroon

Hermann Tagne, A. Le Bris, D. Monkam, Clément Mallet

To cite this version:

Hermann Tagne, A. Le Bris, D. Monkam, Clément Mallet. Classification of time series of Sentinel-2 images for large scale mapping in Cameroon. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Aug 2020, Nice, France. pp.633-640, 10.5194/isprs-archives-XLIII-B3-2020-633-2020 . hal-02927468

HAL Id: hal-02927468

https://hal.science/hal-02927468v1

Submitted on 1 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CLASSIFICATION OF TIME SERIES OF SENTINEL-2 IMAGES FOR LARGE SCALE MAPPING IN CAMEROON

Hermann Tagne[1,2,3,*], Arnaud Le Bris[1], David Monkam[2], Clément Mallet[1]

[1] University Gustave Eiffel, IGN-ENSG, LaSTIG, 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; mapping; Satellite; time series; Sentinel-2.

ABSTRACT:

Sentinel-2 satellites provide dense image time series exhibiting high spectral, spatial and temporal resolution. These images are in particular of utter interest to map Land-Cover (LC) at large scale. 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 national topographic geodatabase. The ι^2 framework is adopted and tested for the specificity of the country. Here, experiments focus on generic 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, allowing to provide a first step towards finer-grained map retrieval.

1. INTRODUCTION

1.1 The need for Land-Cover (LC) mapping

Land cover (LC) is defined as the description of the biophysical material over the surface of the Earth and immediate subsurfaces and man-made structures. Understanding and monitoring LC is important especially in environmental studies. Thus, updating LC maps is essential for countries in development and implementation of environmental policies (Grekousis, Mountrakis, 2015). Human activities related to serving specific societal and individual needs are the main drivers of contemporary LC dynamics (Grekousis et al., 2015).

1.2 Sentinel-2 for LC mapping

The launch of the Sentinel-2 satellites in June 2015 enabled the free access to an unprecedented amount of optical images. The large swath, the short revisit time, the high spatial resolution of about 10 m, and the spectral bands from visible to short wave infra-red have already become essential to monitor large territories for a great variety of environmental applications.
Sentinel-2 (S2) satellites provide a global cover of continental surfaces every five days in 13 spectral bands. This now enables the implementation of novel LC map production systems for the delivery of up to date and accurate information with the appropriate timeliness (Whitcraft et al., 2015a, Whitcraft et al., 2015b). The enhanced temporal resolution ensures a better monitoring of land use and cover with better opportunities to obtain cloudless mosaics and to discriminate natural classes (vegetation types, crops, etc.) (Immitzer et al., 2016). The wide spectral resolution facilitates the thematic identification of LC, while the high spatial resolution allows for the identification of quite small objects or landscapes structures (built-up areas, large roads, specific textures).

* Corresponding author

1.3 Needs of the National Cameroonian Mapping Agency

The national Cameroonian mapping agency aims at implementing a fully operational automatic LC map production system using both time series of Sentinel-2 images and available aerial images. So far, the topographic database is established on the basis of the visual interpretation of the latter images and field surveys. The nomenclature of the national Cameroonian reference topographic and land cover database is highly detailed (9 themes, 47 classes, and 136 sub-classes) and does not fully match the current requirements of a yearly basis with detected changes, as it now appears to be a crucial feature (Wulder et al., 2018). Therefore, it is here targeted at investigating how Sentinel-2 and aerial images can be jointly beneficial for this task since they exhibit complementary strengths.

1.4 Scope of this work

This study focuses on the first step of this exploratory workflow: the analysis of Sentinel-2 images for yearly LC map generation. It aims at assessing state-of-the-art classifiers to this specific context.
Indeed, a first difficulty states in the lack of reference data at the spatial resolution of Sentinel images for the training of a supervised classifier. The proposed solution consists in a direct use of existing Very High Resolution topographic database with the existing ι^2 LC classification framework.
Other challenges consist in working on areas with numerous cloudy days and significant intra-annual variability (Mertens, Lambin, 2000).

2. DATA AND STUDY SITE

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.

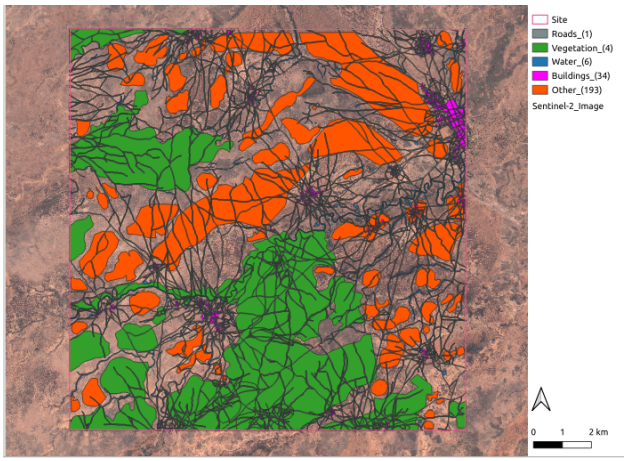

Figure 1. Study area and reference data.

2.1 Study site

Current experiments and analyses are limited to one specific site, located near Petté in the East-North of Cameroon, (see Figure 1). Covering 189.072 km^2, it was selected both for its challenging conditions and the availability of reference and Copernicus data (Figure 1). 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 Figure 2). All Levels 2A and 2B data available through the Copernicus Open Access Hub[1] over the study area for the year 2018 were used. For each month, 4-7 images were available (Figure 2). More specifically, a focus was made on a sub-area of 10 km $\times$ 10 km.

2.2.2 ***Reference data*** The proposed approach relies on existing geodatabases available over the area to build the training and reference datasets, needed for the supervised classification and the subsequent validation of the LC maps. Two data sources have been used in order to build a unique reference dataset :

- The topographic LC database of the Cameroonian National Cartographic Institute (BDINC, simplified for our purpose);
- OpenStreetMap data, made available through the QUICKOSM[2] plugin of QGIS software (Herbreteau et al., 2018).

In further experiments, five generic classes are considered for the classification: *Roads*, *Vegetation*, *Water*, *Buildings* and *Other*. The *Other* class corresponds to the rest of Land-Cover objects in the study area except the four others ones. Remaining unlabelled areas correspond to LC types not included in the specifications of the reference database.

[1] https://scihub.copernicus.eu
[2] https://plugins.qgis.org/plugins/QuickOSM/

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 Iota2 or ι^2 supervised classification framework (Inglada et al., 2017) is adopted. 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 already demonstrated its high efficiency in performing country wide LC classification out of multi-temporal high resolution Landsat and Sentinel imagery at pixel level. 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 Forests

A Random Forests (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:

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

Each step is detailed in the followings sections.

Iota2 configuration for each month and the year 2018

Satellite	Mois	I	IS	TR	B	SpeR	SpaR	F	M	T	V	K	m	max_depth	min_samples
Sentinel-2	January	6	1368 × 1385	5 days	B2 (Blue)	0.46–0.52	10	SB	60	1051919	1051919	1000	0.2	25	5
	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)	0.70–0.71	20								
	May	6			B6 (Red-edge)	0.73–0.74	20	SB-NDWI							
	June	6			B7 (Red-edge)	0.77–0.79	20								
	July	6			B8 (Near infra-red)	0.78–0.90	10	SB-SAVI							
	August	5			B8A (Near infra-red)	0.85–0.87	20		56						
	September	6			B11 (Near infra-red)	0.93–0.95	60	SB-NDRI	66						
	October	7			B12 (Shortwave infra-red)	1.57–1.66	20		76						
	November	6							66						
	December	6													

I: total number of images per month. **IS**: Images Size (in pixel). **TR**: temporal resolution. **B**: bands. **SpeR**: spectral resolutions (in µm). **SpaR**: spatial resolutions (in m). **F**: features. **SB**: spectral bands. **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). **T**: number of training pixels (all polygons is used to train models so 100% of pixels). **V**: number of validation pixels (all polygons is used for validation so 100% of pixels). **K**: number of trees. **m**: number of features randomly selected at each node (20% of pixels is used to split a RF node). **max_depth**: maximal depth of each tree. **min_samples**: minimal number of samples per node.

Figure 2. *Iota2* (ι^2) configuration for monthly and yearly classifications over the study area.

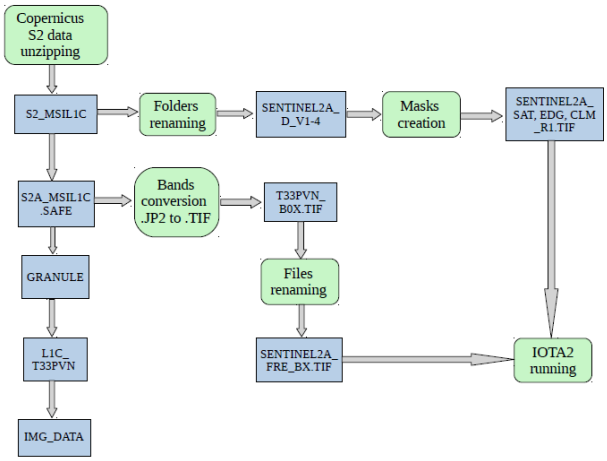

Figure 3. Data preparation.

3.3.1 ***Data preparation*** This step (Figure 3) consists in adapting the Copernicus Sentinel-2 data to fit the input format requirements of ι^2.
First, images bands are converted from $.jp2$ 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 that will be 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;
- Soil Adjusted Vegetation Index (SAVI);
- Normalised Difference Red Edge (NDRE).

Selecting *batchProcessing* parameter in ι^2's 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. Here, for a year long time series, it amounts to 760 features: 10 spectral bands plus 4 indices on 70 dates (see Figure 2). 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 this classifier 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 classes *Roads*, *Vegetation*, *Water*, *Buildings* and *Other* are considered in these experiments. (The *Other* class corresponds to the rest of Land-Cover objects in the study area except the four others ones.) Training samples are extracted for these classes from the reference dataset described in Section 2.2.2, and especially 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 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 (Figure 1). These areas are different (see Figure 4) depending on classes and so the class samples are unbalanced.

Build RF model In supervised classification the balance between class samples is important. There are many ways to manage class balancing in $iota^2$, using either data augmentation to enrich the training sample set or class weighting when training the classifier. 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 choice of parameters as demonstrated in (Pelletier et al., 2016). Thus, it is here applied with the standard ι^2 configuration (see fig. 2): 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.

3.3.4 ***Prediction*** This step simply consists in classifying Sentinel-2 data according to model obtained at 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 classes listed above.

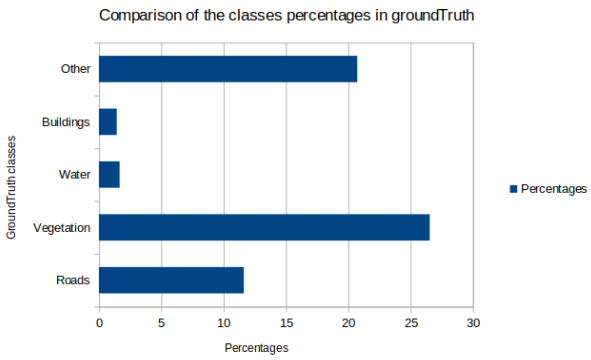

Figure 4. Percentages of classes in the reference data over the study area.

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 from four to seven Sentinel-2 images.

In this step, 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. 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 (supposed to be all well coregistered).

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 (from month or year long time series, and fusion of monthly 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 9). 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 5 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. These metrics were calculated over one run using the same training and validation sets (of 1,051,919 pixels each). Kappa coefficients are generally very good with values above 0.8 (0.88-0.93 for monthly classifications, 0.94 for the yearly one and 0.87 the fusion). Table 4 shows the confusion matrix for the fusion of monthly classifications. It can be highlighted that overall accuracy and Kappa coefficient values are 0.91 and 0.87 respectively. One can observe that the yearly classification yields better results than monthly classifications and their fusion for these two global metrics. Most monthly classifications results are also better than the fusion one.
This is slightly different when considering per class metrics. The classes *Vegetation* and *Other* obtained the best results for the all classification modes. Compared to monthly results, the fusion can slightly improve results (especially considering underdetection), mostly for the other classes, but the best results are still obtained for yearly classifications.The results are even more improved for the classes with few training samples (*Water* and *Building*). The class *Vegetation* is much better recognised using the two classification modes and the fusion of monthly classifications, mainly because of its higher training samples (see Figure 4).
It can also be shown from Figure 5 the improvement for these minority classes (*Water* and *Buildings*) mostly states in the recall: underdetection is reduced. Tables 1, 2 and 3 show the confusion matrices for the best month (May), the baddest month (December) and the year 2018 respectively. They allow a detailed analysis of the errors for both modes. Several aspects of the improvement yielded by the yearly classification can be highlighted:

- No confusion between *Other* and *Vegetation* in both directions;
- The confusion of *Water* with *Vegetation* is reduced in one direction;
- The confusion of *Water* with *Other* is reduced in one direction from 15.61% in June 2018 to 0.67% in the year 2018.

Table 4 shows the confusion matrix for the result from the fusion of the twelve monthly classifications. Several aspects of his improvement can also be highlighted:

- No confusion between *Other* and *Vegetation* in both directions;
- The confusion of *Water* with *Vegetation* is reduced in one direction;

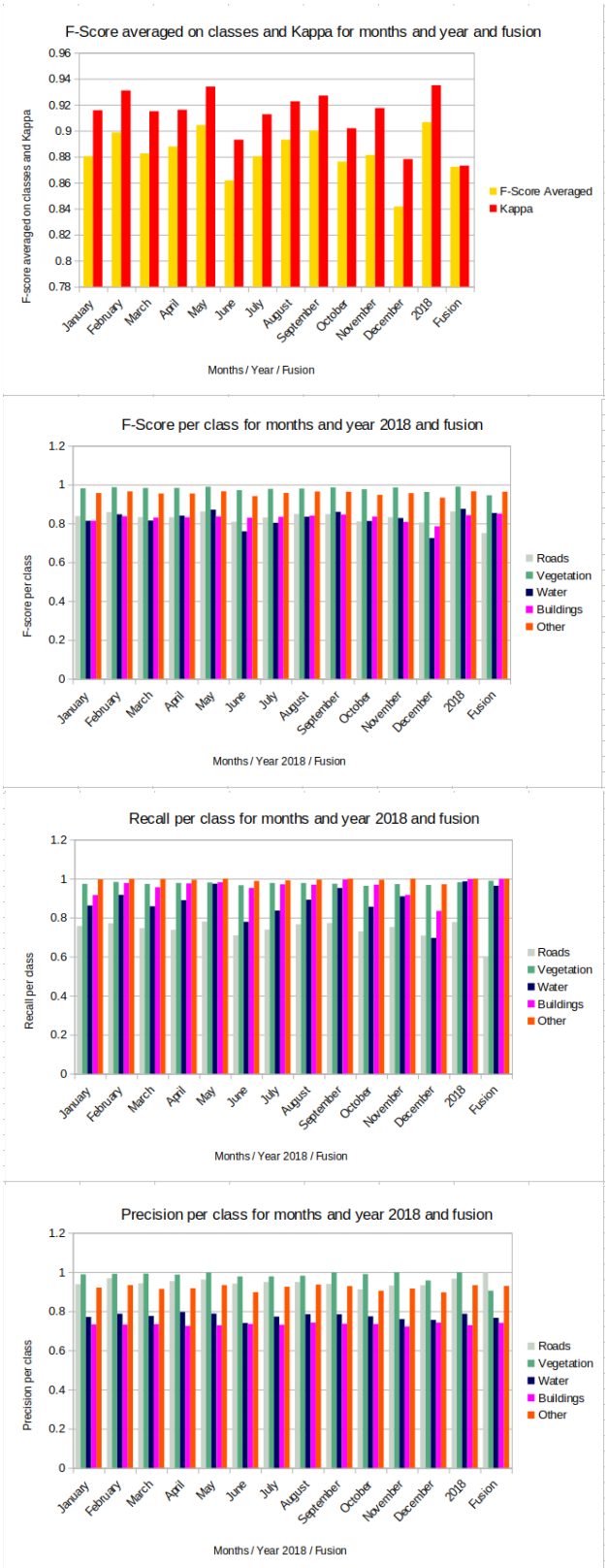

Figure 5. Global (F-Score averaged on classes and Kappa) and per class (F-Score, recall and precision) quality metrics for the study area for months and year 2018 and fusion.

- The confusion of *Water* with *Other* increases to 2.57% compare to the yearly classification.

It can be seen from the confusion matrices most misclassifications occur between *Roads* and *Vegetation* classes, as well as the *Roads* classified as *Other* class. This is partly caused by the fact that bare ground areas (that can be included in these classes in the reference databases) are often classified as *Roads* class.

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. Figure 12 shows the confidence map obtained for the LC map in the yellow squared area of Figure 9 for the year 2018. One can observe that confidence is influenced by the proportion occupied by the classes samples in the reference data (see Figure 4). For example, the *Vegetation* class in that specific yellow zone is easily recognised in the yearly confidence classification. Besides, areas corresponding to LC not included in the reference databases generally exhibit 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. Figure 13 shows the distribution of confidence values calculated for only the learning samples categorie for December, May and yearly classifications. One can observe that all the learning samples were correctly classified (in yellow) so there was not incorrectly classified learning samples (in blue) for each of these specific classifications. However, these correctly classified learning 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 (<200) which have confidence values lower than 40% in the December classification compare to May and the yearly classifications where we have less and less pixels having these confidence values respectively. On the opposite, we have more and more pixels having confidence values which evolve from 60% to 100% in May and yearly classifications respectively compare to the December classification where the number of pixels having these confidence values is decreasing from 1000 to 0. These results are consistent with the quantitative evaluation arguing that the yearly classification exhibits better results than monthly classifications. December also exhibits the worst accuracy among the per month results. This difficulty to classify the December 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, while 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 legend, the study area is globally not difficult to classify using 10 m S2 images even if the small size of *Buildings* class makes it more difficult to classify. Figures 6, 7 and 8 illustrate the classification of the yellow squared area in the South-West of the study area for year 2018, December 2018 and May 2018 respectively. One can observe an important over-detection of the *Roads* class, not necessarily revealed by the metrics, as roads are mostly over-detected over unlabelled

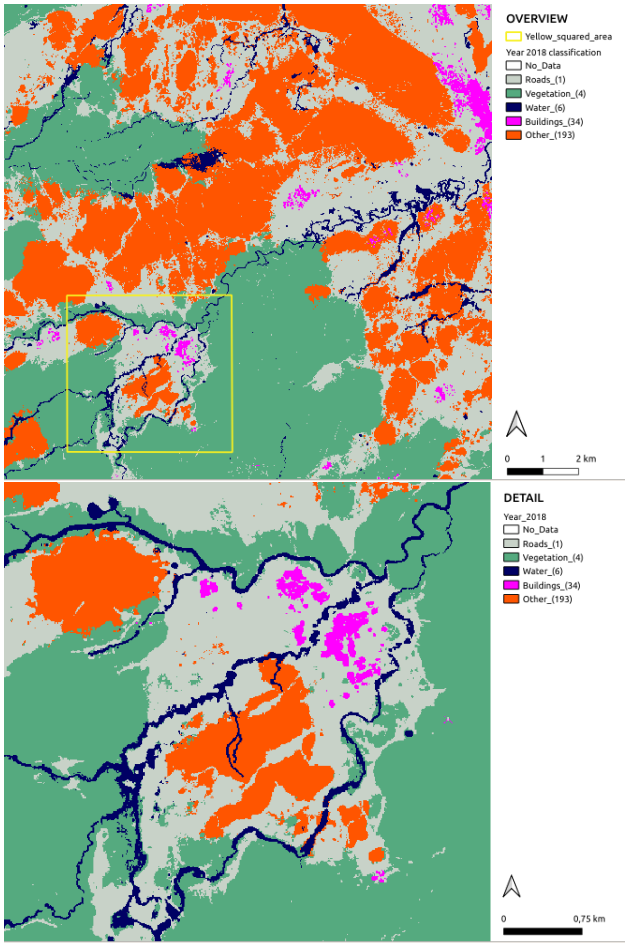

Figure 6. Land-Cover map and zoom of the study yellow squared area for year 2018.

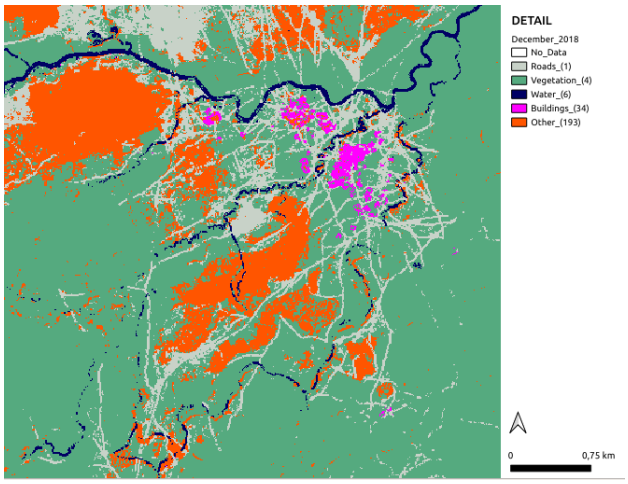

Figure 7. Zoom of the study yellow squared area for dec. 2018.

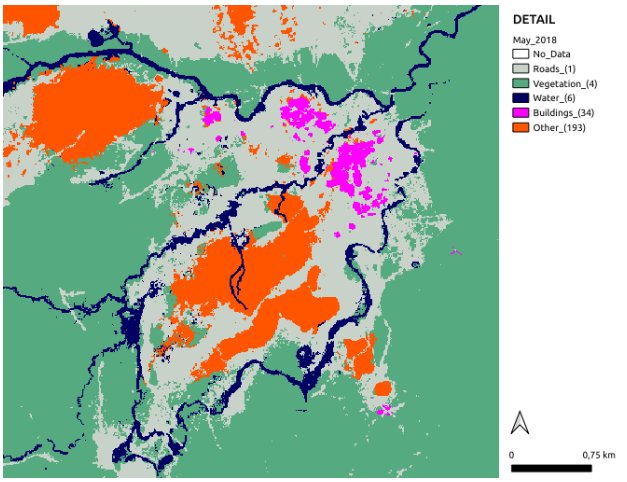

Figure 8. Zoom of the study yellow squared area for may 2018.

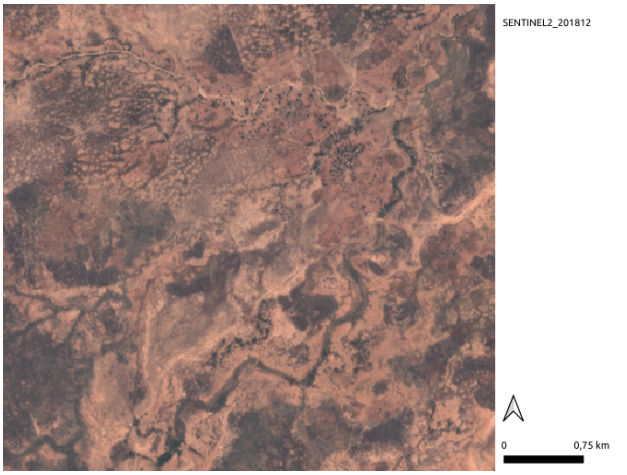

Figure 9. Sentinel-2 image for the study yellow squared area for dec. 2018.

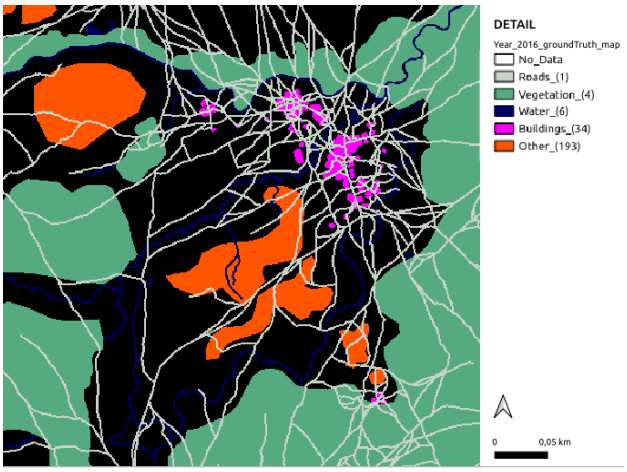

Figure 10. Ground truth map: zoom of the study yellow squared area for year 2016.

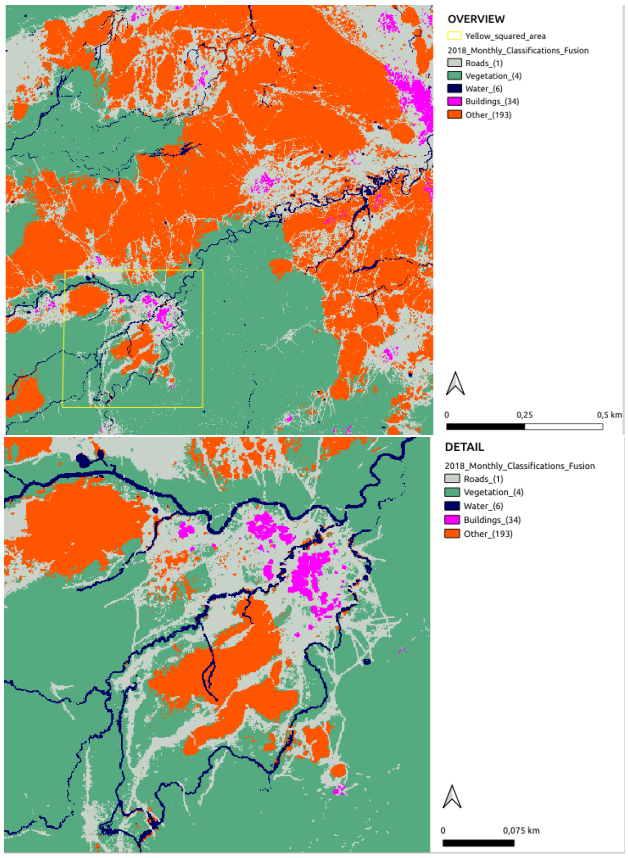

Figure 11. Monthly Land-Cover map fusion and zoom of the study yellow squared area.

areas corresponding to LC types not included in the reference database. December gives the baddest result per month, even though the over-detection of roads is less important. This result is consistent with the previous quantitative assessment.

Table 1. Confusion matrix **May 2018**
OA = 95.80% ; **K** = 0.93

	Roads	Veg.	Water	Build.	Other
Roads	**77.98**	0.66	0.43	3.93	16.97
Veg.	0.91	**98.00**	1.05	0.01	0.00
Water	0.10	1.29	**97.36**	0.00	1.23
Build.	0.39	0.14	0.00	**98.10**	1.35
Other	0.00	0.03	0.00	0.00	**99.96**

Table 2. Confusion matrix **December 2018**
OA = 92.40% ; **K** = 0.88

	Roads	Veg.	Water	Build.	Other
Roads	**70.81**	4.37	0.33	3.13	21.33
Veg.	1.51	**96.70**	0.91	0.01	0.84
Water	0.15	18.03	**69.59**	0.00	12.22
Build.	0.12	6.00	0.00	**83.42**	10.44
Other	0.00	2.87	0.00	0.00	**97.12**

Table 3. Confusion matrix **Year 2018**
OA = 96.00% ; **K** = 0.94

	Roads	Veg.	Water	Build.	Other
Roads	**77.73**	0.54	0.45	3.99	17.26
Veg.	0.80	**98.09**	1.07	0.01	0.00
Water	0.06	0.74	**98.52**	0.00	0.67
Build.	0.11	0.02	0.00	**99.74**	20.11
Other	0.00	0.00	0.00	0.00	**99.99**

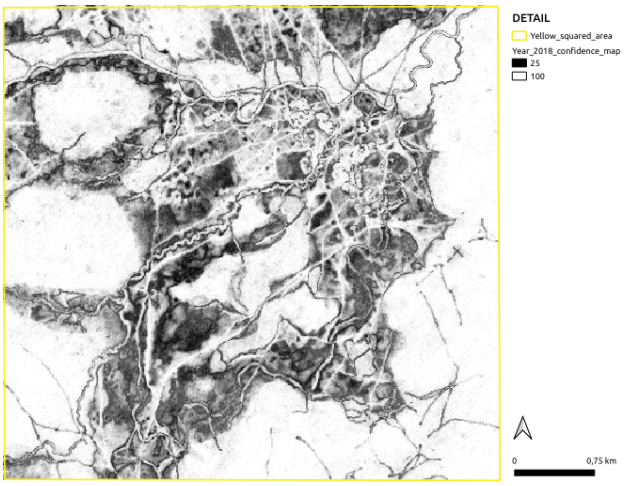

Figure 12. Confidence map: zoom of the study yellow squared area for year 2018. The darker the pixel, the lower the confidence.

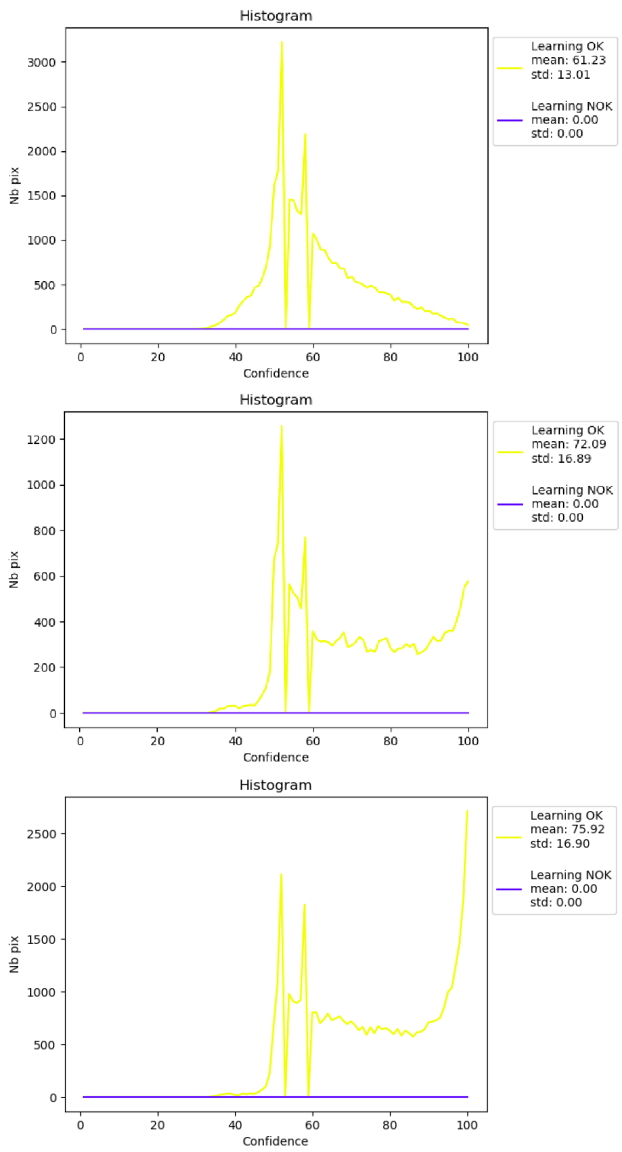

Figure 13. Confidence statistics for the study area for december, may and year 2018 from up to down.

Table 4. Confusion matrix **Fusion**
OA = 91.78% ; **K** = 0.87

	Roads	Veg.	Water	Build.	Other
Roads	**60.01**	23.29	0.31	2.82	13.55
Veg.	0.00	**98.86**	1.13	0.00	0.00
Water	0.00	1.09	**96.29**	0.00	2.57
Build.	0.01	0.03	0.00	**99.83**	0.11
Other	0.00	0.00	0.00	0.00	**100.00**

5. CONCLUSION

This paper presented some experiments of fully automatic production of LC maps at regional scale out of Sentinel-2 images time series using the ι^2 supervised classification chain. This approach uses all available Sentinel-2 image data (no scene selection in terms of appropriate dates or cloud cover) and uses existing databases 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 satellites image data) and not expensive (requiring neither field surveys for model calibration, nor human operators for decision making, and using open and freely available imagery). 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 quite simple, including *Roads*, *Vegetation*, *Water* and *Buildings* classes, as well as an additional *Other* class corresponding to all other database objects of the study area. Despite obtained accuracy is higher than most comparable (in terms of areas covered and number of classes) state-of-the-art approaches, *Roads* and *Water* classes (corresponding to thin objects) suffered from poor recognition. On the opposite, better results (90%) were obtained with the *Vegetation* class. It will now be beneficial to develop the same study only on this particular changing class for also fulfilling the user requirements in terms of crop type mapping and easily compare crop type maps of successive years. S2 time series is supposed to be especially relevant for this task.

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

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.

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.

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.

Grekousis, G., Kavouras, M., Mountrakis, G., 2015. Land Cover Dynamics and Accounts for European Union 2001-2011. *Third International Conference on Remote Sensing and Geoinformation (RSCy2015)*, 9535, 953507.

Grekousis, G., Mountrakis, G., 2015. Sustainable Development under Population Pressure: Lessons from Developed Land Consumption in the Conterminous US. *PLoS ONE*, 10(3), e0119675.

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.

Herbreteau, V., Rvillion, C., Trimaille, E., 2018. Orfeo toolbox applications. N. Baghdadi, C. Mallet, M. Zribi (eds), *QGIS in Remote Sensing*, 1, ISTE/Wiley, chapter 7, 257–286.

Immitzer, M., Vuolo, F., Atzberger, C., 2016. First experience with Sentinel-2 data for crop and tree species classifications in Central Europe. *Remote Sensing*, 8, 166.

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*, 95(9), 19-35.

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.

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

Whitcraft, A., Becker-Reshef, I., Justice, C., 2015a. A framework for defining spatially explicit Earth Observation requirements for a global agricultural monitoring initiative (GEOGLAM). *Remote Sensing*, 7, 1461-1481.

Whitcraft, A., Becker-Reshef, I., Killough, B., Justice, C., 2015b. Meeting Earth Observation requirements for global agricultural monitoring: An evaluation of the revisit capabilities of current and planned moderate resolution optical earth observing missions. *Remote Sensing*, 7, 1482-1503.

Wulder, M., Coops, N., Roy, D., White, J., Hermosilla, T., 2018. Land cover 2.0. *International Journal of Remote Sensing*, 39(12), 4254-4284.