EXPLOITATION OF SPECTRAL AND TEMPORAL INFORMATION FOR MAPPING PLANT SPECIES IN A FORMER INDUSTRIAL SITE

R. Gimenez^{12*}, G. Lassalle³, R. Hédacq³, A. Elger², D. Dubucq³, A. Credoz³, C. Jennet³, S. Fabre¹

¹ ONERA DOTA BP74025 2 av. Edouard Belin, FR-31055 Toulouse cedex 4 - rollin.gimenez@onera.fr

² Laboratoire d'Ecologie Fonctionnelle et environnement, ENSAT - Avenue de l'Agrobiopole 31326 Castanet-Tolosan cedex, France ³ TOTAL S.E., CSTJF - avenue Larribau, 64000 Pau, France

KEY WORDS: Hyperspectral, multispectral, multitemporal, vegetation, classification, anthropogenic impacts, machine learning.

ABSTRACT:

Characterization and seasonal (periodic) monitoring of plant species distribution in the context of former industrial activity is crucial to assess long-term anthropogenic footprint on vegetated area. Species discrimination has shown promising results using both HyperSpectral (HS) and MultiSpectral (MS) images. Airborne HS instruments enable high spatial and spectral resolution imagery while time series of satellite MS images provide high temporal resolution and phenological information. This paper aims to compare supervised classification results obtained with non-parametric (Random Forest, RF, Support Vector Machine, SVM) and parametric methods (Regularized Logistic Regression, RLR) applied on both kinds of images acquired on an industrial brownfield. The studied site is a complex vegetated environment due to species diversity: 8 dominant species are retained. The performance obtained by preliminary feature selection based on principal component analysis and vegetation indices, to improve separability of spectral or temporal information according to species, is analysed. The best performance is obtained by RLR method applied on HS data without feature selection (global accuracy of 93 %). Feature selection is found to be a necessary step to perform classification with time series of MS images. Species that are difficult to distinguish from the HS image, namely *Salix* and *Populus*, are well separated using Sentinel-2 images (precision around 70%).

1. INTRODUCTION

Anthropogenic activities can result in various kinds of environmental footprint including soil contamination. Since soil contaminants, such as hydrocarbons or heavy metal, can damage biodiversity by impacting the plant physiology (Lassalle et al., 2019) and species spatial repartition and assemblage (Oniya et al., 2019), it is of first importance to know the distribution of species growing in a context of former industrial activity.

HyperSpectral (HS) imaging is a widely explored topic that proved to be suited to vegetation classification. Indeed, such images detect subtle spectral differences in relation to biophysical and biochemistry vegetation traits specific to species (Fassnacht et al., 2016) or assemblages of species in case of complex species communities (Burai et al., 2015). Used in conjunction with dimension reduction techniques, machine learning algorithms are among the most efficient methods to perform vegetation classification (Gewali et al., 2018). Support vector machine or random forest algorithms usually lead to the best model performance (Guidici, Clark, 2017). Applied on HS images with spatial resolutions from 1 m to 5 m, these methods make it possible to classify 4 to 8 trees species with accuracies varying between 60 % and 90 % (Fassnacht et al., 2014; Dabiri, Lang, 2018; Richter et al., 2016) as well as complex alkali vegetation with more than 79 % of accuracy (Burai et al., 2015). Nevertheless, dimension reduction techniques do not always improve results (Dalponte et al., 2012) and can make their interpretability very challenging (Fassnacht et al., 2016).

Studies have also shown that considering the phenology of the vegetation tends to improve classification accuracy (Fassnacht et al., 2016). Using this, MultiSpectral (MS) satellites like Sentinel-2 with high temporal resolution enable the development of efficient classification (Macyntre et al., 2020).

Even if spectral and spatial resolutions are generally lower on images captured with such instruments, the exploitation of plants' phenology by machine learning algorithms provide performance up to 90 % on tree species (Denisova et al., 2019; Clark et al., 2018) and up to 74% on homogeneous shrubs vegetation (Macyntre et al., 2020).

A few studies comparing the efficiency of a single-date HS image with time series of MS images for tree classification have demonstrated that similar or even higher accuracies can be derived from the latter (Clark, 2020; Griegorova et al., 2020).

However, these studies are limited to tree species classification and compare results obtained with MS and HS instruments with the same range of spatial resolution (10-30 m).

Our objective is thus to compare classification results obtained by machine learning methods applied on airborne HS imagery with very high spatial resolution and temporal series of MS satellite image Sentinel-2 in a brownfield site. The brownfield is a complex vegetated environment owing to various kinds of vegetation and strata, from tall trees to low-layer vegetation such as grass. Here, the vegetation classes include single species (*e.g.*, for trees) and assemblage of species (*e.g.*, grass mixture).

2. STUDY SITE

The study area was an industrial brownfield located in a temperate region covering approximately 2.45 km². Part of the site was exposed to activities for over 20 years that resulted in different kinds of impacts on soil and vegetation. Since then, vegetation had colonized most of the area (Figure 1).

Vegetation was mainly made of deciduous trees, shrubs, and grassland among which fourteen species or assemblage of species were identified. The same species were also studied at a

^{*} Corresponding author

non-impacted site with similar soil properties located next to the brownfield (Figure 1).



Figure 1. Industrial brownfield (surrounded in red) and reference site (surrounded in blue) extracted from BD ORTHO® 25 cm.

3. MATERIALS

3.1 HyperSpectral images

HS radiance images were acquired over the study area in the absence of cloud using HySpex cameras in July 5, 2017.

These images have a spatial resolution of 1 and 2.5 m and a spectral resolution of 5.2 and 7.8 nm in the visible-near-infrared (VNIR: 400-1000 nm) and short-wave infrared domains (SWIR: 1000-2500 nm) respectively. The SWIR image was registered to the VNIR image using the Gefolki algorithm and resampled to 1-m spatial resolution using a nearest neighborhood filter to provide a hypercube covering the VNIR-SWIR domain (Brigot et al., 2016). Because of the lack of knowledge about the composition of the local industrial atmosphere, the spectral radiances were converted to spectral reflectances using the Empirical Line Method (ELM) (Smith, Milton, 1999). A Savitsky-Golay filter was then applied to the image to improve Signal-to-Noise Ratio (SNR) (Erudel et al., 2017). Finally, bands within the interval 500-2500 nm with atmospheric transmission below 80 % were not retained due to their low signal-to-noise ratio. The processed image was then a georeferenced spectral reflectance hypercube composed of 227 spectral bands.

3.2 MultiSpectral images

MS images were acquired from the Sentinel-2 satellites in 2017. Images were selected only if they were cloud-free georeferenced, surface reflectance images (correction level 2A) available on the Theia platform (Theia, 2021). One image per month obtained from Sentinel-2A (S-2) satellite was kept (Table 1). Finally, spectral bands with a spatial resolution of 20 m were resampled at 10-m with a nearest neighborhood filter.

	Sentinel 2A	HySpex	
Band	8	227	
number			
Spatial	10 m	1 m	
resolution			
Time	1 date per month	1 date	
frequency			
Selected	01/26; 02/25; 03/10;		
dates	04/19; 05/26; 06/18;	07/05	
	07/18; 08/14; 09/23;		
	10/23; 11/22; 12/25		

Table 1. Characteristics of preprocessed images

3.3 In-field data

A field survey was carried out at the brownfield and reference site in November 2020 (Figure 1). Species and locations were used to construct a GIS with manual delineation on *BD ORTHO*® 25 cm resolution images.

From the initial fourteen identified species, 8 species, or assemblage of species for grasses, were considered as classes because of their predominance. The number of pixels extracted from images of each class is noted down in table 2.

Dominant genus or mixture of species	HS pixels		MS pixels	
	Ref.	Imp.	Ref.	Imp.
Platanus (Plat.)	2487	18771	18	106
Salix (Sal.)	321	1125	16	13
Populus (Pop.)	4046	1678	131	31
Acer (Ace.)	375	0	16	0
Quercus (Que.)	326	1265	12	12
Reynoutria (Rey.)	609	1022	8	8
Rubus (Rub.)	1567	4012	12	36
Grass mixture (Gra.)	625	1341	8	24

 Table 2. Predominant species on each area and corresponding pixel number for MS and HS images (Ref.: reference site; Imp.: impacted site).

On the HS image, the most represented classes were *Platanus*, with a total of 21258 pixels followed by *Populus* and *Rubus* (more than 5000 pixels), Grass mixture, *Reynoutria*, *Quercus*, *Salix* (more than 1000 pixels) and finally *Acer* with only 375 pixels. MS pixels per class were far below these numbers. Only *Platanus* and *Populus* classes were represented by more than 100 pixels. Other species only had a maximum of 48 pixels (*Rubus*) and a minimum of 16 pixels (*Acer* and *Reynoutria*). Even if *Acer* trees were only found in the reference area, this class was kept to evaluate false alarms.

4. METHOD DESCRIPTION

Non-parametric machine learning algorithms which usually performed well for species classification (Burai et al., 2015; Fassnacht et al., 2016; Erudel et al., 2017; Macyntre et al., 2020), such as Support Vector Machine algorithm (SVM), Random Forest (RF) and Regularized Logistic Regression (RLR) classifier, were compared. Feature extraction and selection methods, such as Principal Component Analysis (PCA) and Vegetation Indices (VI) were considered to solve the Hughes phenomenon and to increase species specificity (Hughes, 1968). Evaluation metrics (confusion matrix, overall, user's, producer's accuracies) were compared to define the most performant method. The analyse was led by class to characterize the contribution of each kind of information (spectral or temporal).

All classifications were performed using Python scikit-learn package (Pedregosa et al., 2011). The different stages of the method are detailed in sections 4.1 to 4.4

4.1 Training and testing sets build

It was proven that dependent training and testing sets led to biased results (Gewali et al., 2018). To ensure independence, three subsets were considered according to their corresponding area (reference site, impacted site, or entire area representing reference and impacted sites). This division allowed confronting different classification scenarios (Table 3).

First, training and testing sets were defined at the polygon scale, to ensure better independence, over the entire site with a ratio of 50%. Since polygons were different in number of pixels and illumination conditions, these classification scenarios were performed 5 times (Pelletier et al., 2019). Then, training was computed on polygons of the reference site and evaluation on those of the impacted site.

4.2 Feature selection

Feature selection was suited for each data type (MS and HS).

Principal component analysis (PCA) is a feature extraction method often used for the reduction of spectral bands. PCA reduces redundancy across features by projecting them into a lower-dimensional space in which the variance over the dataset is preserved (Jolliffe, Cadima, 2016). In our case, PCA was exploited both on time series of MS images, by computing it for each date, and on the HS image. The resulting features were a group of virtual bands highlighting variance over vegetation. Their variations underlined phenology changes with multitemporal exploitation (Macyntre et al., 2020). Components with higher noise levels were removed by visual inspection. 10, 20 or 40 principal components were retained for the HS image. The PCA applied on the MS multitemporal images led to 2 components to explain 98% of variance.

Spectral vegetation indices are arithmetic operations on spectral bands or their transformations. They provide information on vegetation characteristics, such as its physical state or its biochemistry, due to bands sensitivity to specific traits (chlorophyll content, biomass...) (Xue, Su, 2017). Exploited from a temporal perspective, they also allow to identify differences in phenology across vegetation (Beck et al., 2007). Here, different indices were tested on the time series of MS images. Vegetation indexes suited for Sentinel-2 temporal series and efficient for vegetation characterization in an impacted area, namely the Normalized Difference Vegetation Index (NDVI), the Plant Senescence Reflectance Index (PSRI), and the Inverted Red-Edge Chlorophyll Index (IRECI) were computed alone and combined (Fabre et al. 2020). Since only vegetation was classified in this study, two NDVI variations, the Green Normalized Difference Vegetation Index (GNDVI) (Gitelson, Merzlyak, 1998) and the Wide Dynamic Range Vegetation Index (WDRVI) (Gitelson, 2004), and an index linked with chlorophyll content in leaves (Lichtenthaler et al., 1996) were also tested. GNDVI is more sensitive to chlorophyll concentration than NDVI and adapted to tall and dense vegetation like trees (Gitelson, Merzlyak, 1998). WDRVI is another NDVI variation designed to be sensitive to changes in leaf area index (LAI) under conditions of high density. This led to several classification scenarios summarized in Table 3.

Images (number)	Training	Test	Data input
HySpex (1)	Reference	Impacted	Spectral sig. PC
	Entire area (50% of polygons)	Entire area (50% of polygons)	Spectral sig. PC
Sentinel-2 - monthly (12) -seasonal	Reference	Impacted	Spectral sig. PC VI
(4)	Entire area (50% of polygons)	Entire area (50% of polygons)	Spectral sig. PC VI

Table 3. Scenarios of classification (PC: Principal Components,
Spectral sig.: Spectral signature, VI: Vegetation Index).
Individual or combined VI are used.

4.3 Algorithms and parameters

A part of the pixels extracted from the training polygons was used to set parameters with grid search by 10-fold crossvalidation. These pixels represented 20%, 50%, 80% of the training polygons and are called in the next "sample size". Because of unbalanced classes, training weights were defined as inversely proportional to class frequencies to penalize mistakes on the minority classes.

4.3.1 Support Vector Machines (SVM)

SVM is a non-parametric classifier that relies on constructing a hyperplane that maximizes the margin of separation between two classes. To create that hyperplane, the original feature space is projected on a higher dimensional space where classes are linearly separable by a kernel function. Since these algorithms are designed for binary classifications, multiple class problems are managed with one-against-rest or one-against-one and voting strategies (Vapnik, 1998).

Two kernels were tested (linear, Gaussian (rbf)) with a multiclass strategy of one-against-rest. Many values of the C parameter, which corresponds to the regularization strength, were evaluated: $C \in \{1, 10, 100, 500, 1000, 5000, 10000\}$. The kernel size, defined by the γ parameter, was tested for $\gamma \in \{0.0001, 0.001, 0.01, 0.1, 1\}$.

4.3.2 Random forest (RF)

RF is a non-parametric ensemble algorithm. The main idea consists of using a set of regression trees, a weak classifier with high variance and low bias, created with a bagging approach (multiple draws of training samples with replacement) to improve classification performance. Nodes within a tree are defined to split data using the best predictor possible among a defined number of features. Then, the class decision is taken by averaging probabilities obtained within all trees (Breiman, 2001).

The number of decision trees in the forest was here set among $\{100, 200, 500, 1000, 2000, 5000\}$ with a number of considered features equal at maximum to the square root of the number of features and a maximum depth of the tree varying in $\{2, 3, 5, 10, 15, 20, 25, 30, 40\}$ (Belgiu, Dragut, 2016).

4.3.3 Regularized logistic regression (RLR)

RLR classifier is based on logistic regression with an additional regularization term, called the penalty. In this study, the penalties confronted were the ℓ 1-norm and the ℓ 2-norm. Logistic regression with ℓ 1 regularization, also called lasso regression, controls the selection of variables used during classification whereas ℓ 2 logistic regression, or ridge regression, handles collinearity between variables (Prospere et al., 2014). Many values of the C parameter, which corresponds to the regularization strength, were tested: C \in {0.01, 0.1, 1, 5, 10, 50, 100, 500, 1000, 5000}.

4.4 Classification accuracy evaluation

For each classifier, parameters with best overall accuracy (OA), which evaluates the number of correct classifications over all classifications, across the 10-fold cross-validation were selected and evaluated on the test set with several metrics based on a confusion matrix.

User's accuracy (UA), also called precision, can be defined as the fraction of relevant classes retrieved when produce's accuracy (PA), or recall, is the fraction of retrieved classes that are relevant. Both scores could be resumed with the F1 score (1). In the case of entire area scenarios, mean and standard deviation scores were calculated.

$$F1 = 2 * \frac{PA * UA}{PA + UA} \tag{1}$$

Results were then analysed regarding spectra derived from polygons. For each class, median spectra, and mean spectra were calculated.

5. RESULTS

5.1 HS classifications

5.1.1 Training performance

The overall accuracies over the training sets are given in Table 4. High accuracies were obtained for all scenarios using spectral signatures and with all classifiers. Changing sample sizes did not change parameters returned by the 10-fold cross-validation. Random forest was outperformed by other algorithms in these scenarios even with optimized parameters (Figure 2). Other classifiers were statistically similar in terms of accuracies and variance. The PCA led to a decrease of around 2 % of the global score regardless of the retained component number.



Figure 2. Mean performance with RF for different parameters (sample size = 20%; training: entire area).

Training	Entire area	Reference area
	(mean of 5 iterations)	
Classifier	Spectral	Spectral
	sig.	sig.
RF	92	91
SVM-linear	96	98
SVM-rbf	97	98
RLR-ℓ1	97	98
RLR-ℓ2	96	97

Table 4. Mean overall accuracies obtained during cross-validation for the different classifiers with sample size of 20%

5.1.2 Performance assessment

The overall accuracies over the testing sets are given Table 5. In all scenarios, RLR provided the best results in terms of accuracy whatever the training set. PCA led to a small decrease in classification performance (decrease between 2 and 5 % according to the classifier).

Site	Entire area	Impacted area
Classifier	Spectral	Spectral
	sig.	sig.
RF	87	61
SVM-linear	93	76
SVM-rbf	93	74
RLR-ℓ1	93	87
RLR-ℓ2	93	80

 Table 5. Overall accuracy of different classifiers for different HS scenarios.

With training on the entire area, all algorithms provided high accuracies. The main misclassifications were for the *Quercus* class, for which approximatively 30 % of pixels were classified as *Populus*. Figure 3 shows the intraclass standard deviation for the two classes furthest apart in Euclidean distance, *Populus* and *Reynoutria* (distance of 0.11 in reflectance) and the closest, *Populus* and *Quercus* (distance of 0.006 in reflectance). Due to the high intra-class variance (Figure 3), many overlaps were observed between classes, explaining the remaining errors.



Figure 3. Mean spectrum and intra-class standard deviation for the two most distant classes and the two closest classes in Euclidean distance.

Training algorithms over the reference area and evaluating them over the impacted one led to a general decrease of classification accuracy (ranging from 6% with RLR- ℓ 1 to 26% with RF). The most impacted classes were *Salix*, *Populus*, and *Rubus* (Table 6).

	Plat	Sal	Pop	Ace	Que	Rey	Rub	Gra
Plat	17656	0	348	6	0	15	285	14
Sal	3	107	882	0	0	0	133	0
Pop	92	10	1477	27	24	0	46	2
Ace	0	0	0	0	0	0	0	0
Que	19	0	569	0	675	0	2	0
Rey	36	1	45	4	2	784	125	25
Rub	108	23	117	16	75	81	3572	20
Gra	0	0	541	0	0	16	0	782

Table 6. Confusion matrix obtained with RLR- $\ell 1$ for assessment on the impacted area. See species in Table 1.

While *Salix* trees were well-classified in the first scenario, only 10% were then recognized, mainly due to confusion with *Populus*. This confusion can be explained by spectral differences observed between areas for these species (Figure 4).



Figure 4. Mean spectral signatures of *Salix* and *Populus* over the reference area (unfilled points) and over the impacted area (filled points).

Indeed, in the NIR domain (750-1300 nm), *Salix* average spectrum was under the one of *Populus* on the reference area. On the impacted area, their relative positions were exchanged. *Salix*'s mean spectrum was then above the *Populus* one. Moreover, *Salix* average spectrum corresponding to the impacted area was superimposed on that of *Populus* on the reference area for the entire SWIR domain. Same variations were observed for *Reynoutria* and *Rubus* on the whole spectrum. While the mean reflectance of *Reynoutria* was lower

than the one of *Rubus* in the reference area, it was higher in the impacted one. Some false alarms were observed for *Acer*, which was only present in the training area (Table 1). Finally, *Quercus*'s mean spectrum was extremely close to the *Populus* one on the reference area, explaining misclassifications obtained in all scenarios (average Euclidean distance of 0.0045 in reflectance while it was 0.015 on the impacted area).

5.2 MS classifications

Since numerous combinations were tested for multitemporal scenarios, only those with important results are presented in this section.

5.2.1 Training performance

Even if overall performance was under those obtained with HS scenarios, training still led to high accuracies (Table 7). However, changing the sample size and the classifier strongly impacted the results. In contrast to the HS scenario, classification based on logistic regression was less effective than those based on SVM and RF. To reach a performance of the order of those obtained with HS (around 90 % with a sample size of 20 %), it was necessary to use the RF or SVM classifier and an important training set size. Parameters returned during cross-validation were systematically the same as those issued from HS only with a sample size of 80 %.

Nonetheless, due to the small number of pure pixels available on Sentinel-2 spatial resolution, variance in term of parameters and scores were far more important than that of the HS training set while applied on 50 % of polygons of the entire area.

Training	Reference area				
Training size	20%	80%			
RF	70	86	91		
SVM-linear	65	78	83		
SVM-rbf	66	86	89		
RLR- ℓ1	58	70	71		
RLR- ℓ2	54	59	61		

Table 7. Mean overall accuracy obtained during cross-validation on reference area for different training sizes with monthly spectra.

5.2.2 Performance assessment

Application to all classes

The overall performance was greatly increased by limiting the date number to build the time series or extracting spectral characteristics.

The seasonal time series was created by selecting one date per season (03/10 for spring, 06/18 for summer, 09/23 for autumn and 12/25 for winter). During classification over the entire area, this date selection provided results 22 to 28% higher than those obtained with monthly time series. The progress achieved by extracting spectral features led to even better results, reaching up to 38 %. However, no progress was observed by combining temporal selection and spectral extraction (seasonal PCA, seasonal indices) on the entire dataset.

Neither spectral feature extraction method outperformed the other. Nevertheless, combining spectral vegetation indices conducted to higher accuracies than exploiting a single vegetation index (performance increase around 10 %).

This phenomenon was explained by the high differences across vegetation. The best combination was found with GNDVI, WDRVI, PSRI, and IRECI. GNDVI and WDRVI are general

vegetation indices, performing well on dense vegetation, with different dynamics. Also, IRECI is defined to estimate the slope of the red edge, a spectral characteristic directly linked to chlorophyll content (Frampton et al., 2013). On the contrary, PSRI is an index specifically designed for senescent vegetation (e.g., grass mixtures) sensitive to the ratio between carotenes and chlorophylls (Fabre et al., 2020). Using these 4 indices, 7 Sentinel-2 bands were exploited to highlight vegetation differences.

It was difficult to retain a classification method because the best method changed according to the classification scenario:

- RLR was retained with monthly indices,
- RF led to the best results applying on PCA,
- SVM and RLR- *l*1 had similar performance.

With training and evaluation on distinct areas, PCA results with RF and one image per month were slightly over those obtained with RF and monthly indices or RLR and dates selection. Exploited with spectral feature extraction, RF outperformed other algorithms while RLR provided the best results with all spectral features. As observed with HS classification, higher scores were obtained with evaluation over the entire area. Nonetheless, looking at the spectra and the training results (Table 7), it seemed that the reason was quite different and might be due to a lack of training pixels (Table 2). Classes represented by fewer pixels tended to be identified with lower precision in some classifications. For example, one of the five iterations of RF with monthly PCA applied on the entire area provided precision of 84% for Populus, 83% for Rubus, 80% for Platanus and grass mixture while the reached precisions were only 50% for Reynoutria, 35% for Salix, 25% for Quercus, and 12% for Acer.

Entire area								
(mean of 5 iterations)								
Scenario	Monthly	Seasonal						
	spectra	indices	PCA	spectra				
RF	33	67	71	54				
SVM-linear	34	68	61	60				
SVM-rbf	33	65	68	61				
RLR-ℓ1	36	72	61	60				
RLR-ℓ2	32	67	68	56				
	Impacted area							
RF	25	65	57	43				
SVM-linear	44	43	34	59				
SVM-rbf	22	35	39	45				
RLR-ℓ1	43	58	32	62				
RLR-ℓ2	45	39	40	60				

 Table 8. Overall accuracy of different classifiers for different multitemporal MS scenarios. Combination of PSRI, GNDVI, WRDVI and IRECI indices.

Application to a selection of classes

To analyse if temporal information could improve performance for the most difficult classes to discriminate with HS data (F1score under 85%), supervised classification methods were applied on S-2 data to discriminate these classes in the case of training on the reference area and evaluation on the impacted one. The selected classes were *Populus*, *Acer*, *Quercus*, and Grass mixture.

As observed with all species, both temporal selection and spectral feature extraction improved results. The temporal selection was still outperformed by spectral feature extraction (OA around 10 % higher). No algorithm stood out in these scenarios.

In that case, combining selection on both dates and spectral features greatly improved overall accuracy in comparison to the performance obtained for all the classes, leading to OA ranging from 62 to 75 % (increase between 1 % and 23 %). Again, PCA results were slightly over those obtained with spectral indices (difference between 0 and 8 %). PCA had indeed the advantage of being more integrative than spectral indices.

Encouraging results were obtained to distinguish *Salix* and *Populus*. For example, 69% (10% with HS) and 74% (88% with HS) of precision scores were returned with RLR- ℓ 1 and monthly PCA on *Salix* and *Populus*, respectively (Table 9). Variations of the first principal component of different remaining classes on both areas are shown in Figure 5. As illustrated, 1-PC variations of *Quercus* and grass mixture appeared to be slightly affected by the change of area. While the trends of Quercus 1-PC were the same as in the reference area, Grass mixture 1-PC variations seemed affected from autumn to winter. During the same period, strong variations were constated for both *Salix* and *Populus* between the areas. Nevertheless, the respect of the global trend during other seasons seemed to allow distinguishing them correctly.

	Sal	Pop	Ace	Que	Gra
Sal	8	5	0	0	0
Pop	0	24	0	7	0
Ace	0	0	0	0	0
Que	0	2	0	10	0
Gra	0	0	0	8	16

Table 9. Confusion matrix obtained with RLR- l1 on reduced
classification scenario with seasonal PCA



Figure 5. Seasonal evolution of the first principal component for *Salix, Populus, Quercus* and Grass mixture.

6. DISCUSSION

Obtained performance with both HS and MS images shows that classification at the species level is feasible in a heterogeneous ecosystem composed of various vegetation type. Former works on such an environment only used one kind of instrument and have resulted in a maximum overall accuracy of 76 % (Kamal, Phinn, 2011; Macyntre et al., 2020).

Other species are present in both reference and impacted areas and should be considered in the following for mapping species over the entire site. Moreover, the number of samples must be raised at Sentinel-2 resolution to increase training set size (Burai et al., 2015). The management of class imbalance must be further explored (Zhu, 2005; He, Garcia, 2009; Erudel et al., 2017). A rejection class will then be considered to treat the entire area (Aval, 2018).

Many studies concluded that SVM and RF applied on HS data provided high-quality species mapping. Nonetheless, in our case, RLR led to better species identification used with HS resolutions. If a similar study at the ground level led to the same conclusion (Erudel et al., 2017), to our knowledge, classification topics based on logistic regression are still rare and need more attention. The recent general trend in vegetation classification is focused on deep learning methods (Zhang, 2020). However, these methods require a high number of samples, difficult to obtain on small areas with the usual metric to decametric spatial resolutions. RLR could be an alternative in such cases.

Our results show that feature selection is a crucial step. If not considered in this study, several works have shown that using spatial information, by performing classification at polygon scale or by adding information from neighbouring pixels as virtual bands, greatly improved performance both with MS (Denisova et al., 2019) and HS images (Dian et al., 2015). If feature extraction obtained by PCA did not perform well with our HS scenario, several studies, both on tree species (Dabiri, Lang, 2018) and complex alkali vegetation (Burai et al., 2015) conducted to the contrary conclusion with metric spatial resolution. More spectral and spatial features should be investigated in future works. Spectral indices applied on HS images had proven in the literature to improve classification accuracy (Erudel et al., 2017; Maschler et al., 2018). In the next future, our work will consider optimised selection of spectral indices.

On the contrary, feature reduction methods improved results on MS scenarios both temporally and spectrally. It was noticed during the study that changing the date or the selected vegetation indices could lead to significant changes. Thus, it is necessary to work not only on the type or number of features but also on their different combinations. Only seven spectral indices have been tested but it is not impossible that other indices could increase the performance of the classification. Specific metrics were designed to directly deal with phenology in a multitemporal way by calculating temporal correlation (Griegorova et al., 2020). In our study, only per-date features were computed. More works on spectro-temporal features may improve MS classification accuracies.

This study indicates that using both HS and Multitemporal MS images could allow classifying species even in the case of high intra-class variance. The hierarchical classification scheme used greatly improved the final precision obtained for the different classes. A next step would be to use fusion methods to exploit both HS spectral and spatial information and multitemporal MS temporal information in the same classification scheme.

When the reference and the impacted area were distinguished, the precision decreases. This proved that even if species were identical, different spectral characteristics, that might be related to their health status, were measured. Subsequently, the species will be related to the anthropogenic impacts to analyse the species distribution on impacted soil and their health status.

7. CONCLUSION

The results demonstrate that using both HS and Multitemporal MS images could help to distinguish different species in a complex environment.

If HS classification performed well with training on both impacted and reference areas, changes in spectral signatures potentially due to anthropogenic impact led to a decrease in accuracy when training was done only on the reference area and evaluation on the impacted one. Regularized logistic regression showed higher performances than SVM or RF dealing with HS image.

With the time series of MS images, feature selection was a necessary step to perform classification. No feature selection

method stood out in this study. Provided multitemporal information allowed to distinguish potentially impacted species that remained difficult to classify under a single date study.

ACKNOWLEDGMENTS

The authors gratefully acknowledge P. Déliot (ONERA) and L. Poutier (ONERA) for their implication in image acquisitions and preprocessing. This work was co-funded by TOTAL R&D in the frame of DEMETER Project.

REFERENCES

Aval, J., 2018. Automatic mapping of urban tree species based on multi-source remotely sensed data. (Doctoral dissertation, UNIVERSITE DE TOULOUSE).

Beck, P. S. A., Jönsson, P., Høgda, K. A., Karlsen, S. R., Eklundh, L., Skidmore, A. K., 2007. A ground validated NDVI dataset for monitoring vegetation dynamics and mapping phenology in Fennoscandia and the Kola peninsula. *International journal of remote sensing*, 28(19), 4311-4330.

Belgiu, M., Dragut, L., 2016. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.*, 114, 24-31

Breiman, L., 2001. Random Forests. Mach. Learn., 45, 5-32.

Brigot, G., Colin-Koeniguer, E., Plyer, A., Janez, F., 2016. Adaptation and Evaluation of an Optical Flow Method Applied to Coregistration of Forest Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 9 (7), 2923–2939.

Burai, P., Deak, B., Valko, O., Tomor, T., 2015. Classification of herbaceous vegetation using airborne hyperspectral imagery. *Remote Sensing*, 7(2), 2046-2066.

Clark, M. L., Buck-Diaz, J., Evens, J., 2018. Mapping of forest alliances with simulated multi-seasonal hyperspectral satellite imagery. *Remote Sensing of Environment*, 210, 490-507.

Clark, M. L., 2020. Comparison of multi-seasonal Landsat 8, Sentinel 2 and hyperspectral images for mapping forest alliances in Northern California. *ISPRS Journal of Photogrammetry and Remote Sensing.*, 159, 26-40.

Dabiri, Z., Lang, S., 2018. Comparison of independent component analysis, principal component analysis, and minimum noise fraction transformation for tree species classification using APEX hyperspectral imagery. *ISPRS International Journal of Geo-Information*, 7(12), 488.

Dalponte, M., Ørka, H. O., Gobakken, T., Gianelle, D., Næsset, E., 2012. Tree species classification in boreal forests with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5), 2632-2645.

Denisova, A. Y., Kavelenova, L. M., Korchikov, E. S., Prokhorova, N. V., Terentyeva, D. A., Fedoseev, V. A., 2019. Tree species classification in Samara Region using Sentinel-2 remote sensing images and forest inventory data. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 16(4), 86-101.

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B3-2021 XXIV ISPRS Congress (2021 edition)

Dian, Y., Li, Z., Pang, Y., 2015. Spectral and texture features combined for forest tree species classification with airborne hyperspectral imagery. *Journal of the Indian Society of Remote Sensing*, 43(1), 101-107.

Erudel, T., Fabre, S., Houet, T., Mazier, F., Biottet, X., 2017. Criteria Comparison for Classifying Peatland Vegetation Types using In Situ Hyperspectral measurements. *Remote Sensing*, 9(7), 748.

Fabre, S., Gimenez, R., Elger, A., Rivière, T., 2020. Unsupervised Monitoring Vegetation after the Closure of an Ore Processing Site with Multi-Temporal Optical Remote Sensing. *Sensors*, 20, 4800

Fassnacht, F. E., Neumann, C., Förster, M., Buddenbaum, H., Ghosh, A., Clasen, A., Koch, B., 2014. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6), 2547-2561.

Fassnacht, F. E., Latifi, H., Sterenczak, K., Modzelewsk, A., Lefsky, M., Waser, L.T, Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment.*, 186, 64-87.

Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variable in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing.*, 82, 83-2

Gewali, U. B., Monteiro, S. T., Saber, E., 2018. Machine Learning Based Hyperspectral Image Analysis: A Survey. *arXiv e-prints*, arXiv-1802.

Gitelson, A., Merzlyak, M., 1998. Remote Sensing of Chlorophyll Concentration in Higher Plant Leaves. *Advances in Space Research.*, 22, 689-692.

Gitelson, A. A., 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of plant physiology*, 161(2), 165-173.

Griegorova, O., Brovkina, O., Saidov, A., 2020. An original method for tree species classification using multitemporal multispectral and hyperspectral satellite data. *Silva Fenninca*, 54, 2.

Guidici, D., Clark, M. L., 2017. One-Dimensional convolutional neural network land-cover classification of multi-seasonal hyperspectral imagery in the San Francisco Bay Area, California. *Remote Sensing*, 9(6), 629.

He, H., Garcia, E.A., 2009. Learning from Imbalanced Data, *IEEE Transactions on Knowledge and Data Engineering*, 21, 9, 10.1109/TKDE.2008.239.

Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. *Trans. Inform. Theory*, 14, 55–63.

Jolliffe, I. T., Cadima, J., 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.*, 374(2065), 20150202.

Kamal, M., Phinn, S., 2011. Hyperspectral data for mangrove species mapping: A comparison of pixel-based and object-based approach. *Remote Sensing*, 3(10), 2222-2242.

Lassalle, G., Elger, A., Credoz, A., Hédacq, E., Bertoni, G., Dubucq, D., Fabre, S., 2019. Toward Quantifying Oil Contamination in Vegetated Areas Using Very High Spatial and Spectral Resolution Imagery. *Remote Sensing.*, 11, 2241.

Lichtenthaler, H. K., Gitelson, A., Lang, M., 1996. Non-Destructive Determination of Chlorophyll Content of Leaves of a Green and an Aurea Mutant of Tabacco by Reflectance Measurements. *Journal of Plant Physiology.*, 148 (3-4), 483-493.

Macyntre, P., Niekerk, A. V., Mucina, L., 2020. Efficacy of multi-season Sentinel-2 imagery for compositional vegetation classification. *International Journal of Applied Earth Observation and Geoinformation.*, 85, 101980.

Maschler, J., Atzberger, C., Immitzer, M., 2018. Individual tree crown segmentation and classification of 13 tree species using airborne hyperspectral data. *Remote Sensing*, 10(8), 1218.

Onyia, N., Balzter, H., Berrio, J.C., 2019. Spectral Diversity Metrics for Detecting Oil Pollution Effects on Biodiversity in the Niger Delta, *Remote Sensing.*, 11 (22), 2662

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825-2830.

Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sensing.*, 11(5), 523.

Prospere, K., McLaren, K., Wilson, B., 2014. Plant Species Discrimination in a Tropical Wetland Using in Site Hyperspectral Data. *Remote Sens.* 6, 8494-8523

Richter, R. Reu, B., Wirth, C., Doktor, D., Vohland, M., 2016. The use of Airborne Hyperspectral Data for Tree Species Classification in a Species-rich Central European Forest Area. *International journal of applied earth observation and geoinformation.*, 52, 464-474.

Smith, G. M.; Milton, E. J., 1999. The Use of the Empirical Line Method to Calibrate Remotely Sensed Data to Reflectance. *Int. J. Remote Sens.*, 20 (13), 2653–2662.

Theia Data and Services Center for Continental Surfaces. Available online: https://www.theia-land.fr/pole-theia-2/ (accessed in February 2021)

Vapnik, V.V., 1998: *Statistical Learning Theory. Volume 1.* Wiley: New York, NY, USA.

Xue, J., Su, B., 2017. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal* of Sensors., 2017.

Zhu, X. J. 2005: Semi-supervised learning literature survey. Technical Report 1530, Department of Computer Sciences, University of Wisconsin-Madison: Madison, WI, USA.