

OPTIMAL DATES FOR DECIDUOUS TREE SPECIES MAPPING USING FULL YEARS SENTINEL-2 TIME SERIES IN SOUTH WEST FRANCE

N. Karasiak^{1*}, M. Fauvel², J.-F. Dejoux², C. Monteil¹, D. Sheeren¹

¹ DYNAFOR, Université de Toulouse, INRAE, Castanet-Tolosan, France
nicolas.karasiak@inrae.fr, david.sheeren@ensat.fr

² CESBIO, Université de Toulouse, CNES/CNRS/INRAE/IRD/UPS, Toulouse, France

Commission III, WG 10

KEY WORDS: sentinel-2, satellite image time series, tree species, forest, map, biodiversity, spatial autocorrelation, France

ABSTRACT:

The free to use Sentinel-2 (S2) sensors with 5-day revisit time at high spatial resolution in 10 spectral bands is a revolution in the remote sensing domain. Including 6 spectral bands in the near infrared, with 3 dedicated for the red-edge (where the vegetation significantly increases), these European satellites are very promising for mapping tree species distribution at a national scale. Here, we study the contribution of three one-year S2 Satellite Image Time Series (SITS) for mapping deciduous species distribution in the southwest of France. The annual cycle of vegetation (called phenology) can contribute to the identification of tree species. For some specific dates, species can have different phenological behaviours (senescence, flowering...). To train and validate the maps, we used the Support Vector Machine algorithm with a spatial cross-validation method. To train the algorithm with the same number of samples per species, we decided to undersample each class to the smallest class using a K-means clustering method. Moreover, a Sequential Feature Selection (SFS) has been implemented to detect the optimal dates per species. Our results are promising with high accuracy for Red oak and Willow (average score of the three one-year respectively $F_1 = 0.99$, $F_1 = 0.94$) based on the optimal dates. However, it appears that the performances when using the each full SITS are far below the optimal dates models (average $\Delta F_1 = 0.32$). We did not find, except for Willow and Red oak, that the optimal dates were the same for each year. Perspectives is to find an algorithm robust to temporal or spectral noise and to smooth the time series.

1. INTRODUCTION

In the context of global warming and biodiversity loss, forests are one of the most important ecosystems to protect (Thompson et al., 2011). The diversity of the tree species and their age is one of the main criteria for estimating potential biodiversity indicator (Larrieu, Gonin). Despite its crucial role in preserving biodiversity and mitigating global warming, this ecosystem is affected by two main problems. On the one hand, tree species are prone to diseases. Indeed from 2003 to 2012, more than 85 millions hectares of forest have been affected by insect pests (van Lierop et al., 2015). On the other hand, the share of the world's forests is decreasing year after year and is being replaced by agriculture or grazing with serious consequences on biodiversity (Brockerhoff et al., 2008). These previous authors also pointed out that the replacement of natural or semi-natural forests by forest plantations is always preferable to other land covers, with the exception of the original forest. Whether it is degradation or deforestation, there are solutions to slow down these losses. For example, to prevent natural forest diseases, the higher the number of tree species in a forest, the greater the resilience (Guo et al., 2019). By protecting the trees, we also protect the flora and fauna biodiversity.

Hence, knowing the spatial distribution of trees is important whether for foresters, climatologists, or ecologists. For decades, national maps have been produced by photo-interpretation using high-resolution optical sensors. It all began with aerial photography from aircraft (Smith, 1976), and then with the development of very high spatial resolution satellite sensors, maps were produced at a large scale. In 2019, the National Institute of

Geographic and Forest Information (IGN) published its second version of the main tree species in France (BDForêt[®] v2). It took nearly 20 years because most of the work is manual, carried out by photo-interpretation of false-color images (the infrared band is put in the red channel).

To switch from manual mapping to a supervised classification, it is essential to rely on phenological events of tree species. Indeed, deciduous trees show very marked events throughout the year such as leaf flush, flowering or senescence (leaf fall and leaf coloration) (Badeau et al., 2017). These events modify the reflectance of the forest during the year by influencing the biochemical or biochemical attributes of the canopy (Yang et al., 2016; Miller et al., 1991). SITS can help to monitor the cycle and should increase the spectral separability between deciduous tree species (Fassnacht et al., 2016).

Despite huge progress in machine learning, it appears that mapping tree species distribution over large areas from space is still an opening question and so a challenging task (Fassnacht et al., 2016). In order to have an up-to-date and large scale map of tree species, using sensor with high spatial, temporal and spectral resolution such a Sentinel-2 is highly recommended (Fassnacht et al., 2016).

While some works have already investigated the potential of Sentinel-2 to map temperate forest species with cloud-free images (Grabska et al., 2019; Bolyn et al., 2018; Immitzer et al., 2019), only few authors used a dense time series of satellite images (SITS) with a cloud detection and filtering protocol (Sheeren et al., 2016; Karasiak et al., 2017). Typically, SITS-based works use lower spatial resolution imagery such

* Corresponding author

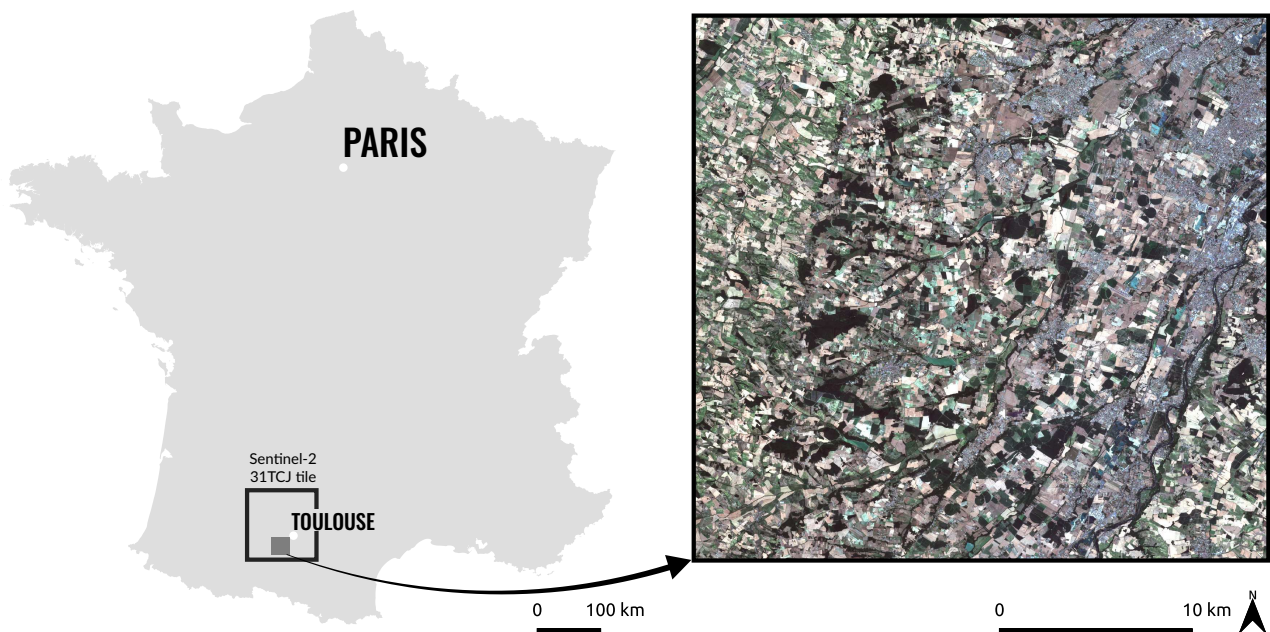


Figure 1. Location of the study area in France with the size of the Sentinel-2 31TCJ tile. On the right, the 2019-07-25 image of Sentinel-2 processed in level 2A with flat reflectance.

as that from the Modis sensor (Yan et al., 2015) or Landsat-8 (Pasquarella et al., 2018). SITS using higher spatial resolution sensors are not focused on tree species distribution but on land-cover mapping (Inglada et al., 2017).

Results between studies are often contradictory, both in terms of optimal dates and best predicted species. Besides, estimation of the quality seems to be very sensitive to spatial autocorrelation between the training and validation set (Karasiak et al., 2019) but it is rarely taken into account.

Because of specific disruptions that may appear in a single year (clouds, diseases, climate events such as drought...), there is a need to use several independent years to have a better understanding of the results. This should provide an understanding of the stability of the optimal dates for mapping tree species distribution. However, to the best of our knowledge, the identification of key dates for each available deciduous species has not been studied, neither using several years, neither using a dense SITS.

The main objectives of our research are:

1. To assess the potential of full years Satellite Image Time Series for mapping deciduous tree species.
2. To identify the best dates for each species.
3. To evaluate the accuracy stability among three independent years.
4. To identify the most suitable species for country scale mapping

5. To understand the quality differences between the full time series and the model using only the best dates.

2. MATERIALS

The aim of this study is to assess the contribution of optimal dates in the mapping of deciduous species. In order to map at large scale, we need to have similar temporal sampling across a whole country and it is therefore not possible to select only cloudless dates for each part of the territory. The dimension of the data (number of dates and spectral bands) must be the same all over the country.

2.1 Study area

The study area is located in the southwest of France, below the city of Toulouse (Figure 1). It is part of the Western European broadleaf forests going from this study area to finish at Dresden (Digital map of European ecological regions, 2000). Our study area is covered by about 10% of forests as the land is dominated by crops. The climate is sub-Atlantic characterized by sunny autumns, hot dry summers, and mild rainy winters (2018 climate data: average temperature = 15.1°C ; precipitation = 700 mm).

2.2 Satellite Images

Images came from Sentinel-2 satellite and are composed of four spectral bands at 10-m spatial resolution

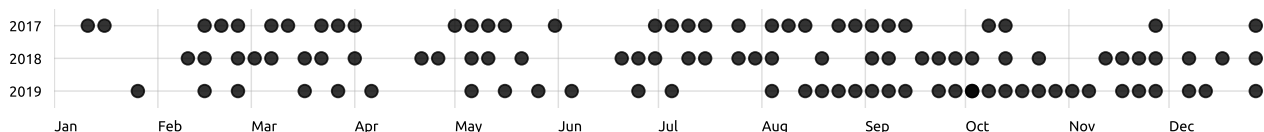


Figure 2. Acquisition dates of Sentinel-2 images for year 2017, 2018, and 2019. Sentinel-2 images are not sampled evenly over time due to excessive cloud cover. This means that some level 1C images were not processed at level 2A by the Theia pole.

Species	Size	Stands
Broadleaf		
Silver birch (<i>Betula pendula</i>)	67	3
Oak (<i>Quercus robur/pubescens/petraea</i>)	247	12
Red oak (<i>Quercus rubra</i>)	205	9
Aspen (<i>Populus spp.</i>)	139	7
European ash (<i>Fraxinus excelsior</i>)	130	5
Black locust (<i>Robinia pseudoacacia</i>)	88	9
Willow (<i>Salix spp.</i>)	52	4
Eucalyptus* (<i>Eucalyptus spp.</i>)	143	5
Conifer		
Corsican pine (<i>Pinus nigra subsp. laricio</i>)	201	9
Maritime pine (<i>Pinus pinaster</i>)	157	8
Black pine (<i>Pinus nigra</i>)	52	2
Silver fir (<i>Abies alba</i>)	53	4
Douglas fir (<i>Pseudotsuga menziesii</i>)	50	8
Cypress (<i>Cupressus</i>)	42	1

*the only sempervirent broadleaf species in our study area.

Table 1. List of tree species with their sample size which is equivalent to the number of Sentinel-2 pixels ($n = 1626$). In **bold**, deciduous species used to estimate the contribution of each date to the time series. The number of forest stands is based on the French National Forest Inventory (IGN BDForêt® v.2)

(Blue: 458–523 μm , Green: 543–578 μm , Red: 650–680 μm , Near InfraRed: 785–899 μm) and 6 bands at 20-m going from the red-edge which is the part of electromagnetic spectrum just after the red and before the near infrared (re1: 698–713 μm , re2: 733–748 μm , re3: 773–793 μm) to the near and mean infrared (NIR: 1565–1655 μm , MIR: 2100–228 μm). Radiometric resolution is 12 bits and images were acquired with a field of view of 100 km. As our study area is located on the middle of the Sentinel-2 tile (from west to east) no Bidirectional Reflectance Distribution Function (BRDF) effect should remain as it appears more on image borders.

The S2 images were downloaded on the pole Theia for year 2017, 2018 and 2019 (<http://www.theia-land.fr/en/products/sentinel-2>). These data are available directly at level 2A (i.e. Top-of-canopy reflectance products orthorectified) and they include a cloud and shadow mask (Baetens et al., 2019). These masks came from the MAJA algorithm which includes atmospheric correction by combining multi-temporal and multispectral criteria in order to estimate the optical thickness of aerosols (Hagolle et al., 2015). If the image was not computed by Theia, it means that the product was not valid (it can be due to too many clouds). Thus, 37 dates were acquired for 2017 and 2018, 32 for 2019. Each time series has different temporal resolution due to the clouds over our study area (Figure 2).

2.3 Forest mask

A forest mask produced for the Haute-Garonne department in 2018 by the IGN, BDForêt® v.2, was used to select forest pixels in the SITS (i.e. forest stands with a minimum area of 0.5 hectares) and to exclude non-forested areas. Although the BDForêt® v.2 was released in 2018, it was made by photo-interpretation using aerial photographs from 2013 (IGN BDOOrtho®). To update the mask and to avoid forest cuts in the ground references, we manually modified it to keep only forest stands that were uncut till the most up-to-date 2018 very high resolution image BDOOrtho®.

2.4 Ground references of tree species

Four field surveys were conducted in the 72 main forests between November 2013 and January 2017 to identify the forest species in the study area. Only dominant broadleaf and coniferous species location were gathered. To ensure data consistency, references were acquired from a Garmin GPSMap 62st GPS receiver (3-5 m accuracy) in the centre of an area covering approximately 900 m² (i.e. nine contiguous pixels from the Sentinel-2 satellite). Only the pixels in the center of these areas were used for classification.

A total 1626 field references was collected for 14 species, including 8 deciduous and 6 coniferous (Table 1). For some species, identification was limited to the genus because of the existence of many cultivars (as for poplar) or the difficulty in determining the exact species (oak, willow or eucalyptus). The class distribution is imbalanced, from 42 cypress and 52 willow samples to 247 oak and 205 red oak samples. However, this dataset reflects the abundance of species in these forests.

3. METHODS

3.1 Pre-processing of the SITS

For each of the three years, a SITS has been computed using linear interpolation when a cloud or cloud shadow was detected by the MAJA algorithm (Section 2.2). The process was performed using the time series gap-filling function (Inglada, 2016) of the Orfeo ToolBox (Grizonnet et al., 2017). Since 2016, Inglada et al. (2017) has been using this prep-processing of the SITS to map French land cover from Sentinel2 every year.

3.2 Undersampling of the reference data

As our set of reference samples reflect the reality of the forests, the effective per species is therefore very imbalanced. The two main species, Oak (*Quercus robur/pubescens/petraea*) and Corsican pine (*Pinus nigra subsp. laricio*) have more than 200 samples each, contrary to Willow (*Salix spp.*) and Cypress (*Cupressus*) which have about 50 samples. Training samples for the two first cited species tend to be 5 to 6 times higher than the rarest species in our dataset.

For each species, we decided to undersample our dataset to the size of the smallest class using their feature representations (i.e. spectral bands) (Jain et al., 1999). A simple approach is to downsize using a random selection by fixing the number of desired output samples. However, to keep the richness of our small dataset and not to lose crucial information that could be under-represented, we preferred to use a K-means to group samples from their similar features (Hartigan, Wong). This technique is recommended and performs very well with small scale datasets (Lin et al., 2017). In concrete terms, European ash (*Fraxinus excelsior*) with 130 samples has many more samples than cypress ($n=42$), which is the minority class. So, the undersampling algorithm generated a K-means with 42 clusters and used the coordinates of these cluster centroids as the new European ash samples. This processing was handled by the *imbalanced-learn* library (Lemaître et al., 2017)

3.3 Feature selection

We implemented a feature reduction approach (Green et al., 1988) called forward Sequential Feature Selection (SFS) (Whitney, 1971). This method is an interesting way to know the most

important dates in a time series. Also, keeping the whole information (i.e. all the features) will lead to the Hughes phenomenon, knowing that the larger the dimension the lower the quality (Hughes, 1968). A previous study has shown the advantage of reducing the size of the input features when using dense SITS (Karasiak et al., 2019).

To estimate the contribution of each date to classify tree species, we learnt for each species a dedicated model using the SFS. Selection of dates is made using the ten spectral bands per date of Sentinel-2. The addition of dates is done iteratively, one by one, and the best date is added to an initially empty set. This protocol has been used systematically to evaluate the performance of the model in predicting the different deciduous tree species. F_1 score was the metrics used as criteria to evaluate each date for each species. The combination of dates with the highest score per species will be presented in the results part.

3.4 Validation by spatial cross-validation

Every model was learned and evaluated using cross-validation. Given the possible optimistic bias related to spatial autocorrelation in the datasets (Pugh, Congalton; Hammond, Verbyla; Mannel et al., 2011), we opted for a spatialized leave-one-out cross-validation method (Le Rest et al., 2014; Karasiak et al., 2019).

Spatial cross-validation involves selecting validation samples that are spatially separated by a given distance from the training samples. To get the distance, Moran's Index (Moran, 1950) was computed on each spectral band by varying lineary the distance from 1 pixel (10-m) to 50 pixels (500-m). Pixels of non forests were masked using our forest/non forest mask made from 2018 very high resolution images (See section 2.3). The more the variability of the defined neighborhood equals the total variability of the data set the more Moran's I will be close to zero. This metrics gives us an index ranging from -1 to 1, where 1 corresponds to the maximum of spatial autocorrelation, and 0 the ideal case without autocorrelation. Considering that below Moran's $i = 0.2$, the spatial autocorrelation can be considered insignificant (Dale, Fortin), we've rounded up the average value where Moran's $i \leq 0.2$ in the three SITS and we found a distance of $d = 340$ -m. This means that for each species, the validation sample is at least distant by 340-m from the training samples.

The Spatial Leave-One-Out cross-validation is composed of 42 folds, as many as samples per class after undersampling. The results of the 42 predictions is averaged to have a global estimation of the error considering the F_1 value of the predicted species.

3.5 Classification protocol

For each year we built as many models as deciduous species to map. In combination with an undersampling and spatial cross-validation strategy, the Support Vector Machine (SVM) algorithm (Vapnik, 1998) was used to train our models. This classifier is known to perform very well with few samples per class and a high number of features (Melgani, Bruzzone; Mountrakis et al., 2011). In this study, we selected the Radial Basis Function (RBF) kernel which is the most frequently used and has already been proven to be effective in the case of similar classification problems (Kavzoglu, Colkesen; Ferreira et al., 2016). The learning process was computed using the *scikit-learn* python library Pedregosa et al. (2011), and the vector of features

were standardized (i.e. centering and scaling to unit variance) before the training process. The parameters of SVM fitted were the regularization parameter (C) know as the penalty parameter and the kernel bandwidth (γ) were tuned by cross-validation. The grid search was constructed with $C = \{0.01, 0.1, \dots, 1^{10}\}$ and $\gamma = \{1^{-9}, 1^{-8}, \dots, 1^3\}$.

This work is based on the python library we developed: *Museo ToolBox*(Karasiak, 2020). This library contains the spatialized cross-validation, the Sequential Feature Selection approach and many more. Moreover, it is fully compatible with the *scikit-learn* library.

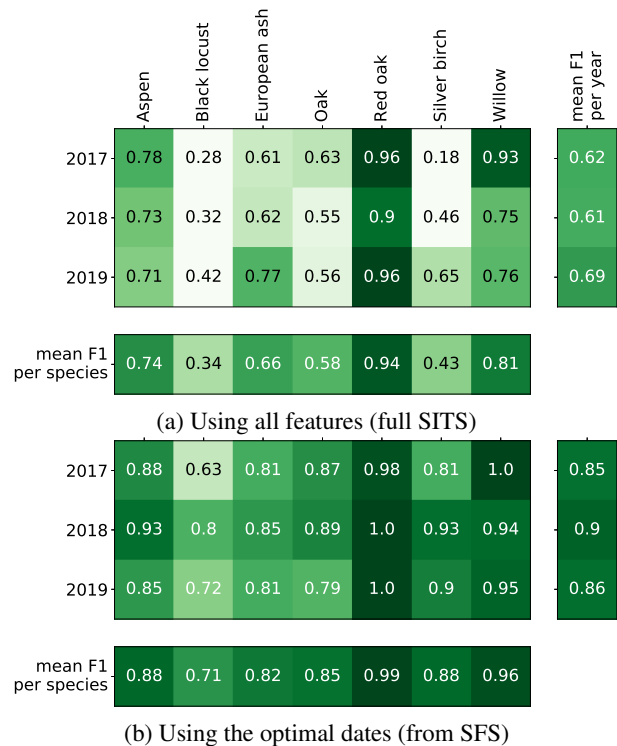


Figure 3. Best F_1 score per species and per year based on (a) all the features, or (b) the optimal dates after a Sequential Feature Selection.

4. RESULTS

In this section we analyse the performances of each model dedicated to a species and their different results when using all features (full SITS) or the optimal number of dates from SFS. The quality of each model is quantitatively measured using the F_1 score metrics of the selected species. The F_1 score for the optimal features per species is the combination of dates having the highest performance.

4.1 Optimal dates versus all features

The classification performances of the three years are presented for each approach in Figure 3. Generally speaking, the performances between the three years were relatively similar. However, there is a significant difference in quality between the two methods. The average F_1 of the seven deciduous species when using the full SITS is 0.64 while with the SFS method the results are 0.23 points higher, i.e. 0.87. For both datasets, the F_1 differences between the best year and the worst year is about 0.05 F_1 . In 2019, when the algorithm was trained using all the dates,

the best deciduous average F_1 was 0.69 while learning with a reduced dataset using SFS was 0.86.

Looking at the species for both datasets Red oak has, except once, the highest accuracy ($F_1 = 0.99$ with optimal dates, $F_1 = 0.93$ with the full SITS) followed by willow ($F_1 = 0.96$, $F_1 = 0.81$ for respectively optimal dates and full SITS). Black locust and Silver birch seem to be particularly affected when learning with the full SITS and loses about 0.4 of F_1 points. The comparison between the two methods show a high difference in the quality evaluation ($\Delta F_1 = 0.32$). However, the best or less well discriminated species remained the same regardless the method.

4.2 Optimal dates

In the following sections, we will only detail the results of the optimal dates approach based on the SFS since it has shown the best results and allows us to identify the most important dates for each species (Section 3.3).

The F_1 contribution per date was computed by subtracting the highest score with n dates by the highest score using $n-1$ dates. Generally speaking, only a small number of dates is needed to reach the best quality per species. With about five dates, each species reaches its maximum F_1 or was really close (± 0.02) (Figure 4).

4.2.1 Similar trends per year The only common and close date selected for the three years is in the end of July or beginning of august and only concerns the Willow trees. This tree systematically reach a very high accuracy with only one date in the middle of the summer ($F_1 = 0.9$). When looking at the Red oak species, selected dates are in early november or late november for respectively 2017 and 2018 but may 1 in 2019.

The other trend is for 2017 and 2018 where the first date in F_1 contribution for Oak and Silver birch is in early or end november. Also the F_1 contribution is high with only one date for Oak ($F_1 = 0.6$) and for Silver birch ($F_1 = 0.77$).

4.2.2 Unstable temporal selection The species which are the most subject to have different dates selected among years are Black locust and European ash. However, these two species share a similar pattern: in 2017 their key date is in october/november, in 2018 in march/april and in 2019 for the middle of summer.

In summary, it is tricky to say that a specific date or season is more important than another when looking at algorithm choices over several years in a row. A general trend seems to suggest that late fall or mid-spring may be the most appropriate times to maximize the gaps between species, but not for every species and every year.

5. DISCUSSIONS

5.1 Optimal dates comparison

Thanks to the three one-year of SITS, it is difficult to say whether a date is more important than another. If Immitzer et al. (2019) found broadleaf species were more able to be detected in April, May and June, we can't confirm their conclusion as for some years it seems to be only november, or for some species (Willow) only end of June or early July. Grabska et al.

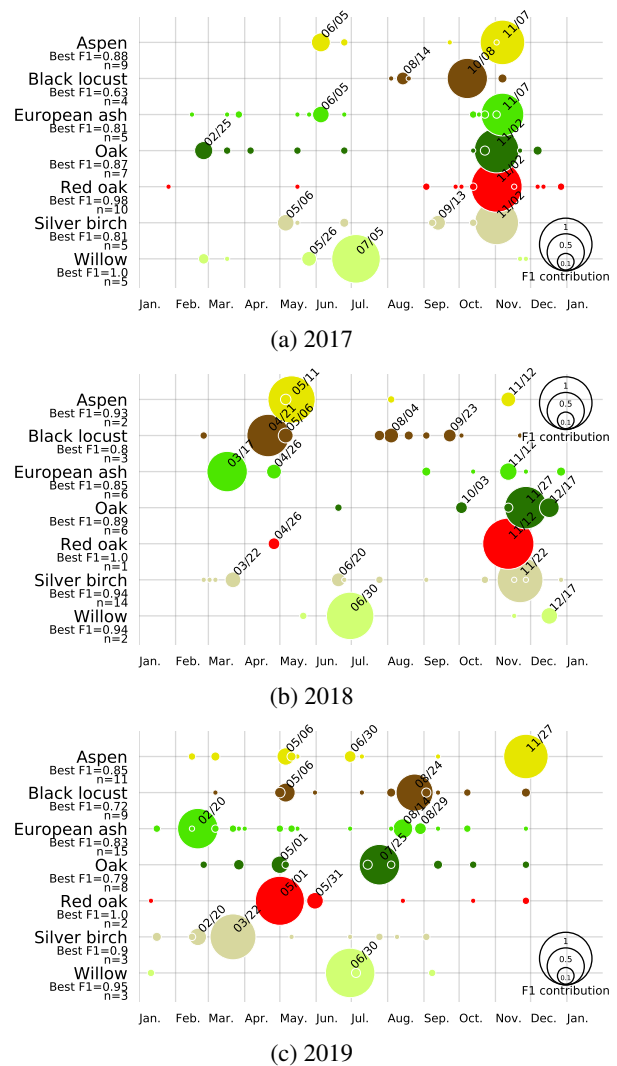


Figure 4. Optimal dates for each species in year (a) 2017, (b) 2018 and (c) 2019. The F_1 contribution per date was computed by subtracting the highest score with n dates by the highest score $n-1$ dates. The larger the circle, the more important the date in the Sequential Feature Selection. If negative score, circle does not appear. Date is only written if the contribution is superior to 0.04 F_1 points. On the left, below the species, is specified the best F_1 score and the number of optimal dates needed to reach it.

(2019) found that the optimal dates for mapping broadleaf and conifers species were two dates in spring (April 30, May 5) and three in autumn (October 14, October 17, November 8), but the differences in accuracy between their several combination of two, three or four dates do not seem significant.

Including our work, the three articles on Sentinel-2 for mapping tree species show significant disagreements over which dates to focus on. A reason why optimal dates are changing from one year to another can be the missing of a key event at a specific date which is not available from one year to another. It can also be relative to a local event, such as the dryness after a warm summer, the soil condition, or a substorey vegetation that begins before the leaf spreading (this could be the reason why February and March are selected as no leaf have flushed at this period).

At last, a bias in the Sequential Feature Selection to consider

is that only one date can be selected for each iteration, but there can be multiple possibility to reach the same accuracy. The best score ($F_1 = 0.87$) from one single date for the Red oak can be reached with three other dates: May 1st, August 24th and November 21st. For other species it is not as pronounced as for the Red oak, but it is common that a date as important as the chosen one can also be available several months apart.

5.2 Species accuracy

Looking at specific species, Grabska et al. (2019) also found difficulties to map Birch trees ($F_1 = 0.23$), but this can be explained by their very small sample size ($n = 25$). Our results contrast with (Bolyn et al., 2018) where accuracy is very high ($F_1 = 0.93$) as the number of samples ($n = 268$). However, as the authors used only 10% of the polygons for the validation and as it included mixed pixels, they think their "validation was probably over-estimated for the application of the classifier over the whole study site".

Like us, Immitzer et al. (2019) also encountered difficulties in mapping European ash ($F_1 = 0.77$). In our cas this species was the second most difficult to predict with the optimal dates ($F_1 = 0.82$).

As opposed to the best dates which differ from one paper to another, it turns out that the easiest or hardest species to predict are rather the same among recent papers.

5.3 Understanding the drop of F_1 score when using the the full SITS

To understand how to reach optimal dates performance with the full SITS, we need to understand errors and success. It is difficult to draw a reliable conclusion about which month or season to use for predicting a species as most important dates differ, sometimes heavily, from year to year.

Understanding these errors should improve the future works that use a full SITS. We chose to focus on the Silver birch trees as the average accuracy using the feature selection is highest in 2017 with only 5 dates ($F_1 = 0.81$) but very low with the full SITS ($F_1 = 0.18$). This species is found in three different forest stands, but two are within the same forest. It turned out that three dates particularly affected the quality of the model (Figure 5), and the reason for this failure is due to a cloud undetected by the MAJA algorithm over this forest. It was therefore not possible for the algorithm to retrieve the spectral signal of the Silver birch in this forest as the other plot was not impacted by a cloud or vice versa. This also confirms that SVM is very sensitive to noise in temporal data, as already suspected (Karasiak et al., 2019). The case Silver birch is not unique. When analysing the quality dropout by comparing the dates that heavily decrease the quality of species prediction, we systematically see underdetected clouds in the full SITS.

The solutions to avoid this quality dropout is either to smooth the time series, or to find an algorithm robust to temporal noise. Random Forest (RF) classifier can deal with noise in the data up to 25% of wrong labels (Pelletier et al., 2017), However, our noise is some a spectral noise, or temporal noise. The impact of this kind of noise has been studied with RF algorithm (Agjee et al., 2018), and authors noted that combining RF and penalized regression such as Ridge Regression significantly increases the noise robustness of the classifier.

5.4 Effect of noisy dates in the SITS

For year 2017, we notice very few optimal dates were in spring. To understand why, we look closely at spring images from 2017 as the season was selected in 2018 and 2019 by the SFS method. We found that for may 6 and 16 the cloud mask were not accurate. Indeed, clouds were underdetected on May 6 and overdetected on May 16 (Figure 6).

This kind of errors has already be stated (Karasiak et al., 2019). The Sequential Feature Selection is appropriate to deal with this error as these dates won't be selected to enhance the prediction. But in order to use the full SITS in case of an country scale product, it will be mandatory to smooth the temporal signature. An interesting way to remove noise would be to use the Whitaker smoother (Eilers, 2003), or Savitzky Golay . However, the first smoother did not show clear benefit for tree species mapping using a one-year SITS (Sheeren et al., 2016) but the authors did not take into account the spatial autocorrelation which can lead to a high overestimation of the predictive potential of the algorithm (Karasiak et al., 2019).

6. CONCLUSIONS

Thanks to the two Sentinel-2 satellites, it has been possible to acquire a significant number of images for year 2017, 2018 and 2019. Using optimal dates based on the Sequential Feature Selection, high performances were observed for Red oak (average $F_1 = 0.99$) and Willow (average $F_1 = 0.96$). However, some species are hard to identify, such as Black locust (average $F_1 = 0.71$).

There is no much differences between years ($\Delta F_1 = 0.05$) but the usage of the full SITS impacts heavily the quality (average $F_1 = 0.64$). The best performances were systematically higher when using the optimal dates found with the Sequential Feature Selection (average $F_1 = 0.87$). The poor performances of the full SITS using SVM algorithm can be explained by noise in the time series due to undetected clouds. However, the very high quality of the optimal dates for 2019 (average $F_1 = 0.86$) using a binding spatial cross-validation suggests there is a real potential of the time series to map forest species at country scale.

The selection of best dates not only provided better quality by reducing the number of images required, but also made possible to analyse the dates selected by the algorithm. The chosen dates were quite different from one year to another. The only species were selected dates were stable between year is Willow and Red oak, which are also the two most accurate species.

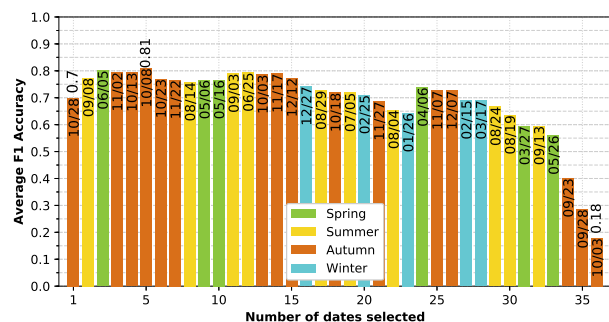


Figure 5. Optimal dates based on SFS to predict the Silver birch in 2017. The first, best and last F_1 score are shown above the bar, and inside is specified the selected date.

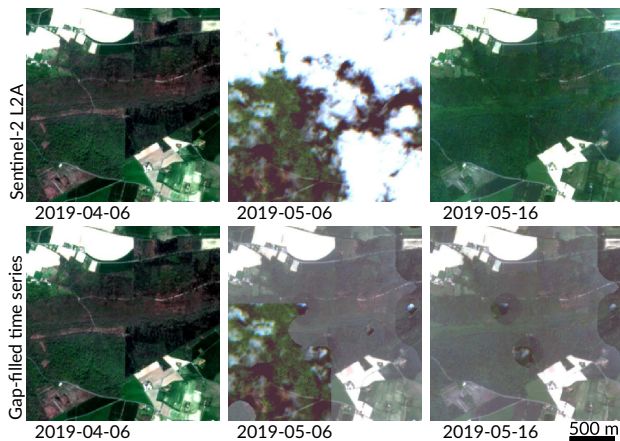


Figure 6. Misdetection of clouds generate temporal noise in the SITS. Here, the 2019-05-16 was uncloudy but as the MAJA algorithm finds a cloud above the forest, the gap-filling method has not been able to take advantage of this cloudless date.

Further research should focus on finding an algorithm robust to temporal noise and methods to smooth the time series. The identification of these methods, if they are successful, should remove the main problems of mapping tree species distribution at a country scale. It should result in a time series accuracy as close as the optimal dates quality. However, there is a need to have more tree species references to be more representative of the whole territory and less sensitive to local errors.

ACKNOWLEDGEMENTS

Special thanks to Pôle Theia and O. Hagolle from CESBIO Lab. for providing the pre-processed Sentinel-2 at level 2A with masks of clouds and cloud shadows.

References

Agjee, N. H., Mutanga, O., Peerbhay, K., Ismail, R., 2018. The Impact of Simulated Spectral Noise on Random Forest and Oblique Random Forest Classification Performance. *Journal of Spectroscopy*, 2018, 1–8.

Badeau, V., Ducatillon, C., Lebourgeois, F., 2017. *Les plantes au rythme des saisons: guide d'observation phénologique*. OCLC: 1027717693.

Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure. *Remote Sensing*, 11(4), 433.

Bolyn, C., Michez, A., Gaucher, P., Lejeune, P., Bonnet, S., 2018. Forest mapping and species composition using supervised per pixel classification of Sentinel-2 imagery. *Biotechnol. Agron. Soc. Environ.*

Brockerhoff, E. G., Jactel, H., Parrotta, J. A., Quine, C. P., Sayer, J., 2008. Plantation forests and biodiversity: oxymoron or opportunity? *Biodiversity and Conservation*, 17(5), 925–951.

Dale, M. R., Fortin, M.-J., 2014. *Spatial Analysis: A Guide For Ecologists*. 2 edn, Cambridge University Press, Cambridge.

Digital map of European ecological regions, 2000.

Eilers, P. H. C., 2003. A Perfect Smoother. *Analytical Chemistry*, 75(14), 3631–3636.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, 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.

Ferreira, M. P., Zortea, M., Zanotta, D. C., Shimabukuro, Y. E., de Souza Filho, C. R., 2016. Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. *Remote Sensing of Environment*, 179, 66–78.

Grabska, E., Hostert, P., Pflugmacher, D., Ostapowicz, K., 2019. Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sensing*, 11(10), 1197.

Green, A. A., Berman, M., Switzer, P., Craig, M. D., 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), 65–74.

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., Cresson, R., 2017. Orfeo ToolBox: open source processing of remote sensing images. *Open Geospatial Data, Software and Standards*, 2(1), 15.

Guo, Q., Fei, S., Potter, K. M., Liebhold, A. M., Wen, J., 2019. Tree diversity regulates forest pest invasion. *Proceedings of the National Academy of Sciences*, 116(15), 7382–7386.

Hagolle, O., Huc, M., Pascual, D., Dedieu, G., 2015. A multi-temporal and multi-spectral method to estimate aerosol optical thickness over land, for the atmospheric correction of FormoSat-2, LandSat, VENS and sentinel-2 images. *Remote Sensing*, 7(3), 2668–2691.

Hammond, T. O., Verbyla, D. L., 1996. Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing*, 17(6), 1261–1266.

Hartigan, J. A., Wong, M. A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, 28(1), 100.

Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1), 55–63.

Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F., Atzberger, C., 2019. Optimal Input Features for Tree Species Classification in Central Europe Based on Multi-Temporal Sentinel-2 Data. *Remote Sensing*, 11(22), 2599.

Inglada, J., 2016. OTB Gapfilling, a temporal gapfilling for image time series library.

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(1), 95.

Jain, A. K., Murty, M. N., Flynn, P. J., 1999. Data clustering: a review. *ACM Computing Surveys*, 31(3), 264–323.

Karasiak, N., 2020. Museo ToolBox: A Python library for remote sensing including a new way to handle rasters. *Journal of Open Source Software*, 5(48), 1978.

- Karasiak, N., Dejoux, J.-F., Fauvel, M., Willm, J., Monteil, C., Sheeren, D., 2019. Statistical Stability and Spatial Instability in Mapping Forest Tree Species by Comparing 9 Years of Satellite Image Time Series. *Remote Sensing*, 11(21), 2512.
- Karasiak, N., Sheeren, D., Fauvel, M., Willm, J., Dejoux, J.-F., Monteil, C., 2017. Mapping tree species of forests in southwest France using Sentinel-2 image time series. *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, IEEE, Brugge, Belgium, 1–4.
- Kavzoglu, T., Colkesen, I., 2009. A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, 11(5), 352–359.
- Larrieu, L., Gonin, P., 2008. L'indice de biodiversité potentielle (ibp) : une méthode simple et rapide pour évaluer la biodiversité potentielle des peuplements forestiers. *Revue Forestière Française*.
- Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., Bretagnolle, V., 2014. Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. *Global Ecology and Biogeography*, 23(7), 811–820.
- Lemaître, G., Nogueira, F., Aridas, C. K., 2017. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*, 18(17), 1–5.
- Lin, W.-C., Tsai, C.-F., Hu, Y.-H., Jhang, J.-S., 2017. Clustering-based undersampling in class-imbalanced data. *Information Sciences*, 409–410, 17–26.
- Mannel, S., Price, M., Hua, D., 2011. Impact of reference datasets and autocorrelation on classification accuracy. *International Journal of Remote Sensing*, 32(19), 5321–5330.
- Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8), 1778–1790.
- Miller, J., Wu, J., Boyer, M., Belanger, M., Hare, E., 1991. Seasonal patterns in leaf reflectance red-edge characteristics. *International Journal of Remote Sensing*, 12(7), 1509–1523.
- Moran, P. A. P., 1950. Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1), 17–23.
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259.
- Pasquarella, V. J., Holden, C. E., Woodcock, C. E., 2018. Improved mapping of forest type using spectral-temporal Landsat features. *Remote Sensing of Environment*, 210, 193–207.
- 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. *Journal of Machine Learning Research*, 12(Oct), 2825–2830.
- Pelletier, C., Valero, S., Inglada, J., Champion, N., Marais Sire, C., Dedieu, G., 2017. Effect of Training Class Label Noise on Classification Performances for Land Cover Mapping with Satellite Image Time Series. *Remote Sensing*, 9(2), 173.
- Pugh, S. A., Congalton, R. G., 2001. Applying Spatial Autocorrelation Analysis to Evaluate Error in New England Forest-Cover-Type Maps Derived from Landsat Thematic Mapper Data. *Photogrammetric Engineering & Remote Sensing*, 67(5), 613–620.
- Sheeren, D., Fauvel, M., Josipović, V., Lopes, M., Planque, C., Willm, J., Dejoux, J.-F., 2016. Tree Species Classification in Temperate Forests Using Formosat-2 Satellite Image Time Series. *Remote Sensing*, 8(9), 734.
- Smith, V. G., 1976. Canadian forest inventory methods. *The Forestry Chronicle*, 52(1), 9–14.
- Thompson, I. D., Okabe, K., Tylanakis, J. M., Kumar, P., Brockerhoff, E. G., Schellhorn, N. A., Parrotta, J. A., Nasi, R., 2011. Forest Biodiversity and the Delivery of Ecosystem Goods and Services: Translating Science into Policy. *BioScience*, 61(12), 972–981.
- van Lierop, P., Lindquist, E., Sathyapala, S., Franceschini, G., 2015. Global forest area disturbance from fire, insect pests, diseases and severe weather events. *Forest Ecology and Management*, 352, 78–88.
- Vapnik, V. N., 1998. Adaptive and Learning Systems for Signal Processing Communications, and control. *Statistical learning theory*.
- Whitney, A. W., 1971. A Direct Method of Nonparametric Measurement Selection. *IEEE Transactions on Computers*, 20(9), 1100–1103.
- Yan, E., Wang, G., Lin, H., Xia, C., Sun, H., 2015. Phenology-based classification of vegetation cover types in Northeast China using MODIS NDVI and EVI time series. *International Journal of Remote Sensing*, 36(2), 489–512.
- Yang, X., Tang, J., Mustard, J. F., Wu, J., Zhao, K., Serbin, S., Lee, J.-E., 2016. Seasonal variability of multiple leaf traits captured by leaf spectroscopy at two temperate deciduous forests. *Remote Sensing of Environment*, 179, 1–12.