Toward an operational monitoring of oak dieback with multispectral satellite time series: a case study in Centre-Val de Loire region of France

Florian Mouret, David Morin, Hilaire Martin, Milena Planells, Cécile Vincent-Barbaroux

▶ **To cite this version:**

Florian Mouret, David Morin, Hilaire Martin, Milena Planells, Cécile Vincent-Barbaroux. Toward an operational monitoring of oak dieback with multispectral satellite time series: a case study in Centre-Val de Loire region of France. 2023. hal-04320235v1

HAL Id: hal-04320235

https://hal.science/hal-04320235v1

Preprint submitted on 18 Apr 2023 (v1), last revised 4 Dec 2023 (v3)

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.

Distributed under a Creative Commons Attribution 4.0 International License

Toward an operational monitoring of oak dieback with multispectral satellite time series: a case study in Centre-Val de Loire region of France

Florian Mouret[1, 2], David Morin[2], Hilaire Martin[3], Milena Planells[2], and Cécile Vincent-Barbaroux[1]

[1]University of Orléans, LBLGC USC1328 INRAe (Ecodiv), BP 6759, rue de Chartres 45067 Cedex 2, Orléans, France
[2]CESBIO, University of Toulouse, CNES/UPS, 18 Av. Edouard Belin, bpi 2801, CEDEX 9, 31401 Toulouse, France
[3]INRAE, UR EFNO, Biodiversity Team, Domaine des Barres, 45290, Nogent-sur-Vernisson, France

Abstract

This paper studies the monitoring of oak dieback in forests of the Centre-Val de Loire region, France. Due to climate change, drought-induced forest dieback has become a major concern in temperate forests, including our study area, where oak is a key species. In order to better assess and adapt the actions needed to mitigate these impacts, access to accurate and regularly updated maps of forest health has become essential. In this context, the main objective of the study is to evaluate the interest of multispectral satellite time series for operational monitoring of forest dieback. Thanks to the in-situ data collected from 2017 to 2022 on about 2700 oak plots, a multi-year mapping of the analyzed region was performed using a supervised classification approach with the Random Forest algorithm. The results show that it is possible to detect oak dieback accurately (average overall accuracy = 80% and average balanced accuracy = 79%). More importantly, this study highlights the importance of measuring the temporal stability of the classification model in addition to standard cross-validation metrics. In this respect, the samples used for training are selected using data augmentation and balancing techniques to achieve better generalization over years. The learned model can also be used for predictive mapping of forest dieback in the coming years, for which the balanced accuracy is slightly reduced to about 70%. A feature analysis is also performed and shows that the shortwave infrared (SWIR) part of the spectrum is the most important for mapping forest dieback. In addition, using the red-edge portion of the spectrum can increase the stability of the model over time. We show that using only two vegetation indices based on continuum removal (CR) of the red edge (CRre) and shortwave infrared (CRswir) parts of the Sentinel-2 spectrum is sufficient to efficiently capture forest dieback. Overall, both in situ data and model predictions showed evidence of forest decline in many areas of the study region. Moreover, our results show that large areas of forest can decline over short periods of time, highlighting the interest of satellite data to provide timely and accurate information on forest status at large scales. This encourages the use and improvement of such approaches in the future.

Keywords
Remote sensing, Sentinel-2, Forest Monitoring, Random Forest, Machine Learning, Climate change, Dieback detection

1 Introduction

Climate change is expected to increase the severity and frequency of forest disturbances, both abiotic (e.g., fire, drought) and biotic (e.g., insects, pathogens). (Dale et al., 2001; Seidl et al., 2017; Turner, 2010). Assessment and monitoring of such disturbances is essential due to the central role of forests in the hydrological cycle (Zhang and Wei, 2021), carbon sequestration and

biodiversity conservation FAO (2020). Forests are also of great economic importance for example by providing wood and timber (Krieger, 2001). As a result, their disturbance can cause significant economic losses (Holmes et al., 2008). Finally, forests are often important in national low carbon strategies (i.e., French Ministry for the Ecological and Solidary Transition (2020)). Management measures (e.g., by reducing competition for water or promoting the establishment of species better adapted to future conditions) are needed to facilitate and mitigate the impacts of this transition on society (Millar and Stephenson, 2015).

In temperate forests, prolonged or exceptional droughts (also called "hotter droughts") can push forests beyond their sustainability threshold (Millar and Stephenson, 2015). The effects of hotter droughts, exacerbated by climate change, range from forest dieback to increased tree mortality to broad-scale forest die-off (Allen et al., 2015). Forest dieback is a complex and evolving phenomenon with multifactorial causes that results in a progressive weakening of trees and stands vigor (Manion, 1981). The factors involved are of several types (predisposing, triggering and aggravating) and partly interchangeable. The symptoms of dieback are essentially visible as a reduction in leaf area and crown (see examples in Figure 1). It has been identified as a factor reducing the resilience of ecosystems (Sangüesa-Barreda et al., 2015). France, like most of the European continent, has been affected by severe droughts in recent years (Moravec et al., 2021; Blauhut et al., 2022), resulting in a generalized weakening of forest health. In particular, the French Department of Forest Health has observed an accentuation of oak dieback on the national territory since 2018, which motivates a timely and accurate mapping of oak dieback to help forest managers. This article focuses on the monitoring of forest dieback using remote sensing satellite data, which is evaluated through a case study conducted in the Centre-Val de Loire region of France, an area known for its oak forests, which are a key component of the local economy and culture.

Figure 1: Illustration of different levels of forest dieback in the Orleans forest (oak trees, 2022). A dead tree is in the center of the image, with declining trees around it.

Remote sensing has been widely recognized as a valuable tool for monitoring forest status (Hansen et al., 2013; Huete, 2012; Torres et al., 2021; Stahl et al., 2023). The obvious advantage of remote sensing satellite data is its ability to cover large areas in a consistent manner. For forest health assessment, multispectral sensors have been most commonly used because they can efficiently capture different types of information about vegetation behavior, i.e., visible, near infrared (NIR), red-edge (RE) and shortwave infrared (SWIR) parts of the spectrum provide complementary information on vegetation status (Torres et al., 2021). In recent years, access to consistent and freely available multispectral satellite data has been facilitated by the opening of the Landsat archive in 2008 (Wulder et al., 2022) and the launch of the two Sentinel-2 (S2) satellites in 2015 and 2017 (Drusch et al., 2012). The arrival of S2 data has opened up unprecedented opportunities: S2 data have a higher revisit time ($\sim$ 5 days with S2-A and S2-B) and a finer spatial resolution (up to 10m) compared to Landsat data (revisit time $\sim$ 7-16 days, spatial resolution up to 30m).

The interest of these improvements has been identified in some cases, for example Abdullah et al. (2019) observed that S2 data were more effective than Landsat-8 data for early detection of bark beetle attacks. Similarly, Bárta et al. (2021) used S2 data for the same task in Norway. These considerations, as well as operational reasons, motivated the use of S2 data in this study.

In the remote sensing literature related to the monitoring of the forest health, many studies have focused on the detection of forest disturbances by looking at anomalies in the remote sensing signal (Barta et al., 2022; Cai et al., 2023; Housman et al., 2018; Mitchell et al., 2017; Mulverhill et al., 2023; Stahl et al., 2023; Yang et al., 2021). As pointed out in Barta et al. (2022), in that case disturbances are characterized by monitoring the magnitude and duration of the remote sensing signal changes. More formally, these approaches can be seen as *prediction-based* anomaly detection techniques (Aggarwal, 2017, Chapter 9.2): they aim to model the normal behavior of the signal based on historical data and define anomalies (or disturbances) as values that significantly deviate from this modeling. Among the benchmark methods based on this idea, one can mention the Break detection For Additive Seasonal Trends (BFAST) (Verbesselt et al., 2010) and the Landsat-based Detection of Trends in Disturbance and Recovery (LandTrendr) (Kennedy et al., 2010) approaches. More recently, the FORDEAD package (Dutrieux et al., 2021a) has been developed to detect bark beetle outbreaks in spruce trees. The main drawback of such approaches is that they require historical data to properly define what is normal behavior. In addition, Stahl et al. (2023) has pointed out that such an approach can struggle to detect "diffuse" disturbances (i.e., subtle changes in spectral reflectance), which is typically the case with drought-induced dieback of oak trees.

Supervised classification approaches can be preferred when the phenomenon under study is subtle and difficult to model (Torres et al., 2021). In recent years, methods based on machine learning (ML) have been increasingly used because they can model complex behaviors and can be easily deployed on a large scale. For example, in the systematic review made by Torres et al. (2021), most of the techniques used to monitor forest health are classification or regression techniques. Among them, the Random Forest (RF) algorithm (Breiman, 2001) was the most used, as it is generally more flexible and provides a relatively more transparent interpretation than most other ML algorithms. The same observation was made by Stahl et al. (2023) in their review related to the attribution (or classification) of forest disturbance types, where it was found that the RF algorithm was used in most cases. In addition, tree-based algorithm are known to be more easily interpretable, which can be useful to better understand the problems being modeled (Huang et al., 2022).

Based on this literature, we decided to tackle the problem of oak dieback detection (*Quercus robur* and *Quercus petraea*) in the Centre-Val de Loire region using supervised ML approaches with S2 data. Our main goals are 1) to separate healthy and declining areas as accurately as possible and 2) to produce maps on a large scale and for several years in an operational monitoring system.

The rest of this paper is organized as follows. Section 2 presents the study area as well as the data used for the analysis (i.e., the reference data and the remote sensing data used to produce the maps). In Section 3, the method used to map the forest dieback is provided, including details on how to handle reference data coming from different years. In Section 4, classification results are provided. In addition, we also study the temporal stability of the classification model and show that such an analysis is crucial in our case since we aim to map forest dieback over different years. An analysis of the features used (importance, temporal range, etc.) is also performed. Section 5 discusses these results and provides some additional insights related to the problem at hand. Finally, Section 6 summarizes and concludes this work.

2 Study Area and Data

2.1 Study Area

Our study area is the Centre-Val de Loire region of France and its surroundings. It was decided to analyze all forests included in the S2 tiles covering the region (the S2 tiles are the one provided by the Theia platform, (`https://theia.cnes.fr`, accessed on 1 March 2023). As shown in Figure 2, the study area is centered approximately at 47°7'N latitude and 1°8'E longitude (Northern France), and is relatively large (11 S2 tiles, 110000 km^2, with 23% of deciduous forest according to the OSO land cover map (Thierion et al., 2022)). The two major sets of soils of the region (base-rich or acidic) affects forest cover distribution (dry or waterlogged variants). Acidic and dry soils support oak forests (*Quercus petraea*, *Quercus robur*), accompanied by Hornbeam (*Carpinus betulus*), Birch (*Betula pendula*), Chestnut (*Castanea sativa*), and resinous (mainly planted) forests. Waterlogged

soils have forests dominated by aspen (*Populus tremula*), alder (*Alnus glutinosa*), and willows (*Salix sp.*) (Cordier et al., 2021). **This study focuses on oaks** (*Quercus robur* and *Quercus petraea*), which are a key species in the region. Centre-Val de Loire Region corresponds to a plateau surrounded by shallow valleys (max. altitude 500m above sea level (ASL), avg. 140m ASL). It is crossed by France's largest river, the Loire, and its main tributaries (Allier, Cher, Indre, Vienne). Moreover, the climate is temperate with an average annual temperatures of 11°C and less than 800 mm of precipitation per year. The few hilly areas (located in the northwest, east and south) have lower temperatures and higher levels of precipitation (Cordier et al., 2021).

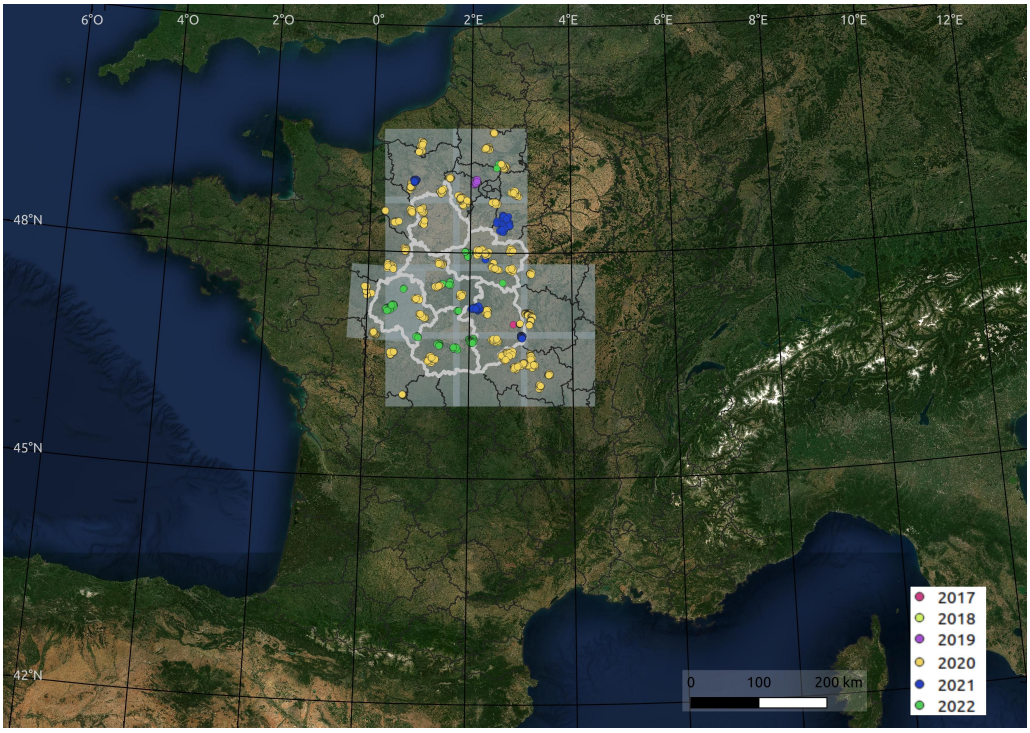


Figure 2: The extent of the studied area is delimited by the grey area (the boundaries between the 11 tiles is highlighted in lighter gray). The frontier of the region Centre-Val de Loire and its departments is in white. Finally, the colored dots locate the reference data, with each color representing a labeling year (for reference data labeled multiple times, the last labeling year is highlighted). The background uses cloudless S2 images.

2.2 Reference data

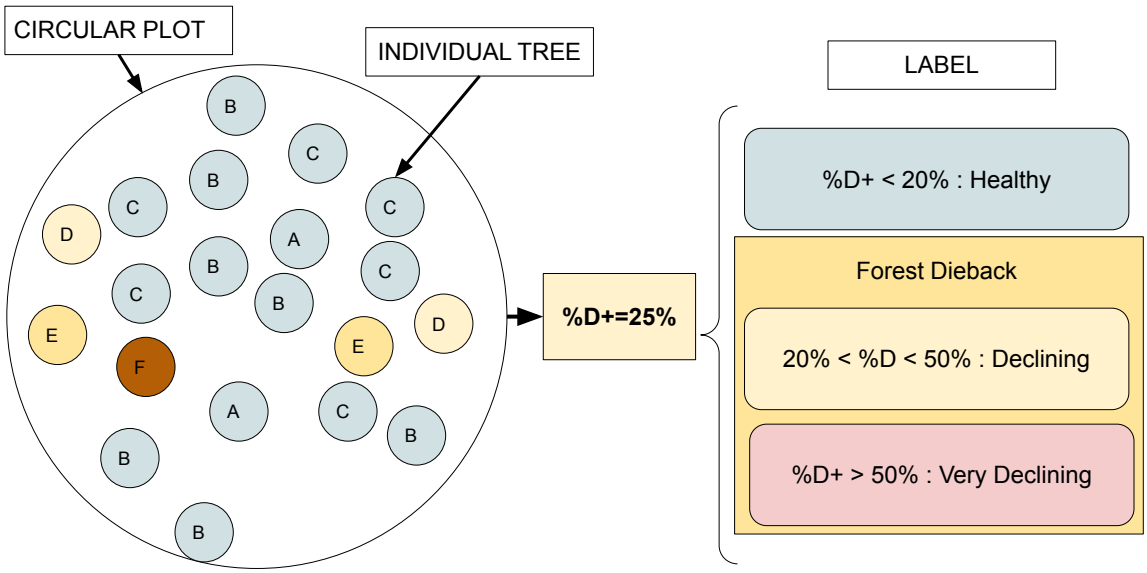
This subsection provides details on the collection and distribution of reference data.

2.2.1 Labeling protocol

The health status of the reference data is assessed using the DEPERIS protocol (Goudet et al., 2018). This protocol is used by the French Forest Health Service and is currently the official protocol for forest dieback monitoring in France (DGAL, 2018).

The DEPERIS protocol assesses the health status of individual trees by combining the percentages of dead branches and missing ramifications. Each tree is assigned a score from "A" (very healthy) to "F" (very declining or dead), with scores of "D" or higher corresponding to declining trees with more than 50% of canopy loss (see examples in Goudet et al. (2018)). Forest mortality is then characterized at the plot-level (a plot consists of 20 trees) by considering the percentage of declining trees. As defined by the French Forest Health Service, a plot is *declining* if more than 20% of its trees are declining, i.e., have a grade equal to or higher than "D". While the main objective of the analysis is to separate healthy from declining plots, for convenience we can add another (optional) category: a plot is *very declining* if more than 50% of its trees are declining. The labeling procedure is illustrated in Figure 3(a) with an example where 25% of the trees in the plot have scores equal to or higher than "D" (declining plot), while Figure 3(b) gives concrete

examples in the forest of Orléans, illustrating that plots that are very close to each other can have different labels.



(a)

(b)

Figure 3: (a) Labeling procedure used to asses plot-level health status. Each tree has a grade from A to F which is given by combining percentages of missing branches and ramifications. (b) Example of plots in Orléans forest labeled in 2022, where cyan / orange / red colors correspond to healthy, declining and very declining plots, respectively. A S2 image acquired in 2020 is used as background.

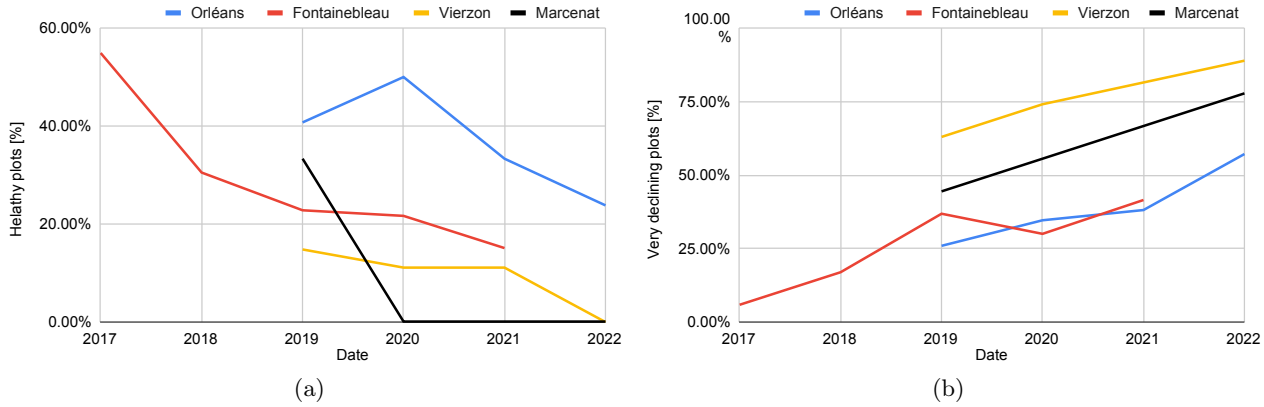
2.2.2 Distribution of the reference data

The distribution of the reference data with respect to the labeling year is given in Table 1. Overall, the proportion of healthy plots vs. plots with forest dieback (declining and very declining plots) is about the same but can vary depending on the labeling year. More than half of the plots were labeled in 2020: this year, a campaign was conducted in France after the successive droughts of 2018/2019/2020 to assess health status of oak forest by a random road sampling. The other plots come from different research campaigns carried out by private or public foresters (see Acknowledgements at the end of this document).

Table 1: 2738 plots were labeled between 2017 and 2022. For each year, the total number of reference data is provided, along with the breakdown into the 3 categories defined in Figure 3.

Year	# Plots	Healthy	Declining	Very declining
2022	300	111	97	92
2021	433	81	131	221
2020	1685	1202	338	145
2019	135	46	40	49
2018	120	50	52	18
2017	65	37	25	3
TOTAL #	2738	1527	683	528
TOTAL %	100.00%	55.77%	24.95%	19.28%

Among the reference data, 4 forest massifs are worth mentioning, since a labeling was carried out several years in a row (the number of labeled plots may vary slightly depending on the year and the forestry interventions). The Fontainebleau massif ($\sim$ 55 plots) was labeled between 2017 and 2021. Orléans ($\sim$ 25 plots), Vierzon ($\sim$ 27 plots) and Marcénat (9 plots) forests were labeled between 2019 and 2022. Figure 4 gives the percentage of healthy plots (Figure 4(a)) and very declining plots (Figure 4(b)) for each massif over time. Looking at this figure, there is evidence of a general increase in dieback. Moreover, the proportion of plots in severe decline is also increasing over time for all these massifs.



(a) (b)

Figure 4: (a) Percentage of healthy plots and (b) percentage of very declining plots with respect to the labeling year in the Orléans (blue), Fontainebleau (red), Vierzon (yellow) and Marcénat (black) massifs. Note that for Marcénat, not data was labeled in 2021.

2.3 Satellite data

This study used data from the Sentinel-2 satellites (S2-A and S2-B). The S2 satellites are operated by the European Space Agency (ESA) as part of the Copernicus mission, the European Union's Earth observation program. S2 satellites are multispectral imaging satellites with 13 spectral bands covering the visible, the near infra-red (NIR) and the shortwave-infrared (SWIR) spectral region and a spatial resolution of 10 to 20m (Drusch et al., 2012), as depicted in Table 2. The MAJA processing chain (Hagolle et al., 2015) was used to produce level 2A images, which are ortho-rectified products expressed in surface reflectance with cloud and shadow masks.

Table 2: Sentinel-2 multispectral bands used for the analysis.

Spectral bands	Central wavelength (nm)	Bandwith (nm)	Resolution (m)
B2: Blue	490	65	10
B3: Green	560	35	10
B4: Red	665	30	10
B5: Vegetation Red Edge (RE)	705	15	20
B6: Vegetation Red Edge(RE)	740	15	20
B7: Vegetation Red Edge(RE)	783	20	20
B8: Near Infrared (NIR)	842	115	10
B8A: Narrow Near Infrared (NIR)	865	20	20
B11: Shortwave Infrared (SWIR)	1610	90	20
B12: Shortwave Infrared (SWIR)	2190	180	20

3 Methods

The proposed approach for multi-year forest dieback mapping is summarized in Figure 5. The remainder of this section details each of the methodological steps in this diagram.

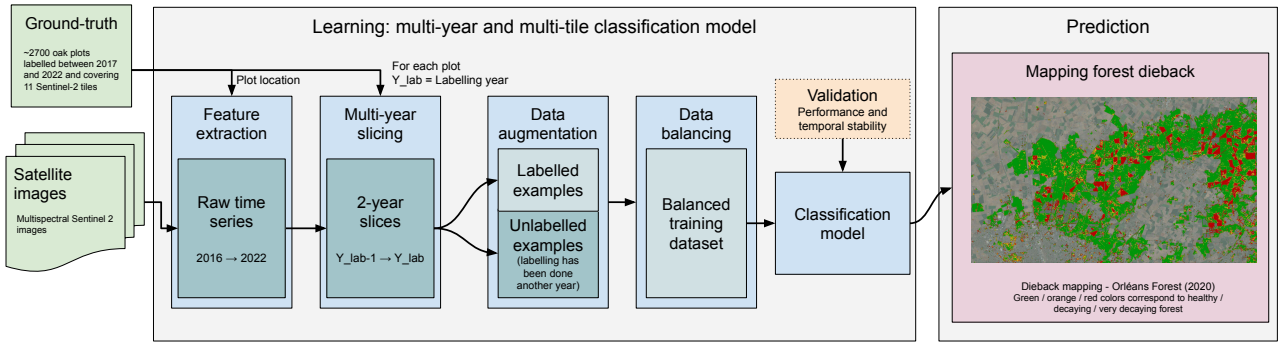


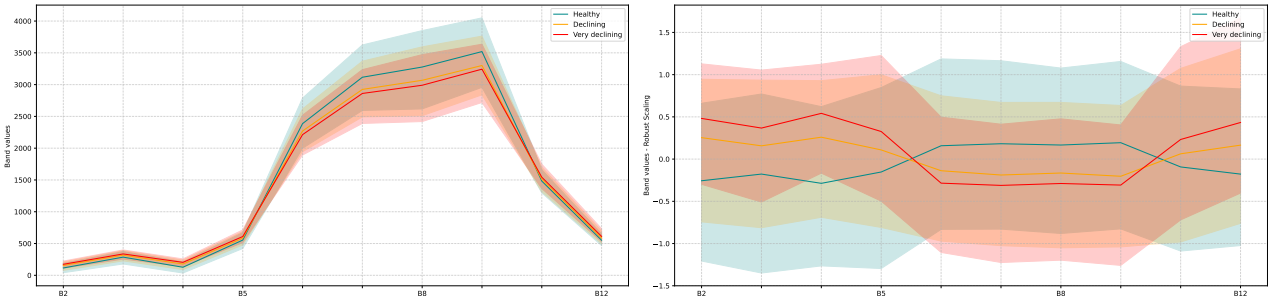
Figure 5: Diagram summarizing the proposed methodological steps for the classification of forest dieback.

3.1 Feature extraction

3.1.1 S2 bands

In the feature extraction step, time series of each S2 band coming from each pixel within the reference plots are extracted (see Figure 3(b) for examples of circular plots). This extraction is done using the *iota*^{2} software (Inglada et al., 2016b), which is a processing chain for the operational production of land cover maps from remote sensing image time series. A linear interpolation (i.e., gapfilling) is performed to reconstruct missing data caused by clouds. Such approach is commonly used in remote sensing and have been applied with success in many application (Inglada et al., 2015; Fauvel et al., 2020; Vuolo et al., 2017). The main advantage of gapfilling is that it can be applied on a large scale due to its low computational cost, which is well suited from an operational point of view. Further analysis on that point is left to future work, see Section 5. During the gapfilling step, the time series from different tiles are also interpolated on the same temporal grid with one acquisition every 10 days for a total of 35 acquisitions per year, illustrating the high temporal resolution of the S2 data. Finally, a road mask is applied to remove learning pixels that were too close to the roads.

As a preliminary analysis, Figure 6 provides the spectral response of the S2 data (the acquisition date is August 30 of the label year) with respect to the declining classes. It can be seen that forest dieback affects the entire S2 spectrum. Looking at the scaled version in Figure 6(b), we can see that pixels of declining plots have on average higher visible reflectances (B2, B3, B4) compared to the healthy pixel. A similar observation can be made for the SWIR bands (B11 and B12). Finally, for the red-edged part of the spectrum (B6, B7, B8a), declining pixels generally have a lower reflectance compared to healthy pixels.



(a) (b)

Figure 6: S2 band values grouped by dieback class (a) in natural scale and (b) after robust scaling (obtained by removing the median and scaling the data by the interquartile range). The acquisition date is August 30th of the labeling year. The colors cyan / orange / red correspond respectively to healthy, declining and very declining plots.

3.1.2 Additional features

In addition to the raw bands, it is classical to also compute more sophisticated features, i.e., vegetation indices (VIs), which can be used more efficiently by the classification algorithm since they contain richer information. A classical example in remote sensing for vegetation analysis is the Normalized Difference Vegetation Indice (NDVI), which is mainly related to the plant vigor (Bannari et al., 1995; Rouse et al., 1974) and is used to detect anomalies in the BFAST Verbesselt et al. (2010) or LandTrendr (Kennedy et al., 2010) algorithms mentioned in the introduction of this article.

In the course of our experiments, we have tested many VIs from the literature, among which two novel indices have been found particularly effective for the classification of forest dieback. These two indices are continuum removal (CR) of the S2 spectrum and are referred to as *CRswir* and *CRre* because they focus on the SWIR and RE portions of the spectrum, respectively. For the sake of brevity, the many tests conducted with other indices are not fully detailed in the manuscript, but a discussion is provided in Section 5. The CRswir and CRre formulas can be expressed as follows:

$$\mathrm{CR}_{\mathrm{swir}} = \frac{\mathrm{B11}}{\mathrm{B8a} + (\lambda_{\mathrm{B11}} - \lambda_{\mathrm{8a}}) \times (\frac{\mathrm{B12}-\mathrm{B8a}}{\lambda_{\mathrm{B12}}-\lambda_{\mathrm{8a}}})} \tag{1}$$

$$\mathrm{CR}_{\mathrm{re}} = \frac{\mathrm{B5}}{\mathrm{B4} + (\lambda_{\mathrm{B5}} - \lambda_{\mathrm{B4}}) \times (\frac{\mathrm{B6}-\mathrm{B4}}{\lambda_{\mathrm{B6}}-\lambda_{4}})} \tag{2}$$

where Bn and $\lambda_{\mathrm{B}n}$ are the reflectance and the wavelength in nanometers of the band n, respectively (see Table 2 for the corresponding values). The CRswir was successfully used for the mapping of bark beetle outbreaks in the FORDEAD package (Dutrieux et al., 2021a,b). Continuum removal aims at isolating individual absorption of interest (Clark and Roush, 1984), and has been mainly used with hyperspectral remote sensing image (Huang et al., 2004). Based on this idea, we propose a similar indicator (CRre) that focuses on the RE part of the S2 spectrum. Potential interests of such an indicator compared to normalized indices such as NDVI are 1) the fact that they combine more than 2 spectral bands and 2) they are not normalized and therefore not subject to saturation effects. They are also very easy and fast to compute when compared to more sophisticated indices. A graphical illustration of CRre and CRswir is given in Figure 7. It can be seen that both indices measure the absorbance of a specific band (B5 for CRre and B11 for CRswir) with respect to its local "convex hull" (B4 and B6 for CRre and B8A and B12 for CRswir). These two indices have the advantage of being complementary in the sense that they are sensitive to two important physical properties of the canopy. The SWIR part of the spectrum is known to be sensitive to leaf water content (Grabska et al., 2020; Olsen et al., 2013), while the red-edge part has been found to be sensitive to canopy chlorophyll content (Bramich et al., 2021; Delegido et al., 2011; Zarco-Tejada et al., 2018, 2003).

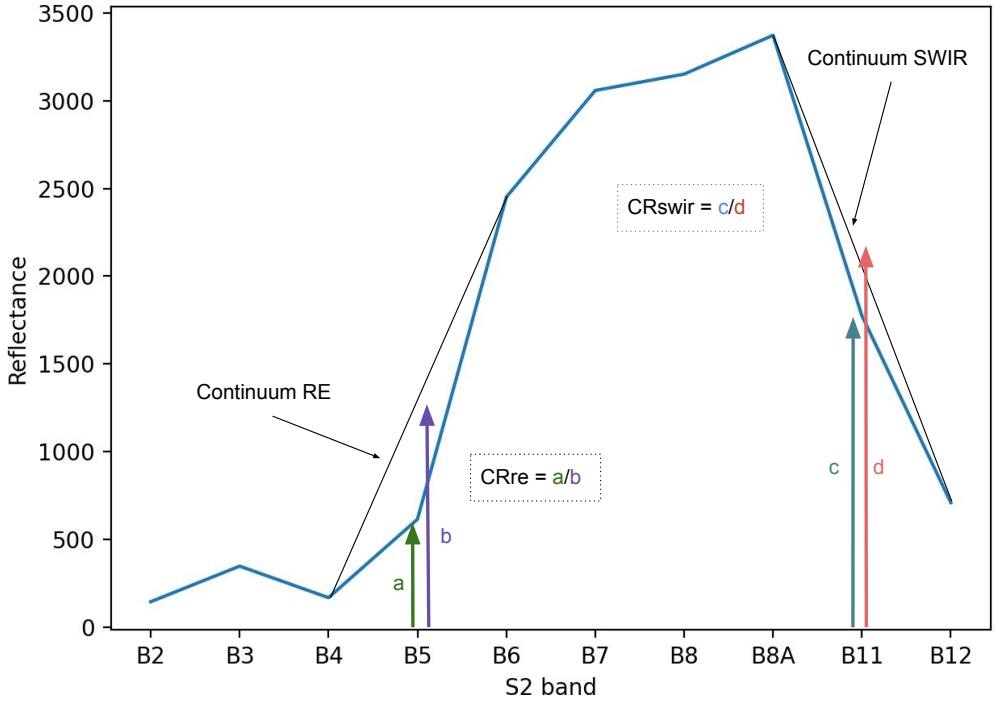


Figure 7: Reflectance of a healthy plot (Orléans forest) in summer (29/06/2020) and its continuum RE and SWIR (inspired by the figure in Dutrieux et al. (2021b))

3.2 Multi-year slicing and creation of the feature matrix

To be correctly used by the classification model, the time series of the plots must be ordered according to their different labeling years. The feature profiles from 2016 to 2022 are sliced according to the labeling date. The length of the temporal slice, denoted N_{years}, can cover for example two years before labeling, as illustrated in Figure 8. These slices, with the same dates (days/months) but different years, are superimposed to build a multi-year classification model. The multi-year mix makes the resulting classification model more robust to seasonal changes between years (e.g., phenological differences).

To illustrate these previous steps, Figure 9 provides time series of the median and interquartile range of the CRswir and CRre indices acquired over 2 years before labeling ($N_{\text{years}} = 2$) and grouped by declining classes. As shown in Figure 8, the different years are mixed according to the year of labeling Y_{lab} (here, Y2 corresponds to Y_{lab} and Y1 to $Y_{\text{lab}} - 1$), which correspond to the features given during the training of the RF algorithm. It can be seen that, despite having data from different years, clear trends are visible, with a gradation between healthy, declining and very declining plots. For both CRswir and CRre, summer (June / September) is the period with the most marked differences between classes. Interestingly, the budding period of the trees also seems to be informative (April/May), with declining tree generally being delayed when compared to the other tree. However, this last observation should be treated with caution due to the possible large phenological variability during this period (see the discussion on this point in Section 5, which provides a possible explanation for why data augmentation mitigates interseasonal variability).

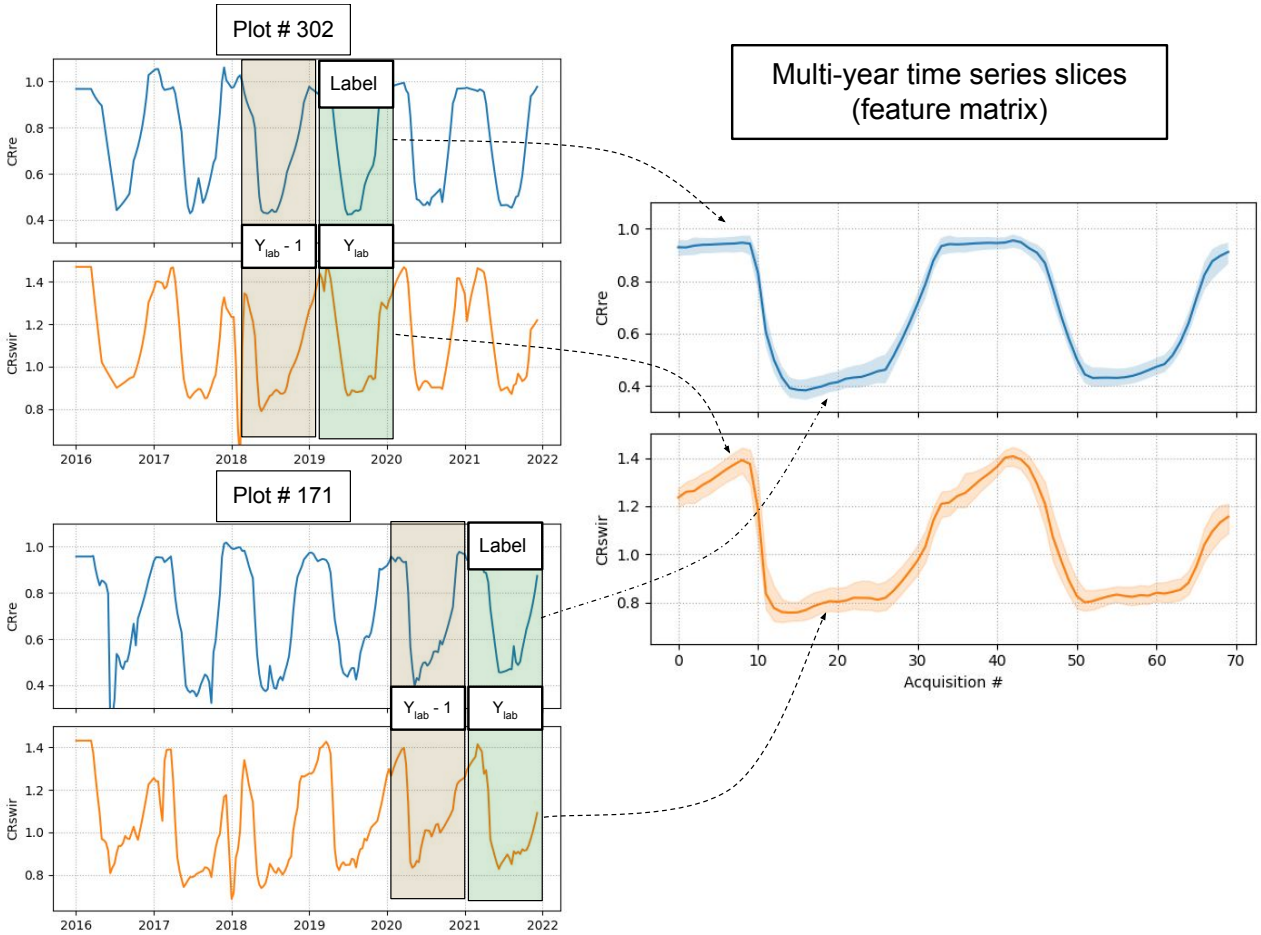
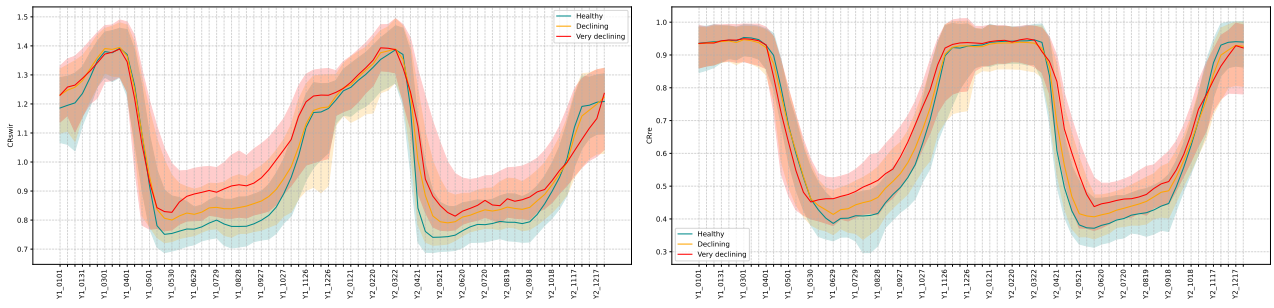


Figure 8: Illustration of the multi-year slicing used to create the feature matrix when using two time series (CRre and CRswir). The left part of the image shows the CRswir (top line) and CRre (bottom line) time series of two different plot over time. Depending on the labeling year of each plot (Y_{lab}), time series slices from different acquisition years (left part) are mixed together (right part) to create the feature matrix. In the right part, median and interquartile range (shaded area) of the whole dataset are displayed. For this example, the length of the slices is fixed to $N_{\text{years}} = 2$.



(a) CRswir (b) CRre

Figure 9: Time series of (a) CRswir and (b) CRre indices of the learning dataset acquired over 2 years prior to labeling (Y1 is the first year of acquisition, Y2 is the year of labeling). The colors cyan / orange / red correspond respectively to healthy, declining and very declining plots based on the percentage of trees with grades lower than D (see detailed examples in Figure 3). The solid line corresponds to the median value of the class and the shaded area to its interquartile range.

3.3 Data augmentation using unlabeled examples from previous and next years

As highlighted in Table 1, the reference data is not uniformly distributed over the different years, 2020 being over represented. We have found that using the raw training data obtained after Section 3.2 is sub-optimal since the classification model is likely to be optimized to classify correctly the samples labeled in 2020 (this can be seen as an imbalance problem (He and Garcia, 2009)). For better generalization, one can increase the number of available samples by using the remaining unlabeled time series (e.g., plot labeled in 2020 but without labels for the other years). A simple procedure with two conditions is applied to select new time series slices based on the available labeling at year Y_{lab}:

1. A very healthy plot (percentage of declining trees lower than 10%) labeled year Y_{lab} was likely to be also healthy the years $Y_{\text{lab}} - 1$ and $Y_{\text{lab}} - 2$ and is added to the training set with its corresponding time slice $(Y_{\text{lab}} - N_{\text{years}}, Y_{\text{lab}} - 1)$ and $(Y_{\text{lab}} - N_{\text{years}}, Y_{\text{lab}} - 1)$.
2. A declining plot (percentage of declining trees higher than 20%) labeled year Y_{lab} was likely to be declining the years $Y_{\text{lab}} + 1$ and $Y_{\text{lab}} + 2$ and is added to the dataset with its correspond time slice $(Y_{\text{lab}} + 1, Y_{\text{lab}} + N_{\text{years}})$ and $(Y_{\text{lab}} + 2, Y_{\text{lab}} + N_{\text{years}} + 1)$.

These two rules are based on the reasonable assumption that forest recovery from dieback is relatively slow (i.e., a declining plot cannot become healthy within 2 years). This also mean that we also assume that there was no silvicultural intervention during this period, or at least that it concerns a limited amount of samples. Obviously, to avoid data leakage the plots already labeled the years $Y_{\text{lab}} - 2$, $Y_{\text{lab}} - 1$, Y_{lab}, $Y_{\text{lab}} + 1$ and $Y_{\text{lab}} + 2$ are not duplicated in the dataset.

3.4 Data balancing

Another imbalance problem in our dataset comes from the label classes themselves. Depending on the year, the different classes to be classified are not evenly distributed (it is especially true after the data augmentation step since the labels of 2020 belong in majority to the healthy class). A naive way to solve this problem might be to balance the global dataset, i.e., using an over- or under-sampling strategy. Nevertheless, we have found that such an approach leads map production potentially oscillating between optimistic and pessimistic predictions over time (see discussion in Section 5). To that extent, we propose to balance the dataset ***each year independently*** using the Synthetic Minority Oversampling Technique (SMOTE), which has been used successfully in many applications (He and Garcia, 2009). In short, the SMOTE technique generates synthetic data based on the similarity in feature space between existing minority instances. The main advantage of over-sampling techniques over under-sampling techniques is that all available examples are kept in the training set. Note that we also tried other variants of the SMOTE algorithm with no improvement in our results (see Section 5). The SMOTE implementation used in this study is the one provided in the Python library imbalance-learn (version 0.10.1) (Lemaître et al., 2017). Finally, it should be mentioned that data augmentation via time lag (i.e., duplication of the time series with a small time lag) or using other sampling strategies (e.g., variants of SMOTE) were also tested without success.

3.5 Classification

For the operational classification step, we used the RF algorithm (Breiman, 2001), which, as presented in the introduction of this article, has shown very successful results when applied to remote sensing data and is therefore widely used in this community (Torres et al., 2021). Note that we also tested other state-of-the-art algorithms, such as XGBoost Chen and Guestrin (2016) or deep learning approaches adapted to time series (Wang et al., 2017). Overall, all of these algorithms provided comparable classification results, and further research on this part is left to future work. From an operational point of view, it was decided to focus on the RF algorithm since a fast C++ implementation is available in the iota2 processing chain based on the shark machine learning library (Igel et al., 2008). The RF algorithm has also the advantage of natively providing feature importance, which is computed as the (normalized) total reduction in Gini impurity introduced by a feature. Feature importance can be used to help us understand how our samples are classified. For our validation experiments, we have used the scikit-learn implementation of RF (version 1.2) since it provides feature importance (Pedregosa et al., 2011). Regarding the hyperparameters chosen for the RF algorithm, which have the advantage of being easy to tune, the number of trees

was set (by gridsearch) to $n_{\text{trees}} = 1000$ and the minimum number of samples in a node was set to nodesize = 30 samples (i.e., if the number of samples in a node is less than this parameter, then the node is not split). During our experiment, we found that the RF was robust in the choice of its parameters, which confirms its ability to perform well without intensive tuning.

3.6 Validation procedure

3.6.1 Scoring metrics

Classification results are validated by repeated stratified cross-validation (CV): 3-fold stratified CV are repeated 10 times, with stratification done to account for the proportion of each class. Obviously, the training and test sets are separated at the plot-level to avoid data leakage, i.e., pixels of the same plot in both train and test sets. Since the RF algorithm works at the pixel level, the labels of the plots are assigned by selecting the **majority class** among the pixels of the plots (using the average probability gives similar results), allowing some heterogeneity in the prediction of the pixels of a given plot.

Regarding the number of classes to be classified, our main goal is to classify as accurately as possible forest dieback, i.e., optimize the **binary classification** between healthy forest and forest dieback. However, from the users' point of view, the feedback we received was formal in highlighting the potential benefit of having 3 classes (as defined in Figure 3) for a more intuitive appropriation of the generated maps. In this sense, we have evaluated both configurations, with the 3-class problem being much more difficult, as highlighted in the next section. Note that in the 2-class problem, the model is first learned with 3 classes and the predictions of the declining classes are merged (using directly 2 classes lead to very close results).

Standard metrics were used, namely **overall accuracy (OA)** and **F1 scores**. OA provides the percentage of correctly classified items, while F1 score is the harmonic mean of precision (percentage of samples correctly labeled in class j) and recall (percentage of samples of class j that were correctly labeled). We also used **balanced accuracy (BA)**, which is defined as the average of the recalls obtained for each class (Pedregosa et al. (2011)). Like OA, BA has the advantage of providing a unique metric that is convenient for comparing different models or feature sets. However, unlike OA, BA is not affected by unbalanced data sets, since the recall is expressed as a percentage for each class. Finally, regarding the imbalance nature of our dataset, we have found that it was not adapted (and misleading) to use global metrics calculated over all test samples because such a metric gives more weight to the 2020 samples. Instead, we separated the samples coming from each labeling year (a global score can be obtained by averaging the scores from the different years).

3.6.2 Temporal stability

The standard CV results presented below can validate the accuracy of a given model to classify the examples of our reference data. However, this validation does not take into account the temporal stability of the model. In other words, a given prediction may be accurate for year Y, but it may fluctuate in years Y-1 and/or Y+1. As an example, we can take a pixel that is classified as "healthy" in 2020, but "declining" in 2019 (and 2021). Such a prediction is unlikely in practice, especially if it is generalized to the whole mapping (e.g., pessimistic prediction for the whole region in 2019 vs. optimistic prediction in 2020). Such behavior can be seen as a form of over-fitting (Cawley and Talbot, 2010), i.e., the model is optimized on the (limited) available data used for validation. This over-fitting is very difficult to mitigate, since it is not directly visible in the CV results. However it is noticeable when comparing the resulting prediction maps over consecutive years since in our case we know that the oak's decline is a gradual process with some temporal stability. Therefore, oscillations in health status are a sign of overfitting.

To measure the temporal stability of the different classification frameworks, we can analyze the prediction of given pixels over time. More precisely, using the same CV procedure as before, we classify the test pixels for the different years (2017 to 2022) and, even if we don't know the exact label of the pixels over time, we can measure for each year the percentage of declining pixels that are classified as healthy the next year, *i.e.*, **the number of declining pixels year $Y-1$ that become healthy year Y divided by the number of declining pixels year $Y-1$**. Obviously, to avoid data leakage, we make sure that all different years of a given test pixel are removed from the training set.

4 Results

This section presents experimental results that have been conducted to validate the proposed method. First, classification performances and temporal stability of the proposed framework are evaluated. Secondly, we focus on explaining these classification results and provide some insights into the multispectral response of forest dieback. For the sake of brevity, this section will focus on the results obtained using the RF algorithm with a selection of features (a discussion of complementary results is provided in Section 5).

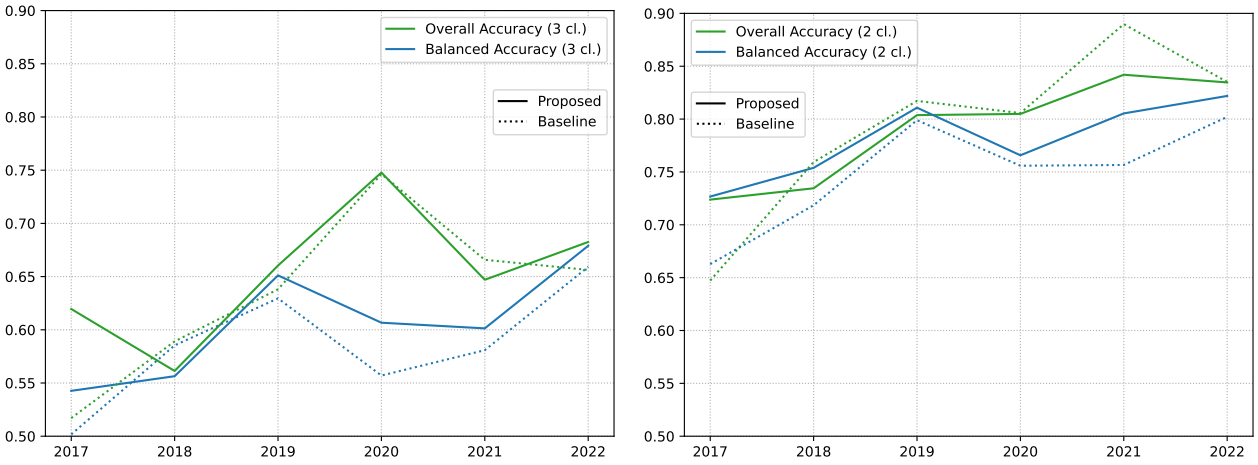
4.1 Classification results

For different feature sets and classification frameworks, Table 3 provides the OA and BA averaged over the labeling years with 95% confidence interval (CI). The so-called *baseline* framework consists of naively training a model without data augmentation and yearly balancing, while the *proposed* framework is the one provided in Figure 5. These results show that an accurate separation between healthy and declining classes can be obtained using S2 data and the RF algorithm. More precisely, it is possible to separate healthy forest and forest dieback (2 class scenario) with high accuracy (OA and BA $\sim$ 78% using the proposed framework). However, it is more difficult to separate declining and very declining plots, hence the drop in accuracy with 3 classes. Looking at the F1 values of each class (see Appendix A.1), we can see that the middle class (declining) is the most difficult to classify, while the very declining class is more accurately detected. We can see that using 2 years of data (instead of 1) improves the classification performance (similar results are obtained using 3 years, see Section 5). Finally, it can be observed that using only the CRswir index is sufficient to obtain accurate classification results, indicating that this index is able to efficiently summarize the S2 signal related to forest dieback (in Section 4.2, the potential interest of adding CRre to increase the temporal stability is highlighted).

Table 3: Overall Accuracy (OA) and Balanced Accuracy (BA) averaged over the different labeling years (confidence interval in parenthesis) for a classification with 3 classes (3 cl.) using the Random Forest algorithm. Results obtained by merging the declining classes (2 cl.) are also provided. The proposed and baseline frameworks are compared with different feature sets (best results in bold).

Framework	Features	N_{years}	OA - 3 cl.	BA - 3 cl.	OA - 2 cl.	BA - 2 cl.
Baseline	CRswir, CRre	2	0.635 (0.007)	0.586 (0.008)	**0.792** (0.007)	0.749 (0.007)
Baseline	S2 bands	2	0.617 (0.008)	0.582 (0.009)	0.772 (0.007)	0.723 (0.007)
Baseline	CRswir	2	0.624 (0.008)	0.572 (0.008)	0.780 (0.004)	0.738 (0.005)
Proposed	CRswir, CRre	2	**0.653** (0.008)	0.606 (0.008)	**0.791** (0.007)	**0.781** (0.007)
Proposed	CRswir, CRre	1	0.642 (0.009)	0.596 (0.008)	0.782 (0.007)	0.771 (0.007)
Proposed	S2 bands	2	0.650 (0.010)	0.608 (0.012)	0.778 (0.009)	0.773 (0.009)
Proposed	CRswir	2	**0.653** (0.010)	**0.610** (0.010)	**0.792** (0.007)	0.779 (0.007)

For a more in depth analysis, Figure 10 provides the OA and BA obtained with 3 (a) and 2 (b) classes of forest dieback when using the RF algorithm with CRswir and CRre features acquired over 2 years before labeling. This results help us to understand the differences observed in the averaged metrics presented in Table 3 and the interest of using BA over OA. The differences in OA and BA are the most important in 2020 and 2021, which correspond to very imbalanced year ($\sim$ 70% of healthy plots in 2020, while it is $\sim$ 18% in 2021). In both cases, the high OA can be misleading because it indicates a good classification of the majority class. When looking at the BA, the results can be less optimistic, but better take into account the classification of the minority samples. Therefore, the significantly better BA obtained with the proposed procedure over the baseline procedure (see Figure 10(a and b)) indicates that the data augmentation and balancing improve the classification model. These results have been confirmed visually during the map production, see the discussion with a visual example in Section 5. Finally, note that there is an exception in 2018 for the 3-class scenario: the better metrics obtained with 2 classes indicate confusion between declining classes, but not between healthy and declining classes, which are better separated by the proposed approach.



(a) 3 classes

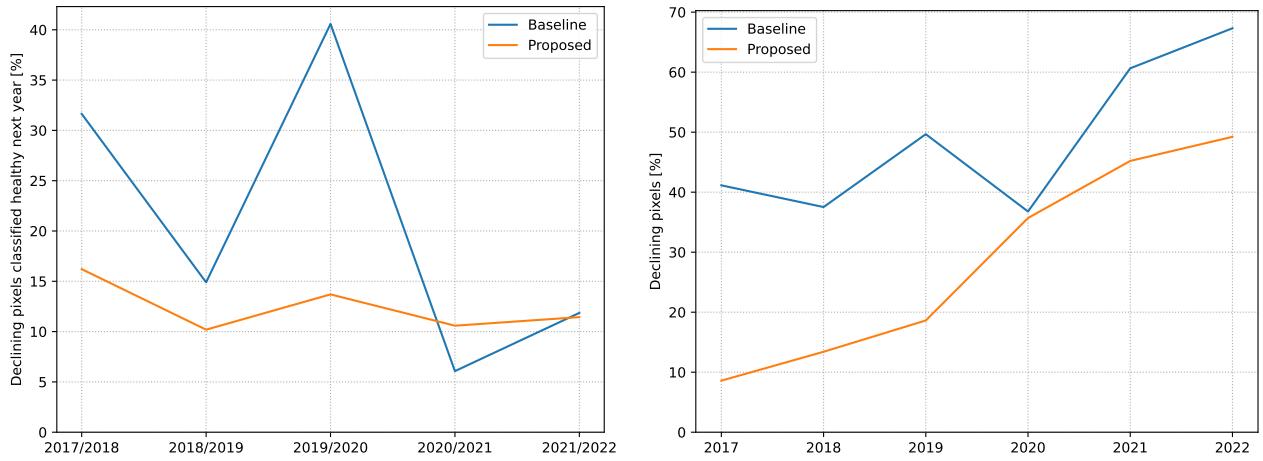
(b) 2 classes

Figure 10: Overall and Balanced accuracy obtained when using (a) 3 classes and (b) 2 classes of forest dieback. The features used as input are the CRswir and CRre acquired over two years ($N_{\text{years}} = 2$). The solid lines correspond to the results obtained with the proposed framework, the dashed line corresponds to the results obtained with the baseline framework. Results are averaged using 3-fold CV repeated 10 times.

4.2 Temporal stability

Previous results have confirmed the interest of S2 data for mapping forest dieback. However, as explained in Section 3.6.2, there is still a risk that the predictions of the model will fluctuate over years. To this end, Figure 11(a) measures the percentage of classified declining pixel that are classified healthy the next year (see definition in Section 3.6.2) and Figure 11(b) provides the percentage of classified declining pixels, using the baseline and proposed frameworks with the CRswir and CRre indices acquired over 2 years before labeling.

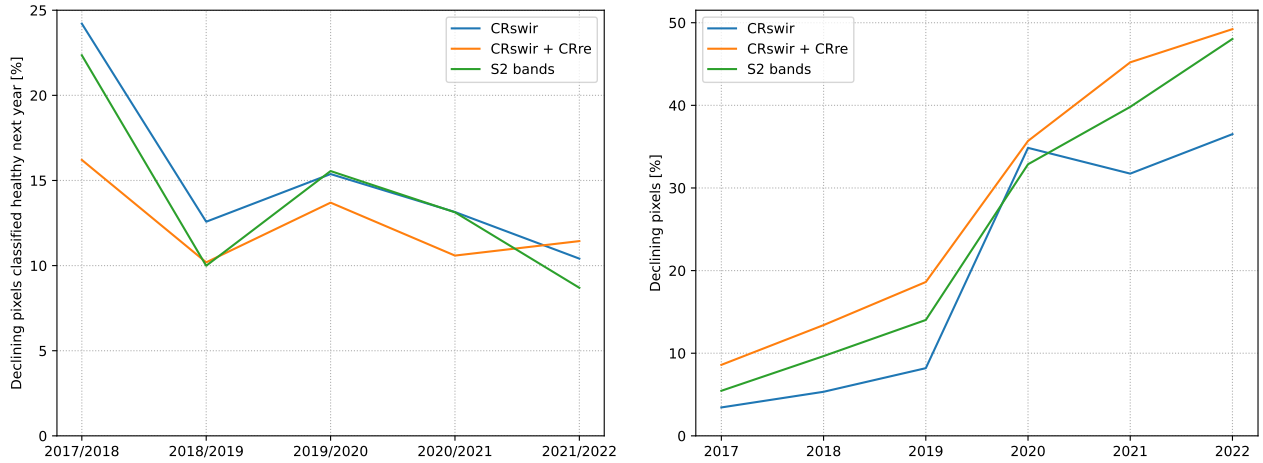
The results displayed in Figure 11(a) show that using the baseline framework leads to potentially very large oscillations, e.g., 40% of the reference pixels change from declining to healthy between 2019 and 2020, indicating that the map generated in 2019 was very pessimistic compared to the one generated in 2020. On the other hand, the use of the proposed framework leads to much more stable results, with a percentage of changes in the range 10-15%. Looking at Figure 11(b), we can see that the percentage of declining pixels oscillates when using the baseline classification. In addition, before 2020, these percentages are higher or close to the one obtained in 2020, which does not correspond to the observations and feedback we received regarding the evolution of the study area. Using the proposed framework, the percentage of declining pixels constantly increases over time, which is more consistent with the successive droughts that started in 2018. Finally, it is important to note that our dataset is not representative of the whole Centre-Val de Loire oak forest, as many plots of our reference data were selected because they were in decline. This explains why the number of pixels in decline is close to 50% in 2022. Looking at the pixels coming from the random road sampling done in 2020 (see Section 2.2.2), which is more representative of the region, the percentage of pixels classified as declining is close to 25%. These results (which can also be appreciated visually, see Section 5) show that in addition to a better theoretical classification on our reference data, using the proposed framework leads to a model with significantly better temporal stability. Moreover, they also show that the use of standard classification metrics obtained via CV may not be sufficient to have a complete overview of a model's behavior.



(a) (b)

Figure 11: (a) Percentage of declining pixels classified healthy next year and (b) percentage of pixels classified as declining. Results are averaged using 3-fold CV repeated 10 times. The features used are the CRswir and CRre acquired over 2 years prior to labeling.

Figure 12 is similar to Figure 11 but focuses only on the proposed framework using either CRswir, CRswir and CRre or all S2 bands as features. Figure 12(a) shows that using CRswir and CRre together leads to higher overall temporal stability. Moreover, Figure 12(b) also shows that using CRswir and CRre leads to the detection of a higher percentage of declining pixels, with a constant increase over time. On the other hand, when using only CRswir, the percentage of declining pixels oscillates slightly after 2020 without increasing, which is not consistent with our knowledge of the study area (especially since 2022 was a particularly severe drought).



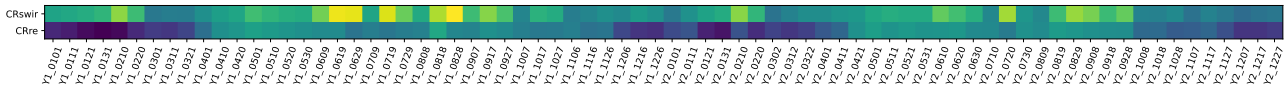
(a) (b)

Figure 12: (a) Percentage of declining pixels classified healthy next year and (b) percentage of pixels classified as declining pixels. Results are averaged using 3-fold CV repeated 10 times. The proposed framework is used with different set of features (CRswir, CRswir and CRre and all S2 bands) acquired over 2 years prior to labeling.

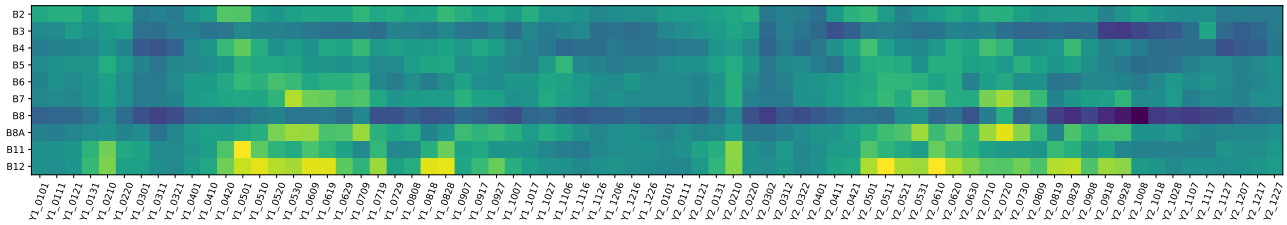
4.3 Feature analysis

This section focuses on analyzing how S2 bands and spectral indices are used to detect forest dieback. Note that as a complement, we also use the Python library SHAP (version 0.41.0) with the Tree Explainer method proposed by Lundberg et al. (2020). The use of SHAP leads to similar conclusions as the following results, therefore for the sake of simplicity we do not provide detailed results here, but it is interesting to point out that both analyses agree.

The feature importance obtained using the CR indices (CRswir + CRre) or S2 bands are shown in Figure 13(a) and Figure 13(b), respectively. The features are displayed in chronological order, where Y_1 is the first year of acquisition and Y_2 the second year of acquisition (i.e., the year of labeling). It can be seen that the SWIR information (B11 and B12 or CRswir) is largely used by the RF algorithm. It also appears that the most important dates are in the summer, between June and August (using the S2 bands, April and May also appear to be important). Finally, the two years of acquisition are used, with great importance given to features of year Y_1, indicating that the decline is visible long before labeling.



(a)



(b)

Figure 13: Log-feature importance obtained when training the model using a) CRswir and CRre indices and b) S2 bands, acquired over 2 years before labeling. The features are ordered in temporal order, Y1 being the first year of acquisition and Y2 the second year of acquisition, i.e., the labeling year.

4.4 Mapping of the study area

Figure 14 provides a map example for the year 2020. The OSO land cover map (Thierion et al., 2022)) is used to select deciduous trees (since no accurate oak mask was available, it was decided to used a deciduous tree to avoid masking too much area). It can bee seen that the southernmost forests are more affected by forest dieback. In the center of the map, the Sologne region is also largely affected (this region mostly consist in small forest patches combined with management objective based on hunting rather than production).

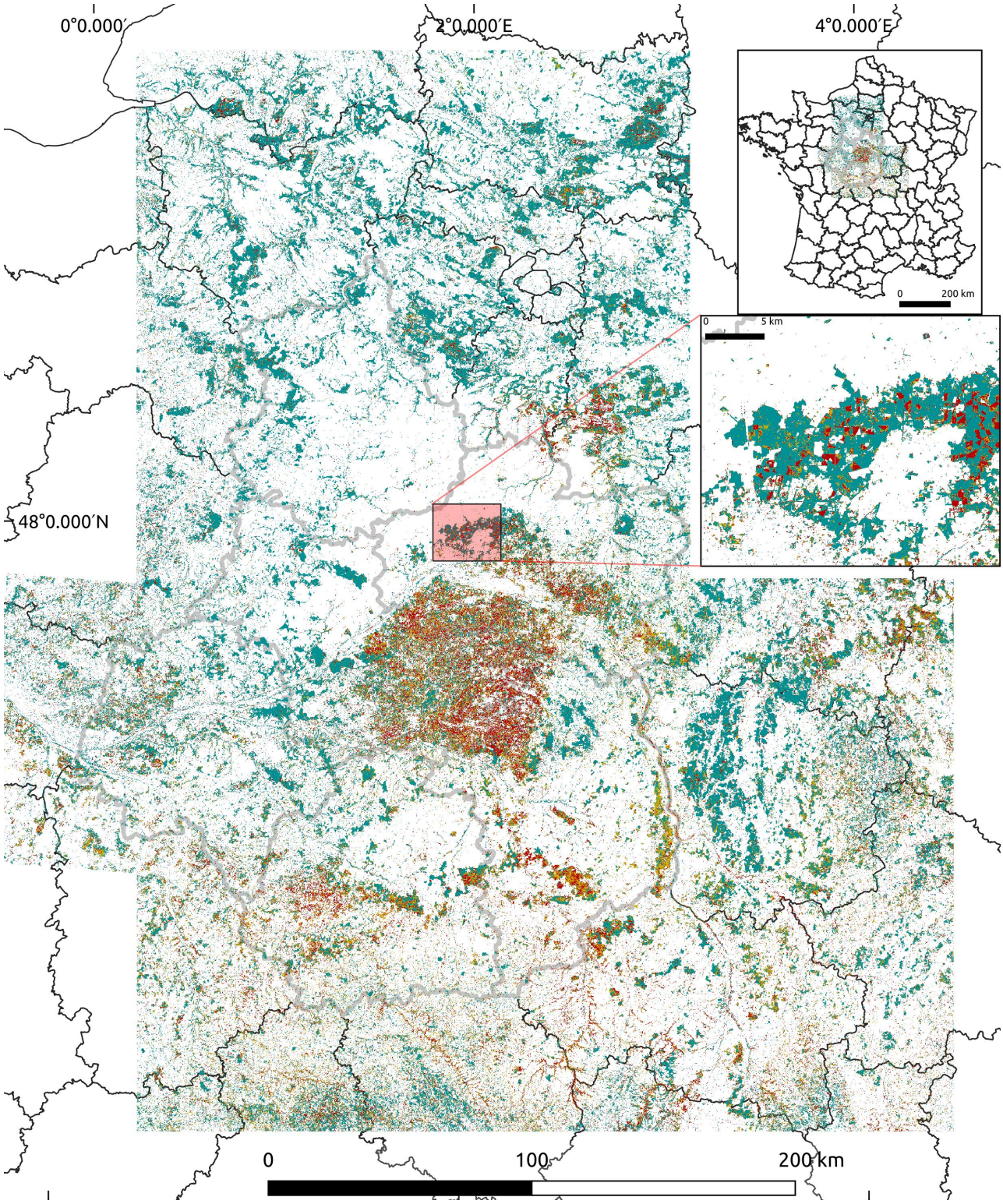
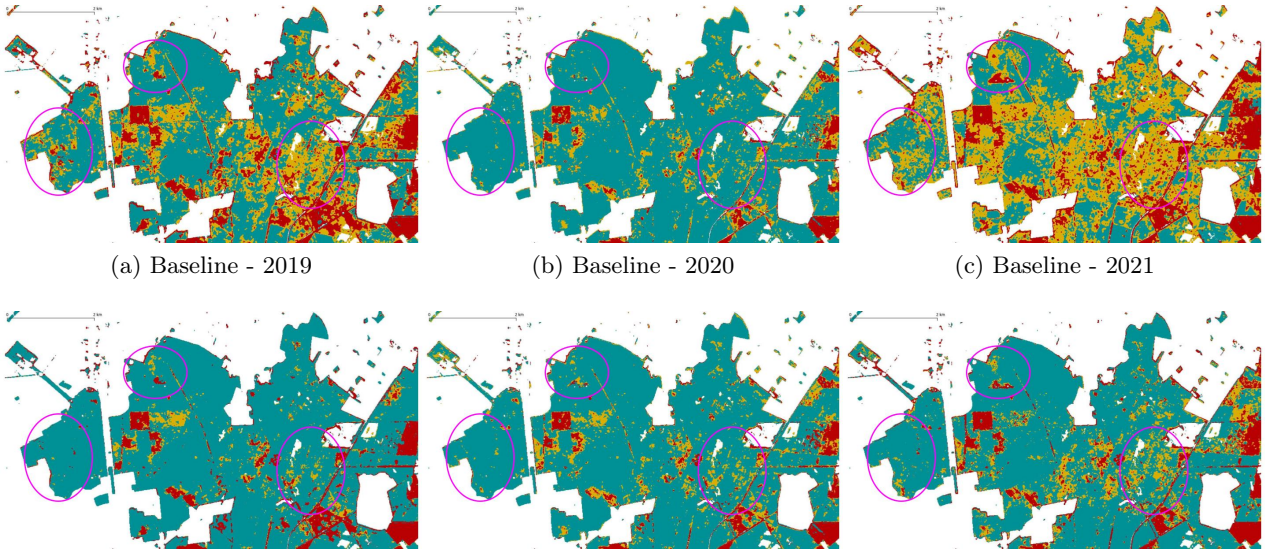


Figure 14: Final map production for the year 2020. The OSO. Healthy, declining and very declining pixels are in cyan, orange and red, respectively. The deciduous trees OSO land cover map is used. A de-zoomed version of the map is shown at the top right. Below is a zoomed version of the pink area, in Orléans forest (see Figure 15.

5 Discussion

5.1 Visual map comparison and interest of data augmentation

Figure 15 provides a visual comparison of mapping obtained with the baseline and the proposed classification approaches. Without data augmentation and balancing, we observed that some areas in the produced maps (especially those circled in pink) can vary significantly between years. On the other hand, proper balancing of the training data can mitigate these variations and lead to much more stable maps. We observed that with the baseline approach, the year 2019 and 2021 were pessimistic when compared to 2020, e.g., as visible in Figure 15(c). This is not in agreement with feedbacks from foresters and the maps from 2020, e.g., it is very unlikely that an area in decline in 2019 become healthy again in 2020 and so on. Finally, looking at these maps, some well delimited red patches are clearly visible. Theses patches correspond to areas where trees have been cut down due to forestry management (or in some cases roads).



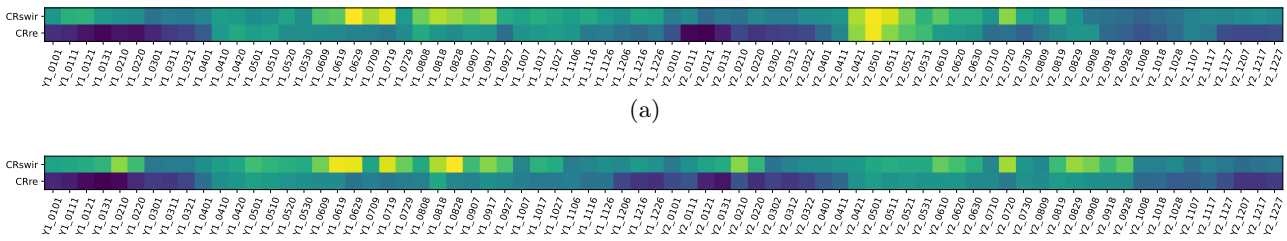
(a) Baseline - 2019 (b) Baseline - 2020 (c) Baseline - 2021

(d) Proposed - 2019 (e) Proposed - 2020 (f) Proposed - 2021

Figure 15: Maps produces in the Orléans forest without data augmentation and yearly balancing (a,b,c) and with data augmentation and yearly balancing (d, e, f) using the RF algorithm with CRswir and CRre indices (acquired over 2 years). Interesting areas are circled in pink. Healthy, declining and very declining pixels are in cyan, orange and red, respectively.

To the best of our knowledge, the most important explanation for the difference in variability with and without data augmentation and balancing is the influence of the budding period (April / May) on the model prediction. To highlight this point, Figure 16 shows the feature importance (in natural scale) obtained using the baseline framework (Figure 16(a)) and the proposed framework (Figure 16(b)). One can clearly see the importance of the April/May period before labeling (Y2) when using the baseline workflow (without data augmentation). On the other hand, the model trained with balanced augmented data relies less on this period and focuses more on dates between June and September.

To illustrate this point, Figure 17 provides the median CRswir and CRre time series of the **healthy pixels** labeled in the Fontainebleau forest grouped by labeling year (we chose this forest because we have the same area visited over time, allowing us to mitigate variations due to spatial location). We can clearly see that the inflexion point in April/May can vary over the years, e.g. the year 2020 was advanced while the year 2021 was delayed. Without a well-balanced dataset, such year-specific differences can be used by the model to "learn" that a given year is (on average) declining (e.g. 2021) or healthy (e.g. 2020). Note that the drop in CRswir in 2018/02/10 is due to undetected clouds in the analyzed area. Finally, we also found that adding dates other than summer (June/September) improve the classification results, which mean that there is interesting information outside the summer period that can be used to detect forest dieback.



(a)

(b)

Figure 16: Feature importance (in natural scale) obtained when training the model using CRswir and CRre indices with a) the baseline framework and b) data augmentation and balancing. The features are ordered in temporal order, Y1 being the first year of acquisition and Y2, corresponding to the labeling year, the second year of acquisition.

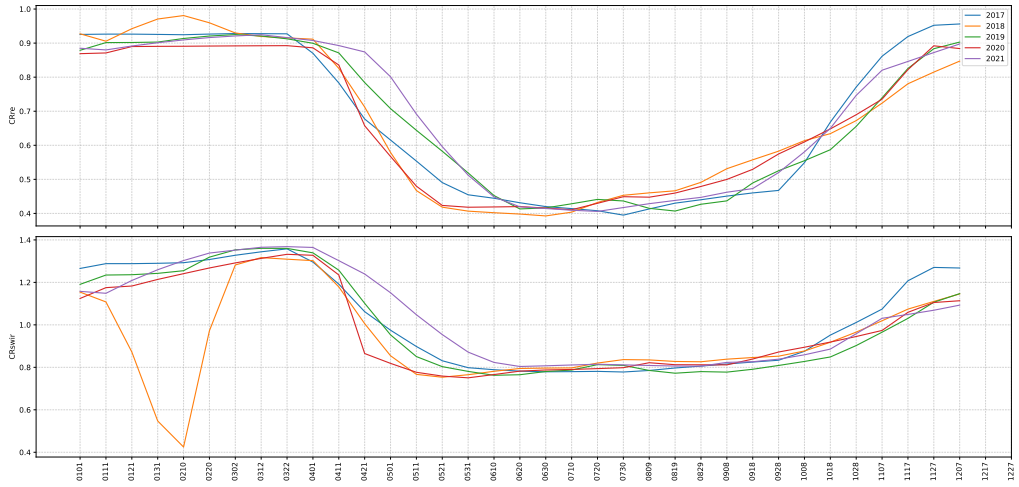


Figure 17: For the different labeling years, median of the CRre and CRswir for the healthy plots in the Fontainebleau forest.

5.2 Spectro-temporal response of S2 data to forest dieback

Our study highlights the crucial importance of the SWIR part of the spectrum to map forest dieback. This is in agreement with the literature, which had already identified this spectral zone. For example, the interest of SWIR for mapping bark beetle attacks was found in Dutrieux et al. (2021a), Bárta et al. (2021) and Huo et al. (2021), while Sapes et al. (2022) found that adding SWIR information increased the accuracy for oak wilt detection. Since drought-induced dieback is associated with increased plant water stress (Allen et al., 2015), the importance of SWIR was also expected in our case ying Huang et al. (2019).

In addition, our analysis shows that other parts of the spectrum (particularly RE) are also affected by forest dieback. This seems also coherent with the literature, since losses in chlorophyll content have been related to oak decline (e.g., Hornero et al. (2021) for Phytophthora-induced symptoms in oak decline). In the case of bark beetle attacks, Zabihi et al. (2021) highlight the potential interest of using RE instead of SWIR, but in our case this was not confirmed. The fact that bark beetle attacks are much aggressive than drought-induced dieback could be an explanation to the reduced importance of RE in our case. Even similar classification scores are obtained when using CRswir only and CRswir with CRre, better stability is obtained when using both indices. This is interesting and could indicate that the RE part of the spectrum can be used to "confirm" oak dieback in certain cases. Other benchmark indices such as the NDVI or the normalized difference water index (NDWI) (Gao, 1996), which combine bands 8A (red-edge) and 12 (SWIR) were tested without improving the detection results (e.g., the results obtained using NDWI instead of CRswir are very close but slightly worse).

Finally, using 2 years of data instead of only 1 year leads to better classification and stability and is consistent with the fact that oak dieback is influenced by previous consecutive years of drought (Rodríguez-Calcerrada et al., 2017). Nevertheless, the fact that acceptable results can be

obtained with only 1 year of data is interesting. Furthermore, no significant gain was observed when adding another year of data. One explanation could be that adding one year of data leads to having too many features (curse of dimensionality) or that this new information is too correlated with the other features.

5.3 Classification and pre-processing methods

Regarding the classification method to be used, we found that the RF algorithm is well suited and provides good results without requiring intensive parameter tuning. The fact that all models tested so far converged to similar results is encouraging (see Appendix A.2) and further research on that point will be an interesting perspective. The study of unsupervised methods such as FORDEAD Dutrieux et al. (2021b) is another perspective which is encouraged by the good results obtained using only the CRswir index. Such methods have the advantage of being usable without reference data. However, their implementation will be challenging given the problem at hand and the very subtle changes observed for declining pixels (for bark beetle attacks, the changes in CRswir are much faster and of higher magnitude).

Regarding the over-sampling technique, it was observed that using the standard SMOTE algorithm lead to the best results. Using variants (e.g., Borderline SMOTE Han et al. (2005) or ADASYN He et al. (2008)) lead to a deterioration of the results. One explanation for this deterioration could be related to potentially blurred class boundaries: while SMOTE over-samples the entire minority class, the variants such as ADASYN focus more on the boundaries. Further investigations on that point could be interesting, since our results highlight the importance of class balancing.

For the imputation of missing values, more sophisticated methods than gapfilling have been tested, e.g., Multiple Imputation by Chained Equations (MICE) (van Buuren and Groothuis-Oudshoorn, 2011) or k-nearest neighbors (KNN) (Troyanskaya et al., 2001), but they did not significantly improve the classification results and were much more time consuming. Further work on that part is also an interesting perspective, since its interest had been reported in various studies related to remote sensing Konrad Turlej et al. (2022); Mouret et al. (2022); Vuolo et al. (2017).

Finally, the greater confusion between declining and very declining plots may be explained by the fact that both classes share the same type of trees, i.e., trees with grades equal to or higher than "D" (see Section 2.2.1). The difference lies in the proportion of those trees in the plot. Nevertheless, the possible separation between declining and very declining plots (even if not perfect) is interesting from a user's point of view. Moreover, the good results obtained in the 2-class scenario (healthy forest vs. forest dieback) indicate that forest dieback generally affects the majority of the S2 pixels in the plot (conversely, the majority of pixels in healthy plots are healthy). Working at the plot level (e.g. using the plot mean or median) was tested and resulted in a degradation of the classification results, indicating that using all pixels of the plot leads to better generalization.

6 Conclusion

This paper presents a supervised classification framework for the operational monitoring of forest dieback using multispectral satellite imagery. Our analysis is conducted through a case study in the Centre-Val de Loire region (France), which focus on the study of oaks, a key essence of the region. Like other temperate forests, our study area has been affected by successive droughts in recent years, and regular mapping of the forest health situation will become increasingly important in the coming year to help the various stakeholders involved in forest management.

Like all supervised models, the proposed framework is based on training a classification model using reference data. The training is performed on time series slices acquired over two years prior to labeling. We propose to use time series from two vegetation indices based on continuum removal (CR) of the spectrum, namely CRre, which focuses on the red edge response of the spectrum, and CRswir, which focuses on the shortwave infrared part of the spectrum. Since our reference data are labeled in different years, the training features are also from different time periods. In order to have a better generalization over time, i.e. to mitigate the influence of phenological variability over years, we have found that data augmentation is a crucial step. In our case, we can take advantage of the fact that 1) healthy pixels were likely healthy in the years prior to labeling, and 2) declining pixels are likely to continue declining in the coming years, to increase the number of samples in our training set. Finally, we oversample the minority class for each labeling year using the SMOTE

algorithm. A random forest model is learned on this augmented dataset and can then be used to predict forest dieback.

Our analysis shows that the use of multispectral data (here Sentinel-2 images) can be used to accurately detect forest dieback (overall and balanced accuracy close to 79%). In the 3-class scenario (healthy, declining, very declining), the separation of declining and very declining classes remains challenging. Moreover, our study highlights the fact that standard cross-validation results may not be sufficient to measure the stability of a classification framework over time. More specifically, we have shown that without data augmentation, and despite relatively good cross-validation results, map production may lack temporal stability across the different years analyzed. A key factor that leads to oscillations between pessimistic and optimistic mapping of the study area is the phenological variation over time, which can be interpreted by a classification model as a sign of forest dieback (or health) without proper balancing of the data. Finally, a feature analysis was carried out, highlighting the importance of the SWIR information for mapping forest dieback, even if the whole S2 spectrum is affected. Other parts of the spectrum, especially the red edge, can be used to increase the temporal stability of the model, but without significantly improving the cross-validation classification results.

These results open up interesting perspectives. First of all, specific work on the classification model could improve dieback detection, especially since interesting results have been obtained using standard deep learning approaches. Other state-of-the-art models could be tested, such as Sparse Gaussian Processes (Bellet et al., 2023) or deep learning models based on transformers such as Lightweight Temporal Self-attention (Garnot and Landrieu, 2020). Since the data augmentation and balancing steps were found to be important for better generalization, further work on this part could increase model stability and performance, for example by using semi-supervised approaches (van Engelen and Hoos, 2019). An advantage of using deep learning approach could also be the potential application to other types of trees via transfer learning (Weiss et al., 2016), e.g. *Pinus sylvestris* is another key essence of the studied area. Unsupervised detection is another interesting perspective, especially since the detection does not rely on labeled data. Given the relatively subtle response of oak trees to drought-induced mortality, a combination of both approaches may be preferable. Finally, additional data coming from synthetic aperture radar using Sentinel-1 satellites could increase the classification performances and is encouraged by the interest found in combining both type of data in many remote sensing applications (Inglada et al., 2016a; Morin et al., 2019).

Acknowledgments: This work was supported by the SYCOMORE program, with the financial support of the Région Centre-Val de Loire (France).

The authors would like to thank B. Tardy, M. Fauvel and J. Inglada for their support and help with the *iota*^{2} software. We thank the CNES for providing us access to its high performance computing (HPC) infrastructure to run the map generation presented in this paper. Finally, we would like to thank all the participants of the RECONFORT project, who provided reference data and feedbacks. More precisely: Office National des Forêts (ONF), Centre National de la Propriété Forestière (CNPF), Unisylva and INRAE EFNO. Finally, we thank J.-B. Féret for its help regarding the CRswir indice, T. Belouard, A. Jolly and M. Chartier for their scientific advices.

Appendix A Additional results

A.1 F1 scores

To complement the results provided in the main document, Table 4 provides the F1 scores for each class. We can see that the proposed framework is better overall, with a significant improvement for the healthy class. Furthermore, the use of all S2 bands leads to a decrease in the F1 score of the middle class. Finally, when compared to the results presented in the main document (especially BA and OA with 2 classes), the lower F1 score for the declining classes indicates a confusion between these two classes and not between healthy and declining classes.

Table 4: F1 scores averaged over the different labeling years (confidence interval in parenthesis) using the Random Forest algorithm for the 3 classes of forest dieback, namely healthy (health.), declining (decl.) and very declining (v. decl.). The proposed and baseline frameworks are compared with different feature sets.

Framework	Features	N year	F1 - health.	F1 - decl.	F1 - v. decl.
Baseline	CRswir, CRre	2	0.699 (0.010)	**0.500** (0.011)	0.557 (0.013)
Baseline	S2 bands	2	0.663 (0.012)	0.489 (0.011)	**0.581** (0.013)
Baseline	Crswir	2	0.684 (0.009)	0.489 (0.014)	0.55 (0.014)
Proposed	CRswir, CRre	2	**0.736** (0.010)	**0.496** (0.013)	**0.573** (0.014)
Proposed	CRswir, CRre	1	**0.732** (0.009)	0.467 (0.013)	**0.575** (0.016)
Proposed	S2 bands	2	**0.741** (0.009)	0.457 (0.015)	**0.581** (0.016)
Proposed	CRswir	2	**0.737** (0.010)	**0.506** (0.011)	**0.579** (0.016)

A.2 Results obtained with other classifiers

Table 5 provides the average OA and BA obtained with other classifiers than the RF algorithm, when using the CRswir acquired over 2 years before labeling. The RF algorithm is compared to the XGBoost algorithm (Chen and Guestrin, 2016) and a Fully Convolutional Network (FCN) (Wang et al., 2017), which are adapted to time series. A minimal tuning was conducted by grid-search, i.e., for XGBoost we set the learning rate, the tree depth and number of estimator to 0.1, 7 and 100, respectively. For the FCN, we used the baseline parameters provided in Wang et al. (2017), i.e., a multi-channel convolutional network of size (128, 256, 128) with learning rate 0.001 and 30 epochs (Wang et al., 2017, Figure 1). Overall, RF algorithm provides the best results without a need for intensive tuning. The results obtained with standard value for the FCN are encouraging and could motivate future work on that point.

Table 5: Overall Accuracy (OA) and Balanced Accuracy (BA) averaged over the different labeling years (confidence interval in parenthesis) for a classification with 3 classes (3 cl.). Results obtained by merging the declining classes (2 cl.) are also provided. Various classifiers are compared using CRswir index acquired over 2 years before labeling.

Classifier	OA - 3 cl.	BA - 3 cl.	OA - 2 cl.	BA - 2 cl.
RF	0.653 (0.010)	**0.610** (0.010)	**0.792** (0.007)	**0.779** (0.007)
XGBoost	0.623 (0.009)	0.597 (0.011)	0.777 (0.009)	0.760 (0.009)
FCN	**0.660** (0.012)	**0.610** (0.012)	0.783 (0.009)	0.757 (0.008)

A.3 Forecasting study

To analyze the potential interest of the proposed model to classify future years, a forecasting study was conducted by testing the model on a given year after training on the other years. Subset of the analyzed were selected by repeating 10 times a stratified 3-fold CV. Average OA and BA (2 classes) are reported in Figure 18. One can see a drop in the accuracy, i.e., average BA over yers is equal to 0.721 instead of 0.792 (see Figure 10(b)). This drop was expected, since the model is

not calibrated on the year to be predicted and because less samples are available (particularly for the years 2020, 2021 and 2022).

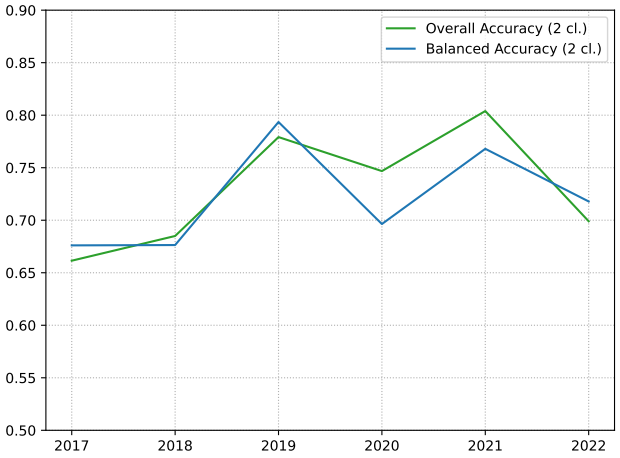


Figure 18: Overall and Balanced accuracy obtained when using (a) 3 classes and (b) 2 classes of forest dieback. The features used as input are the CRswir and CRre acquired over two years ($N_{\text{years}} = 2$).

References

Abdullah, H., Skidmore, A.K., Darvishzadeh, R., Heurich, M., 2019. Sentinel-2 accurately maps green-attack stage of European spruce bark beetle (Ips typographus, L.) compared with Landsat-8. Remote Sens. Ecol. Conserv. 5, 87–106. URL: https://zslpublications.onlinelibrary.wiley.com/doi/abs/10.1002/rse2.93, doi:https://doi.org/10.1002/rse2.93.

Aggarwal, C.C., 2017. Outlier Analysis. 2nd ed., Springer International Publishing, Cham. doi:10.1007/978-3-319-47578-3_3.

Allen, C.D., Breshears, D.D., McDowell, N.G., 2015. On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the anthropocene. Ecosphere 6, 129. URL: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/ES15-00203.1, doi:https://doi.org/10.1890/ES15-00203.1.

Bannari, A., Morin, D., Bonn, F., Huete, A.R., 1995. A review of vegetation indices. Remote Sens. Rev. 13, 95–120. doi:10.1080/02757259509532298.

Barta, K.A., Hais, M., Heurich, M., 2022. Characterizing forest disturbance and recovery with thermal trajectories derived from Landsat time series data. Remote Sens. Environ. 282, 113274. URL: https://www.sciencedirect.com/science/article/pii/S0034425722003807, doi:https://doi.org/10.1016/j.rse.2022.113274.

Bellet, V., Fauvel, M., Inglada, J., 2023. Land cover classification with gaussian processes using spatio-spectro-temporal features. IEEE Trans. Geosci. Remote Sens. 61, 1–21. doi:10.1109/TGRS.2023.3234527.

Blauhut, V., Stoelzle, M., Ahopelto, L., Brunner, M.I., Teutschbein, C., Wendt, D.E., Akstinas, V., Bakke, S.J., Barker, L.J., Bartošová, L., Briede, A., Cammalleri, C., Kalin, K.C., De Stefano, L., Fendeková, M., Finger, D.C., Huysmans, M., Ivanov, M., Jaagus, J., Jakubínský, J., Krakovska, S., Laaha, G., Lakatos, M., Manevski, K., Neumann Andersen, M., Nikolova, N., Osuch, M., van Oel, P., Radeva, K., Romanowicz, R.J., Toth, E., Trnka, M., Urošev, M., Urquijo Reguera, J., Sauquet, E., Stevkov, A., Tallaksen, L.M., Trofimova, I., Van Loon, A.F., van Vliet, M.T.H., Vidal, J.P., Wanders, N., Werner, M., Willems, P., Živković, N., 2022. Lessons from the 2018–2019 European droughts: a collective need for unifying drought risk management. Nat. Hazards Earth Syst. Sci. 22, 2201–2217. URL: https://nhess.copernicus.org/articles/22/2201/2022/, doi:10.5194/nhess-22-2201-2022.

Bramich, J., Bolch, C.J., Fischer, A., 2021. Improved red-edge chlorophyll-a detection for sentinel 2. Ecol. Indic. 120, 106876. URL: https://www.sciencedirect.com/science/article/pii/S1470160X20308141, doi:https://doi.org/10.1016/j.ecolind.2020.106876.

Breiman, L., 2001. Random forests. Mach. Lear. 45, 5–32. doi:https://doi.org/10.1023/A:1010933404324.

van Buuren, S., Groothuis-Oudshoorn, K., 2011. mice: Multivariate imputation by chained equations in R. J. Stat. Softw. 45, 1–67. URL: https://www.jstatsoft.org/v045/i03, doi:10.18637/jss.v045.i03.

Bárta, V., Lukeš, P., Homolová, L., 2021. Early detection of bark beetle infestation in Norway spruce forests of Central Europe using Sentinel-2. Int. J. Appl. Earth Obs. Geoinf. 100, 102335. URL: https://www.sciencedirect.com/science/article/pii/S0303243421000428, doi:https://doi.org/10.1016/j.jag.2021.102335.

Cai, Y., Shi, Q., Xu, X., Liu, X., 2023. A novel approach towards continuous monitoring of forest change dynamics in fragmented landscapes using time series landsat imagery. Int. J. Appl. Earth Obs. Geoinf. 118, 103226. URL: https://www.sciencedirect.com/science/article/pii/S1569843223000481, doi:https://doi.org/10.1016/j.jag.2023.103226.

Cawley, G.C., Talbot, N.L.C., 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. J. Mach. Learn. Res. 11, 2079–2107. URL: http://jmlr.org/papers/v11/cawley10a.html.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in: Proc. SIGKDD, ACM, New York, NY, USA. pp. 785–794. URL: http://doi.acm.org/10.1145/2939672.2939785, doi:10.1145/2939672.2939785.

Clark, R.N., Roush, T.L., 1984. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. J. Geophys. Res. Solid Earth 89, 6329–6340. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB089iB07p06329, doi:https://doi.org/10.1029/JB089iB07p06329.

Cordier, J., Dupré, R., Bellenfant, S., Gautier, S., 2021. Atlas de la flore du Centre-Val de Loire. Inventaires & biodiversité ; 19, Muséum national d'Histoire naturelle, Paris ; Biotope, Mèze.

Dale, V.H., Joyce, L.A., McNulty, S., Neilson, R.P., Ayres, M.P., Flannigan, M.D., Hanson, P.J., Irland, L.C., Lugo, A.E., Peterson, C.J., Simberloff, D., Swanson, F.J., Stocks, B.J., Wotton, B.M., 2001. Climate Change and Forest Disturbances: Climate change can affect forests by altering the frequency, intensity, duration, and timing of fire, drought, introduced species, insect and pathogen outbreaks, hurricanes, windstorms, ice storms, or landslides. BioScience 51, 723–734. URL: https://doi.org/10.1641/0006-3568(2001)051[0723:CCAFD]2.0.CO;2, doi:10.1641/0006-3568(2001)051[0723:CCAFD]2.0.CO;2.

Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors 11, 7063–7081. URL: https://doi.org/10.3390/s110707063, doi:10.3390/s110707063.

DGAL, 2018. Surveillance des dépérissements en forêts. Note de service. Direction générale de l'Alimentation (DGAL). URL: https://info.agriculture.gouv.fr/gedei/site/bo-agri/instruction-2018-433/telechargement.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens. Environ. 120, 25 – 36. URL: http://www.sciencedirect.com/science/article/pii/S0034425712000636, doi:https://doi.org/10.1016/j.rse.2011.11.026. the Sentinel Missions - New Opportunities for Science.

Dutrieux, R., Feret, J.B., Ose, K., De Boissieu, F., 2021a. Package Fordead. Software. Recherche Data Gouv. URL: https://doi.org/10.15454/4TEO6H, doi:10.15454/4TEO6H.

Dutrieux, R., Féret, J.B., Ose, K., 2021b. Mise au point d'une méthode reproductible pour le suivi généralisé des dégats de scolytes par télédétection satellitaire. ONF Rendez-vous techniques , (69–70):37–44URL: https://www.onf.fr/onf/+/cec::les-rendez-vous-techniques-de-lonf-no69-70.html.

van Engelen, J.E., Hoos, H.H., 2019. A survey on semi-supervised learning. Mach. Learn. 109, 373–440. URL: https://doi.org/10.1007/s10994-019-05855-6, doi:10.1007/s10994-019-05855-6.

FAO, 2020. Global forest resources assessment 2020: Main report. Food and Agriculture Organization of the United Nations, Rome URL: http://www.fao.org/3/ca9825en/ca9825en.pdf.

Fauvel, M., Lopes, M., Dubo, T., Rivers-Moore, J., Frison, P.L., Gross, N., Ouin, A., 2020. Prediction of plant diversity in grasslands using Sentinel-1 and -2 satellite image time series. Remote Sens. Environ. 237, 111536. URL: https://www.sciencedirect.com/science/article/pii/S0034425719305553, doi:https://doi.org/10.1016/j.rse.2019.111536.

French Ministry for the Ecological and Solidary Transition, 2020. National Low Carbon Strategy. Technical Report. French Ministry for the Ecological and Solidary Transition: Paris, France. URL: https://www.ecologie.gouv.fr/sites/default/files/en_SNBC-2_complete.pdf.

Gao, B., 1996. NDWI — A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257 – 266. doi:https://doi.org/10.1016/S0034-4257(96)00067-3.

Garnot, V.S.F., Landrieu, L., 2020. Lightweight temporal self-attention for classifying satellite images time series, in: Lemaire, V., Malinowski, S., Bagnall, A., Guyet, T., Tavenard, R., Ifrim, G. (Eds.), Workshop on Advanced Analytics and Learning on Temporal Data (AALTD), Springer International Publishing, Ghent, Belgium. pp. 171–181.

Goudet, M., Saintonge, F., Nageleisen, L., 2018. Quantifier l'état de santé de la forêt, méthode simplifiée d'évaluation. Technical Report. Département de la santé des forêts. URL: https://agriculture.gouv.fr/telecharger/90879.

Grabska, E., Hawryło, P., Socha, J., 2020. Continuous detection of small-scale changes in Scots pine dominated stands using dense Sentinel-2 time series. Remote Sens. 12, 1298. URL: http://dx.doi.org/10.3390/rs12081298, doi:10.3390/rs12081298.

Hagolle, O., Huc, M., Villa 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, VENμS and Sentinel-2 images. Remote Sens. 7, 2668–2691. URL: https://www.mdpi.com/2072-4292/7/3/2668, doi:10.3390/rs70302668.

Han, H., Wang, W.Y., Mao, B.H., 2005. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning, in: Huang, D.S., Zhang, X.P., Huang, G.B. (Eds.), Advances in Intelligent Computing, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 878–887.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850–853. URL: https://www.science.org/doi/abs/10.1126/science.1244693, doi:10.1126/science.1244693.

He, H., Bai, Y., Garcia, E.A., Li, S., 2008. ADASYN: Adaptive synthetic sampling approach for imbalanced learning, in: Proc. IEEE Int. Conf. Neural Netw. (IJCNN), Hong Kong, China. pp. 1322–1328. doi:10.1109/IJCNN.2008.4633969.

He, H., Garcia, E., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21, 1263–1284. URL: https://doi.org/10.1109/tkde.2008.239, doi:10.1109/tkde.2008.239.

Holmes, T.P., Huggett, R.J., Pye, J.M., 2008. Forest economics, natural disturbances and the new ecology, in: The Economics of Forest Disturbances. Springer. volume 9, pp. 15–32. URL: https://doi.org/10.1007/978-1-4020-4370-3_2, doi:10.1007/978-1-4020-4370-3_2.

Hornero, A., Zarco-Tejada, P., Quero, J., North, P., Ruiz-Gómez, F., Sánchez-Cuesta, R., Hernandez-Clemente, R., 2021. Modelling hyperspectral- and thermal-based plant traits for the early detection of Phytophthora-induced symptoms in oak decline. Remote Sens. Environ. 263, 112570. URL: https://www.sciencedirect.com/science/article/pii/S003442572100290X, doi:https://doi.org/10.1016/j.rse.2021.112570.

Housman, I., Chastain, R., Finco, M., 2018. An evaluation of forest health insect and disease survey data and satellite-based remote sensing forest change detection methods: Case studies in the United States. Remote Sens. 10, 1184. URL: http://dx.doi.org/10.3390/rs10081184, doi:10.3390/rs10081184.

ying Huang, C., Anderegg, W.R., Asner, G.P., 2019. Remote sensing of forest die-off in the anthropocene: From plant ecophysiology to canopy structure. Remote Sens. Environ. 231, 111233. URL: https://www.sciencedirect.com/science/article/pii/S0034425719302524, doi:https://doi.org/10.1016/j.rse.2019.111233.

Huang, L., Zhao, X., Liu, Y., Yang, P., 2022. Analysis of the atmospheric duct existence factors in tropical cyclones based on the SHAP interpretation of extreme gradient boosting predictions. Remote Sens. 14, 3952. URL: http://dx.doi.org/10.3390/rs14163952, doi:10.3390/rs14163952.

Huang, Z., Turner, B.J., Dury, S.J., Wallis, I.R., Foley, W.J., 2004. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. Remote Sens. Environ. 93, 18–29. URL: https://www.sciencedirect.com/science/article/pii/S0034425704001920, doi:https://doi.org/10.1016/j.rse.2004.06.008.

Huete, A.R., 2012. Vegetation indices, remote sensing and forest monitoring. Geogr. Compass 6, 513–532. URL: https://compass.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-8198.2012.00507.x, doi:https://doi.org/10.1111/j.1749-8198.2012.00507.x.

Huo, L., Persson, H.J., Lindberg, E., 2021. Early detection of forest stress from European spruce bark beetle attack, and a new vegetation index: Normalized distance red & SWIR (NDRS). Remote Sens. Environ. 255, 112240. URL: https://www.sciencedirect.com/science/article/pii/S0034425720306131, doi:https://doi.org/10.1016/j.rse.2020.112240.

Igel, C., Heidrich-Meisner, V., Glasmachers, T., 2008. Shark. J. Mach. Learn. Res. 9, 993–996.

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., et al., 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. Remote Sens. 7, 12356–12379. URL: http://dx.doi.org/10.3390/rs70912356, doi:10.3390/rs70912356.

Inglada, J., Vincent, A., Arias, M., Marais-Sicre, C., 2016a. Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. Remote Sens. 8, 362. URL: http://dx.doi.org/10.3390/rs8050362, doi:10.3390/rs8050362.

Inglada, J., Vincent, A., Arias, M., Tardy, B., 2016b. iota2-a25386. Software. CESBIO. doi:10.5281/zenodo.58150.

Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — temporal segmentation algorithms. Remote Sens. Environ. 114, 2897–2910. URL: https://www.sciencedirect.com/science/article/pii/S0034425710002245, doi:https://doi.org/10.1016/j.rse.2010.07.008.

Konrad Turlej, C., Ozdogan, M., Radeloff, V.C., 2022. Mapping forest types over large areas with Landsat imagery partially affected by clouds and SLC gaps. Int. J. Appl. Earth Obs. Geoinf. 107, 102689. URL: https://www.sciencedirect.com/science/article/pii/S0303243422000150, doi:https://doi.org/10.1016/j.jag.2022.102689.

Krieger, D.J., 2001. Economic value of forest ecosystem services: a review. Technical Report. The Wilderness Society: Washington, DC, USA.

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, 1–5. URL: http://jmlr.org/papers/v18/16-365.html.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67. doi:10.1038/s42256-019-0138-9.

Manion, P.D., 1981. Tree disease concepts. Hardcover ed., Prentice-Hall.

Millar, C.I., Stephenson, N.L., 2015. Temperate forest health in an era of emerging megadisturbance. Science 349, 823–826. URL: https://www.science.org/doi/abs/10.1126/science.aaa9933, doi:10.1126/science.aaa9933.

Mitchell, A.L., Rosenqvist, A., Mora, B., 2017. Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+. Carbon Balance Manag. 12, 9. URL: https://doi.org/10.1186/s13021-017-0078-9, doi:10.1186/s13021-017-0078-9.

Moravec, V., Markonis, Y., Rakovec, O., Svoboda, M., Trnka, M., Kumar, R., Hanel, M., 2021. Europe under multi-year droughts: how severe was the 2014–2018 drought period? Environ. Res. Lett. 16, 034062. URL: https://dx.doi.org/10.1088/1748-9326/abe828, doi:10.1088/1748-9326/abe828.

Morin, D., Planells, M., Guyon, D., Villard, L., Mermoz, S., Bouvet, A., Thevenon, H., Dejoux, J.F., Le Toan, T., Dedieu, G., 2019. Estimation and mapping of forest structure parameters from open access satellite images: Development of a generic method with a study case on coniferous plantation. Remote Sens. 11, 1275. URL: http://dx.doi.org/10.3390/rs11111275, doi:10.3390/rs11111275.

Mouret, F., Albughdadi, M., Duthoit, S., Kouamé, D., Rieu, G., Tourneret, J.Y., 2022. Reconstruction of Sentinel-2 derived time series using robust Gaussian mixture models. Application to the detection of anomalous crop development. Comput. Electron. Agric. 198, 106983. doi:https://doi.org/10.1016/j.compag.2022.106983.

Mulverhill, C., Coops, N.C., Achim, A., 2023. Continuous monitoring and sub-annual change detection in high-latitude forests using harmonized landsat sentinel-2 data. ISPRS J. Photogramm. Remote Sens. 197, 309–319. URL: https://www.sciencedirect.com/science/article/pii/S0924271623000424, doi:https://doi.org/10.1016/j.isprsjprs.2023.02.002.

Olsen, J., Ceccato, P., Proud, S., Fensholt, R., Grippa, M., Mougin, E., Ardö, J., Sandholt, I., 2013. Relation between seasonally detrended shortwave infrared reflectance data and land surface moisture in semi-arid sahel. Remote Sens. 5, 2898–2927. URL: http://dx.doi.org/10.3390/rs5062898, doi:10.3390/rs5062898.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Rodríguez-Calcerrada, J., Sancho-Knapik, D., Martin-StPaul, N.K., Limousin, J.M., McDowell, N.G., Gil-Pelegrín, E., 2017. Drought-induced oak decline—factors involved, physiological dysfunctions, and potential attenuation by forestry practices, in: Oaks Physiological Ecology. Exploring the Functional Diversity of Genus *Quercus* L.. Springer International Publishing, pp. 419–451. URL: https://doi.org/10.1007/978-3-319-69099-5_13, doi:10.1007/978-3-319-69099-5_13.

Rouse, J., Haas, R., Schell, J., Deering, D., 1974. Monitoring vegetation systems in the great plains with ERTS. NASA special publication 351, 309.

Sangüesa-Barreda, G., Linares, J.C., Camarero, J.J., 2015. Reduced growth sensitivity to climate in bark-beetle infested Aleppo pines: Connecting climatic and biotic drivers of forest dieback. For. Ecol. Manag. 357, 126–137. URL: https://www.sciencedirect.com/science/article/pii/S0378112715004429, doi:https://doi.org/10.1016/j.foreco.2015.08.017.

Sapes, G., Lapadat, C., Schweiger, A.K., Juzwik, J., Montgomery, R., Gholizadeh, H., Townsend, P.A., Gamon, J.A., Cavender-Bares, J., 2022. Canopy spectral reflectance detects oak wilt at the landscape scale using phylogenetic discrimination. Remote Sens. Environ. 273, 112961. URL:

https://www.sciencedirect.com/science/article/pii/S003442572200075X, doi:https://doi.org/10.1016/j.rse.2022.112961.

Seidl, R., Thom, D., Kautz, M., Martín-Benito, D., Peltoniemi, M., Vacchiano, G., Wild, J., Ascoli, D., Petr, M., Honkaniemi, J., Lexer, M., Trotsiuk, V., Mairota, P., Svoboda, M., Fabrika, M., Nagel, T., Reyer, C., 2017. Forest disturbances under climate change. Nature Clim. Change 7, 395–402. doi:https://doi.org/10.1038/nclimate3303.

Stahl, A.T., Andrus, R., Hicke, J.A., Hudak, A.T., Bright, B.C., Meddens, A.J., 2023. Automated attribution of forest disturbance types from remote sensing data: A synthesis. Remote Sens. Environ. 285, 113416. URL: https://www.sciencedirect.com/science/article/pii/S0034425722005223, doi:https://doi.org/10.1016/j.rse.2022.113416.

Thierion, V., Vincent, A., Valero, S., 2022. Theia OSO Land Cover Map 2021. Dataset. CESBIO. doi:10.5281/zenodo.6538910.

Torres, P., Rodes-Blanco, M., Viana-Soto, A., Nieto, H., García, M., 2021. The role of remote sensing for the assessment and monitoring of forest health: A systematic evidence synthesis. Forests 12, 1134. URL: http://dx.doi.org/10.3390/f12081134, doi:10.3390/f12081134.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., Altman, R.B., 2001. Missing value estimation methods for DNA microarrays . Bioinformatics 17, 520–525. doi:10.1093/bioinformatics/17.6.520.

Turner, M.G., 2010. Disturbance and landscape dynamics in a changing world. Ecology 91, 2833–2849. URL: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/10-0097.1, doi:https://doi.org/10.1890/10-0097.1.

Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. Remote Sens. Environ. 114, 106–115. URL: https://www.sciencedirect.com/science/article/pii/S003442570900265X, doi:https://doi.org/10.1016/j.rse.2009.08.014.

Vuolo, F., Ng, W.T., Atzberger, C., 2017. Smoothing and gap-filling of high resolution multi-spectral time series: Example of Landsat data. Int. J. Appl. Earth Obs. Geoinf. 57, 202–213. URL: https://www.sciencedirect.com/science/article/pii/S0303243416302100, doi:https://doi.org/10.1016/j.jag.2016.12.012.

Wang, Z., Yan, W., Oates, T., 2017. Time series classification from scratch with deep neural networks: A strong baseline, in: Proc. IJCNN, IEEE, Anchorage, Alaska, USA. URL: https://doi.org/10.1109/ijcnn.2017.7966039, doi:10.1109/ijcnn.2017.7966039.

Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. J. Big Data 3. URL: https://doi.org/10.1186/s40537-016-0043-6, doi:10.1186/s40537-016-0043-6.

Wulder, M.A., Roy, D.P., Radeloff, V.C., Loveland, T.R., Anderson, M.C., Johnson, D.M., Healey, S., Zhu, Z., Scambos, T.A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C.J., Masek, J.G., Hermosilla, T., White, J.C., Belward, A.S., Schaaf, C., Woodcock, C.E., Huntington, J.L., Lymburner, L., Hostert, P., Gao, F., Lyapustin, A., Pekel, J.F., Strobl, P., Cook, B.D., 2022. Fifty years of Landsat science and impacts. Remote Sens. Environ. 280, 113195. URL: https://www.sciencedirect.com/science/article/pii/S0034425722003054, doi:https://doi.org/10.1016/j.rse.2022.113195.

Yang, Y., Anderson, M.C., Gao, F., Wood, J.D., Gu, L., Hain, C., 2021. Studying drought-induced forest mortality using high spatiotemporal resolution evapotranspiration data from thermal satellite imaging. Remote Sens. Environ. 265, 112640. URL: https://www.sciencedirect.com/science/article/pii/S0034425721003606, doi:https://doi.org/10.1016/j.rse.2021.112640.

Zabihi, K., Surovy, P., Trubin, A., Singh, V.V., Jakuš, R., 2021. A review of major factors influencing the accuracy of mapping green-attack stage of bark beetle infestations using satellite imagery: Prospects to avoid data redundancy. Remote Sens. Appl.: Soc. Environ. 24, 100638. URL: https://www.sciencedirect.com/science/article/pii/S2352938521001749, doi:https://doi.org/10.1016/j.rsase.2021.100638.

Zarco-Tejada, P., Hornero, A., Hernández-Clemente, R., Beck, P., 2018. Understanding the temporal dimension of the red-edge spectral region for forest decline detection using high-resolution hyperspectral and Sentinel-2A imagery. ISPRS J. Photogramm. Remote Sens. 137, 134–148. URL: https://doi.org/10.1016/j.isprsjprs.2018.01.017, doi:10.1016/j.isprsjprs.2018.01.017.

Zarco-Tejada, P., Pushnik, J., Dobrowski, S., Ustin, S., 2003. Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. Remote Sens. Environ. 84, 283–294. URL: https://www.sciencedirect.com/science/article/pii/S003442570200113X, doi:https://doi.org/10.1016/S0034-4257(02)00113-X.

Zhang, M., Wei, X., 2021. Deforestation, forestation, and water supply. Science 371, 990–991. URL: https://www.science.org/doi/abs/10.1126/science.abe7821, doi:10.1126/science.abe7821.