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

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 , and Cécile Vincent-Barbaroux

Abstract—This article studies the monitoring of oak dieback in forests of the Centre-Val de Loire region (France), where drought-induced dieback has become a major concern due to climate change. The main objective of the study is to evaluate the applicability of multispectral satellite time series for operational monitoring of forest dieback. Using in situ data collected from 2017 to 2022 on approximately 2700 oak plots, a multiyear mapping of the analyzed region was performed using the random forest algorithm and Sentinel-2 images. Our results show that it is possible to detect oak dieback accurately (average overall accuracy = 80% and average balanced accuracy = 79%). A spatial cross-validation analysis also evaluates the performance of the model on regions that were never encountered during training, across all years, resulting in a slight decrease in accuracy (∼5%). The study also highlights the importance of measuring the stability and performance of the classification model over time, in addition to standard cross-validation metrics. A feature analysis shows that the shortwave infrared part of the spectrum is the most important for mapping forest dieback, while the red-edge portion of the spectrum can increase the stability of the model over time. Overall, both in situ data and model predictions showed evidence of forest decline in many areas of the study region. Our results suggest 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.

Index Terms—Climate change, dieback detection, forest monitoring, machine learning (ML), random forest, remote sensing, sentinel-2.

I. INTRODUCTION

CLIMATE change is expected to increase the severity and frequency of forest disturbances, whether abiotic (e.g.,

Florian Mouret is with the Université of Orléans, USC INRAE 1328/P2E laboratory (Physiology, Ecology, and Environment), 45100 Orléans, France, and also with the CESBIO, Université de Toulouse, CNES/CNRS/INRAE/IRD/UT3-Paul Sabatier, 31401 Toulouse, France (e-mail: florian.mouret@univ-orleans.fr).

Cécile Vincent-Barbaroux is with the Université of Orléans, USC INRAE 1328/P2E laboratory (Physiology, Ecology and Environment), 45100 Orléans, France (e-mail: cecile.vincent1@univ-orleans.fr).

David Morin and Milena Planells are with the CESBIO, Université de Toulouse, CNES/CNRS/INRAE/IRD/UT3-Paul Sabatier, 31401 Toulouse, France (e-mail: david.morin@univ-tlse3.fr; milena.planells-rodriguez@univ-tlse3.fr).

Hilaire Martin is with the INRAE, UR EFNO, Biodiversity Team, Domaine des Barres, 45290 Nogent-sur-Vernisson, France (e-mail: hilaire.martin@inrae.fr).

Digital Object Identifier 10.1109/JSTARS.2023.3332420

fire, drought) and biotic (e.g., insects, pathogens) [1], [2], [3]. Assessment and monitoring of such disturbances is essential due to the central role of forests in the hydrological cycle [4], carbon sequestration and biodiversity conservation [5]. Forests are also of great economic importance, for example, by providing wood and timber [6]. As a result, their disturbance can cause significant economic losses [7]. Finally, forests are often important in national low carbon strategies (i.e., [8]). 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 [9].

In temperate forests, prolonged, repeated or exceptional droughts combined with hotter temperatures (also called "hotter droughts" [10]) can push forests beyond their sustainability threshold [9]. The effects of hotter droughts, exacerbated by climate change, range from forest dieback to increased tree mortality to broad-scale forest die-off. While extreme climate conditions can impact extensive forest regions, they can intensify the vulnerability of specific trees or areas, leading to localized instances of dieback. These specific areas may become more vulnerable due to a combination of factors, including altered microclimates, variations in soil composition and quality, increased competition, and heightened susceptibility to pests and diseases [10]. Forest dieback is a complex and evolving phenomenon with multifactorial causes that results in a progressive weakening of trees and stands vigor [11]. The symptoms of dieback are essentially visible as a reduction in leaf area and crown and has been identified as a factor reducing the resilience of ecosystems [12]. France, like most of the European continent, has been affected by severe droughts in recent years [13], [14], 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.

Remote sensing has been widely recognized as a valuable tool for monitoring forest status [15], [16], [17], [18]. 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 [17]. In recent years, access to consistent and freely available multispectral satellite data has been facilitated by the opening of the Landsat archive in 2008 [19] and the launch of the two Sentinel-2 (S2) satellites in 2015 and 2017 [20]. The arrival of S2 data has opened up unprecedented opportunities: S2 data have a shorter revisit time ($\sim$ 5 days with S2-A and S2-B) and a finer spatial resolution (up to 10 m) compared to Landsat data (revisit time $\sim$ 7–16 days, spatial resolution up to 30 m). The interest of these improvements has been identified in various cases, e.g., for forest species mapping [21] or early detection of bark beetle attacks [22]. 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 [18], [23], [24], [25], [26], [27], [28]. As pointed out in [23], 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 [29]: 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) [30] and the Landsat-based Detection of Trends in Disturbance and Recovery (LandTrendr) [31] approaches. More recently, the FORDEAD package [32] has been developed to detect bark beetle outbreaks in spruce trees. One drawback of such approaches is that they require historical data to properly define what is normal behavior. Moreover, Stahl et al. [18] 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 [17]. 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. [17], most of the techniques used to monitor forest health are classification or regression techniques. Among them, the randomforest (RF) algorithm [33] 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 Torres et al. [18] 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 algorithms are known to be easier to interpret, which can be useful to better understand the problems being modeled [34]. Note that we also assessed other classification algorithms in our analysis, including the support vector machine (SVM)

algorithm [35] and deep learning methods. A brief discussion on this topic is included in this document.

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 the RF algorithm with S2 data. Our main goal is to establish a new approach to oak dieback monitoring that is able to: 1) separate healthy and declining areas as accurately as possible and 2) produce maps on a large scale and for several years in an operational monitoring system.

The rest of this article is organized as follows. Section II 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 III, the method used to map the forest dieback is provided, including details on how to handle reference data coming from different years. In Section IV, 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 finally performed. Section V discusses these results and provides some additional insights related to the problem at hand. Finally, Section VI summarizes and concludes this work.

II. STUDY AREA AND DATA

A. 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 tiles are the one provided by the Theia platform, (https://theia.cnes.fr , accessed on 17 July 2023). As shown in Fig. 1, the study area is centered approximately at 47°7'N latitude and 1°8'E longitude (Northern France), and covers 11 S2 tiles (110 000 km^2), with 23% of deciduous forest according to the OSO land cover map [36]. Centre-Val de Loire Region corresponds to a plateau surrounded by shallow valleys (max. altitude 500 m above sea mean level (ASML), avg. 140 m ASML). 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 [37]. 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.*) [37]. This study focuses on oaks (*Quercus robur* and *Quercus petraea*), which are a key species in the region.

B. Reference Data

1) Labeling Protocol: The health status of the reference data is assessed using the DEPERIS protocol [38]. This protocol is
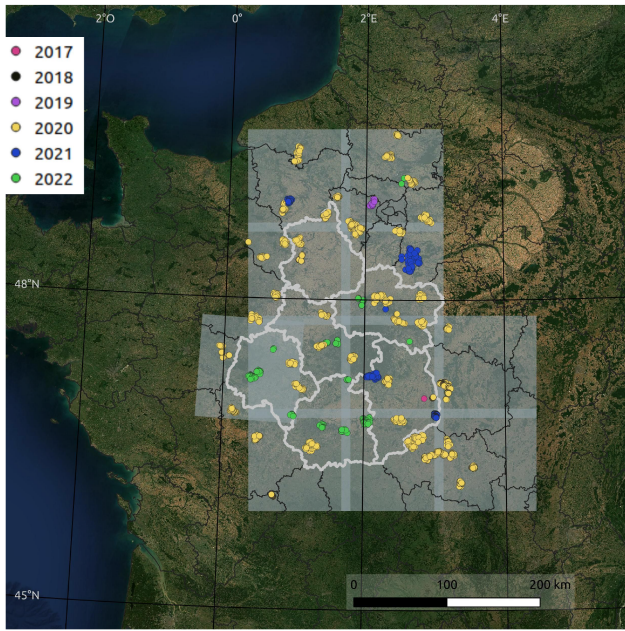
Fig. 1. Extent of the studied area is delimited by the gray area (the boundaries between the 11 tiles is highlighted in lighter gray). The frontier of the region Centre-Val de Loire and its departments is in white. Finally, the colored dots locate the reference data, with each color representing a labeling year (for reference data labeled multiple times, the last labeling year is highlighted). The background uses cloudless S2 images.

Fig. 2. (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) Plots in Montargis forest labeled in 2020 (cyan/orange/red colors correspond to healthy, declining, and very declining plots, respectively). A S2 image acquired in 2020 is used as background.

used by the French Forest Health Service and is currently the official protocol for forest dieback monitoring in France [39]. 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 [38]). Forest decline 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 Fig. 2(a) with an example, where 25% of the trees in the plot have scores equal to or higher than "D" (declining plot), while Fig. 2(b) gives concrete examples in the forest of Montargis, illustrating that plots that are very close to each other can have a different health statuses.

2) Distribution of the Reference Data: The distribution of the reference data with respect to the labeling year is given in Table I and shown in Fig. 1. More than half of the plots were labeled in 2020, 71% of which were labeled healthy. This year, a campaign was conducted in France to assess the health status of oak forest through a random road sampling [40], following the successive droughts of 2018/2019/2020. This campaign is of crucial importance since it is a systematic sampling of the
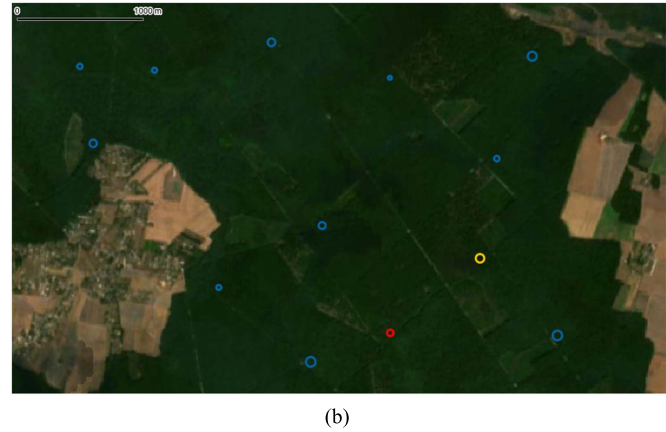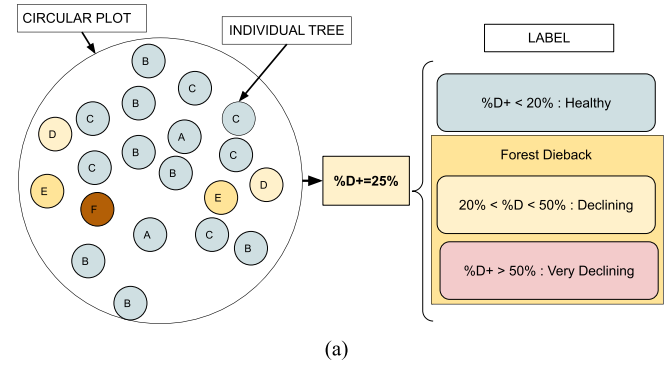
TABLE I
REFERENCE DATA PER YEAR AND DIEBACK CATEGORY

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%

analyzed area [see for example Fig. 2(b)]. Using the same protocol, the plots labeled the other years come from different research campaigns carried out by private or public foresters (see Acknowledgments at the end of this document), with the proportion of healthy plots varying from 17% (2021) to 57% (2017). Their spatial distribution is not uniform throughout the study area, but they are located in different massifs and regions, allowing us to monitor the temporal evolution and diversity of dieback in different characteristic sites. Based on the distribution of the reference data, two factors have been carefully considered for our analysis and the results presented in Section IV. First, since the study area is not completely covered by reference data, except for 2020, it is necessary to estimate the accuracy of the
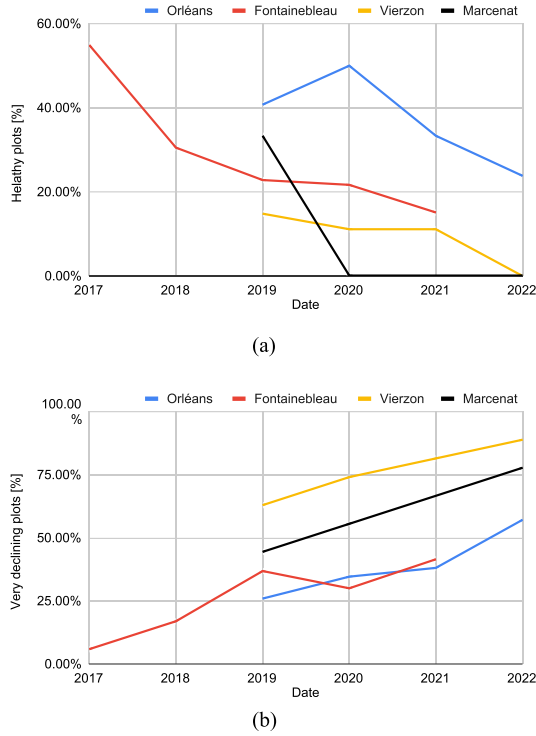
Fig. 3. (a) Percentage of healthy plots and (b) percentage of very declining plots with respect to the labeling year in the Orléans (blue), Fontainebleau (red), Vierzon (yellow), and Marcénat (black) massifs. Note that for Marcénat, no data was labeled in 2021.

TABLE II
SENTINEL-2 MULTISPECTRAL BANDS USED FOR THE ANALYSIS

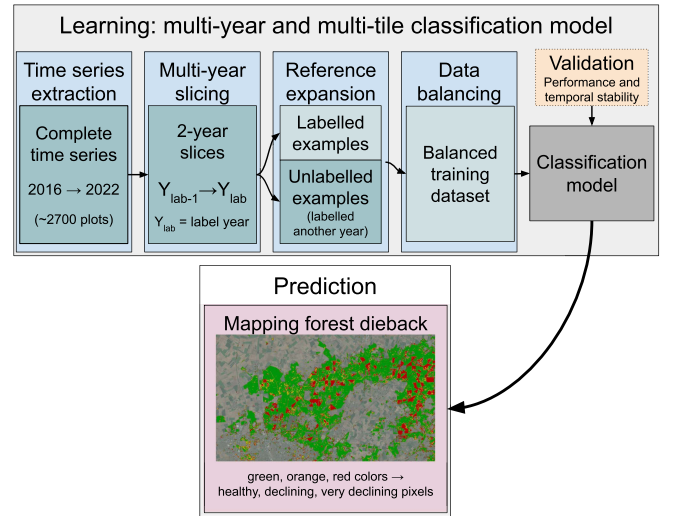| Spectral bands	Central wavelength (nm)	Bandwith (nm)	Resolution (m)
B2: Blue	490	65	10
B3: Green	560	35	10
B4: Red	665	30	10
B5: Vegetation Red Edge (RE)	705	15	20
B6: Vegetation Red Edge(RE)	740	15	20
B7: Vegetation Red Edge(RE)	783	20	20
B8: Near Infrared (NIR)	842	115	10
B8A: Narrow Near Infrared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

Fig. 4. Diagram summarizing the proposed methodological steps for the classification of forest dieback.

maps in areas that are partially or never observed over time. Second, changes in plot locations from one year to the next partly cause variations in the proportions of healthy and declining plots over time (the natural progression of forest dieback is obviously another source of variation), which motivates the use of validation metrics adapted to deal with imbalanced data. Finally, considering the entire dataset, the proportions of healthy, declining and very declining plots are 55.77%, 24.95%, and 19.28%, respectively.

Among the reference data, four forest massifs were labeled several years in a row (the number of labeled plots may vary slightly depending on the year and the forestry interventions). The Fontainebleau massif (∼55 plots) was labeled between 2017 and 2021. Orléans (∼25 plots), Vierzon (∼27 plots), and Marcénat (9 plots) forests were labeled between 2019 and 2022. Fig. 3 gives the percentage of healthy plots [see Fig. 3(a)] and very declining plots [see Fig. 3(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.

C. Satellite Data

This study used data from the Sentinel-2 satellites (S2-A and S2-B), which 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. The ten spectral bands considered in this study have a spatial resolution of 10 to 20 m [20], as depicted in Table II. The MAJA processing chain [41] was used to produce level 2 A images, which are orthorectified products expressed in surface reflectance with cloud and shadow masks. MAJA implements a multitemporal methodology, utilizing the temporal resolution of Sentinel-2 imagery, to identify clouds and their shadows. In addition, MAJA calculates the aerosol optical thickness and water vapor to adjust a specific image for atmospheric effects such as absorption and scattering.

III. METHODS

The proposed approach for multiyear forest dieback mapping is summarized in Fig. 4. The remainder of this section details each of the methodological steps of this diagram. Note that the algorithm operates at the pixel level. Therefore, when the model is trained, all pixels in a plot receive the plot label. This increase the variability of the training set, see Section V for a discussion and perspectives on this point.
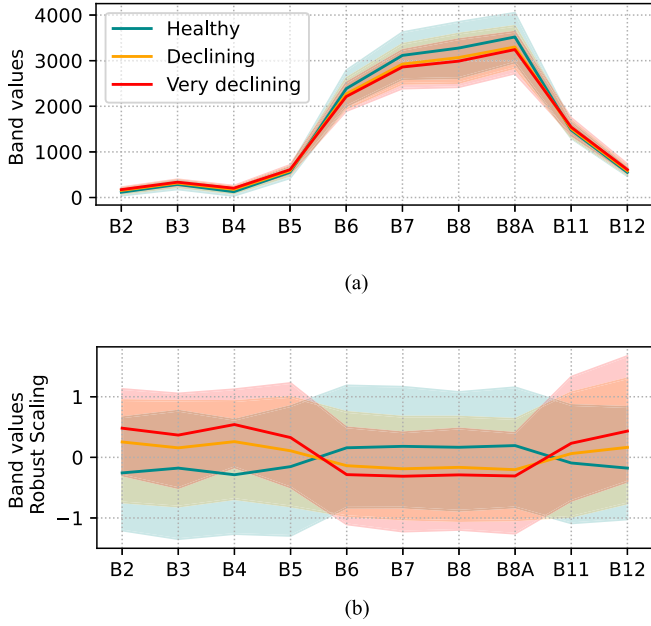
(a)

(b)

Fig. 5. 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.

A. Time Series Extraction

1) S2 Bands: In the time series extraction step, time series of each S2 band coming from each pixel within the reference plots are extracted [see Fig. 2(b) for examples of circular plots]. This extraction is done using the $iota^2$ software [42], 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 [43], [44], [45]. 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 V. 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, Fig. 5 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 Fig. 5(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-edge (B6, B7) and NIR parts (B8, B8a) of the spectrum, declining pixels generally have a lower reflectance compared to healthy pixels.

2) Vegetation Indices: In addition to the raw bands, it is classical to also compute 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 Index (NDVI), which is mainly related to the plant vigor [46], [47] and is used to detect anomalies in the BFAST [30] or LandTrendr [31] algorithms mentioned in the introduction of this article.

In the course of our experiments, we have tested different 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 V. The CRswir and CRre formulas can be expressed as follows:

$$CR_{swir} = \frac{B11}{B8a + (\lambda_{B11} - \lambda_{8a}) \times \left(\frac{B12 - B8a}{\lambda_{B12} - \lambda_{8a}}\right)} \quad (1)$$

$$CR_{re} = \frac{B5}{B4 + (\lambda_{B5} - \lambda_{B4}) \times \left(\frac{B6 - B4}{\lambda_{B6} - \lambda_4}\right)} \quad (2)$$

where Bn and λ_{Bn} are the reflectance and the wavelength in nanometers of the band n, respectively (see Table II for the corresponding values). The CRswir was successfully used for the mapping of bark beetle outbreaks in the FORDEAD package [32], [48]. CR aims at isolating individual absorption of interest [49], and has been mainly used with hyperspectral remote sensing image [50]. 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 two 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 complex indices. A graphical illustration of CRre and CRswir is given in Fig. 6. 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 [51], [52], while the red-edge part has been found to be sensitive to canopy chlorophyll content [53], [54], [55], [56].

B. Multiyear Slicing and Creation of the Feature Matrix

Standard classification algorithms assume that each sample to be classified is characterized by a fixed number of features. Thus, in our case, each pixel must be described by time series of fixed length, ending with the year of labeling. The length of the temporal slice, denoted N_{years}, can cover for example two years before labeling, as illustrated in Fig. 7. These slices, with the
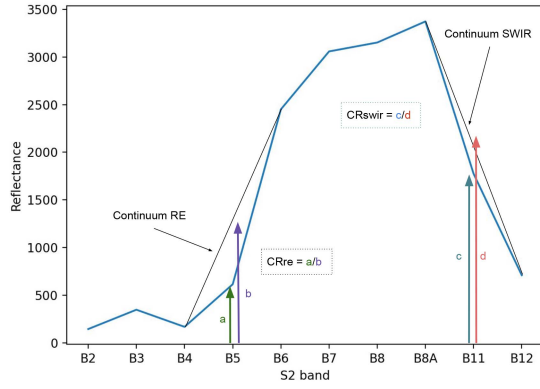
Fig. 6. Reflectance of a healthy plot (Orléans forest) in summer (29/06/2020) and its continuum RE and SWIR (inspired by the figure in [48]).
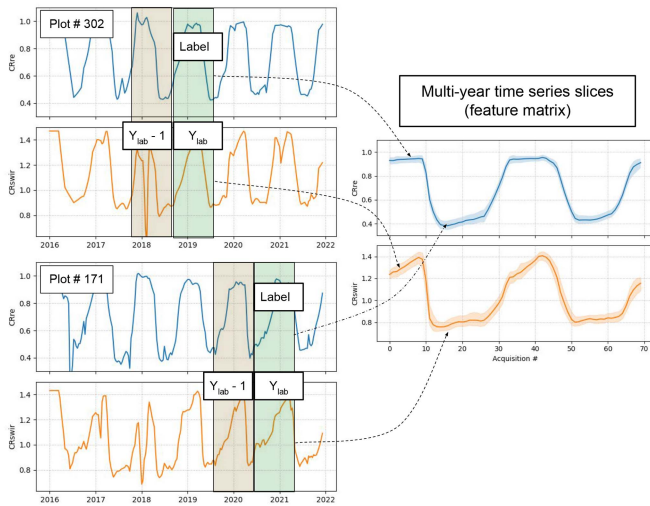

Fig. 7. Illustration of the multiyear slicing used to create the feature matrix when using two indices (CRre and CRswir). The left part of the image shows the CRswir (top blue line) and CRre (bottom orange line) time series of two different plots over time. Each plot is characterized by time series slices of fixed length (colored areas), the end of which depends on the different labeling years of the plots (Y_{lab}). The different time series slices (left part) are used to create the feature matrix (right part). 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 $N_{years} = 2$.

same dates (days/months) but different years, are used to train a multiyear classification model. The multiyear mix makes the resulting classification model more robust to seasonal changes between years (e.g., phenological differences).

To analyze the potential interest of CR indices, Fig. 8 provides time series of the median and interquartile range of the CRswir and CRre indices acquired over 2 years before labeling ($N_{years} = 2$) and grouped by declining classes. The different years are mixed according to the year of labeling Y_{lab} (here, Year 2 is Y_{lab} and Year 1 is $Y_{lab} - 1$), as shown in Fig. 7. It can be observed 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
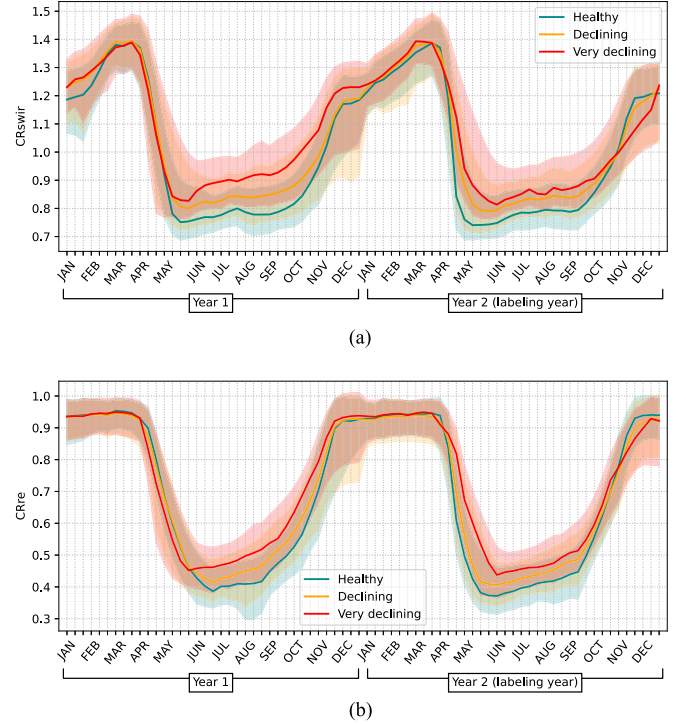

Fig. 8. Time series of (a) CRswir and (b) CRre indices of the learning dataset acquired over 2 years prior to labeling (Year 1 is the year before labeling and Year 2 is the year of labeling). The colors cyan/orange/red correspond respectively to healthy, declining and very declining plots based on the percentage of trees with grades lower than D (see detailed examples in Fig. 2). The solid line corresponds to the median value of the class and the shaded area to its interquartile range.

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 V).

C. Expanded Training References From Unlabeled Time Series Slices

Accessing more training data, when possible, is a common practice to improve the accuracy of ML models [57]. In our case, it seems intuitive that a broader and more consistent characterization of oak dieback patterns over time would be possible with a greater number of examples across years and massifs, especially in areas that are not consistently marked each year. We propose to expand our training dataset with unlabeled time series slices from plots that are only labeled for a specific year (denoted Y_{lab}), such as plots labeled in 2020 but without labels for the other years. To avoid adding noisy or misleading examples, a two-condition procedure is used to select the time series slices for which we have high confidence in the label to be assigned:

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

2) A declining plot (percentage of declining trees higher than 20%) labeled year Y_{lab} was likely to be declining the years $Y_{lab} + 1$ and $Y_{lab} + 2$ and is added to the dataset with its correspond time slice $(Y_{lab} + 1, Y_{lab} + N_{years})$ and $(Y_{lab} + 2, Y_{lab} + N_{years} + 1)$.

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 (which is consistent with our reference data labeled several years in a row, as well as other studies, see for example, [58]). It is also assumed that there was no silvicultural intervention during this period, or at least that it concerns a limited amount of pixels. In Section IV, the interest of this data expansion step is evaluated w.r.t. using only the raw reference data. In Section V, we discuss on possible improvement of this procedure.

Finally, note that this additional examples are never used to validate the performances of the models, they are only used as additional examples to train the model. Moreover, the plots already labeled the years $Y_{lab} - 2$, $Y_{lab} - 1$, Y_{lab}, $Y_{lab} + 1$, and $Y_{lab} + 2$ are obviously not duplicated in the dataset to avoid data leakage.

D. Data Balancing

Depending on the year, the different classes to be classified are not evenly distributed, which is typically known as an imbalance problem [59]. We propose to balance the dataset each year independently (i.e., for each labeling year, we have the same number of samples coming from the three different classes) using the Synthetic Minority Oversampling Technique (SMOTE), which has been used successfully in many applications [60]. In short, the SMOTE technique generates synthetic data based on the similarity in feature space between existing minority instances. The main advantage of oversampling techniques over undersampling techniques is that all available examples are kept in the training set. Other variants of the SMOTE algorithm were tested without improvement in our results (see Section V). The SMOTE implementation used in this study is the one provided in the Python library imbalance-learn (version 0.10.1) [61].

E. Classification

For the operational classification step, we used the RF algorithm [33], which has shown very successful results when applied to remote sensing data and is therefore widely used in this community [17], [62]. As presented in Appendix A-B, we also tested other state-of-the-art algorithms, such as XGBoost [63] or deep learning approaches adapted to time series [64] and all of these algorithms provided comparable classification results (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 ML library [65]. 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

Experiment	Evaluated factor	Sec.
Standard CV	Accuracy for areas covered by references	Section IV-A1
SLOO CV	Accuracy for unseen areas (time/space)	Section IV-B
Pred. over time	Temporal stability of the mapping	Section IV-C
Temporal CV	Forecasting	Appendix A-C

CV stands for cross-validation and sloo for spatial leave one out.

scikit-learn implementation of RF (version 1.2) since it provides feature importance [66]. 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 $n_{trees} = 1000$ 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.

Finally, 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 have highlighted the potential benefit of having three classes (as defined in Fig. 2) for a more intuitive appropriation of the generated maps. In the 2-class problem, the model is first learned with three classes and the predictions of the declining classes are merged (using directly two classes lead to very close results).

F. Validation Procedure

The performance of our classification framework was validated by looking at different factors that could influence the mapping accuracy within the study area. The main experiments conducted to evaluate these factors are summarized in Table III and detailed in what follows, CV stands for cross-validation and SLOO for spatial leave one out. In all experiments, the training and test sets are separated at the plot-level to avoid data leakage and autocorrelation [67], i.e., pixels of the same plot are grouped in the same set. The plot's labels are predicted 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.

1) Accuracy Metrics: Standard accuracy metrics are used, namely overall accuracy (OA), F1 scores, and balanced accuracy (BA). 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). BA is defined as the average of the recalls obtained for each class ([66]). Unlike OA, BA is not affected by imbalanced datasets, since the recall is expressed as a percentage for each class. Therefore, OA can be biased toward the majority class and thus be a misleading performance measure for imbalanced datasets [68], which is typically the case in our study. In this respect, our analysis emphasize the results expressed in BA.

2) Standard CV: To evaluate the accuracy of the model on random subsets of the reference plots, classification results are first validated by repeated cross-validation (CV): fivefold CV are repeated ten times, with stratification done to account for the proportion of each class by year.

3) Sloocv: Spatial dependence between training and test sets has been identified as a potential problem in the evaluation of mapping accuracy in remote sensing. Spatial leave one out cross validation (SLOOCV) can mitigate this problem by removing from the training set all the samples too close from the tested plot [67]. Intuitively, while a standard CV can provide interesting information on areas covered by reference data, it can be overly optimistic on areas without reference data. Conversely, spatial CV has been identified as potentially overly pessimistic on areas covered by reference data since important information is removed from the covariate space [69]. Therefore, we have performed both standard and SLOO CV, the former providing information on areas well covered by reference data and the latter estimating the performance of the model on areas poorly covered by reference data. In the results presented in Section IV-B, we use a spatial buffer of 10 km, which is generally largely sufficient to exclude an entire massif and ensure a good spatial independence of the S2 pixels (see [67]). Similar accuracy results were obtained with a spatial buffer of 20 km but are not presented here for brevity. Since all samples within the spatial buffer are removed, even if they were labeled in a different year, this means that the test areas were never covered by reference data, which is a challenging scenario. The (S)LOOCV was repeated 300 times for each labeling year.

4) Prediction Over Time: To measure the temporal stability of the different classification frameworks, we have analyzed the prediction of reference pixels over time. Using the standard CV procedure, test pixels are classified over the years (2017 to 2022), even if we do not know the exact label of the pixels each year. Then, we have computed the percentage of declining pixels that are classified as healthy the next year.

Changes from class "healthy" to "declining" are acceptable, since an increase in the number of declining pixels is expected over time due. However, changes from "declining" to "healthy" are suspicious, since the recovery of oak dieback is expected to be slow. It is especially true if the change from declining to healthy concern a large part of the study area (e.g., pessimistic mapping in 2019 and optimistic mapping in 2020). In our experiments, we measure the percentage of pixels changing from "declining" to "healthy" over time to evaluate the stability of the classification models, which is calculated as the number of declining pixels year $Y - 1$ that become healthy year Y divided by the number of pixels in the test set.

G. Feature Importance Analysis With Random Forest

The RF algorithm has the advantage of natively providing feature importance, which is computed as the (normalized) total reduction in Gini impurity introduced by a feature. Note that as a complement, we also used the Python library SHAP (version 0.41.0) with the Tree Explainer method proposed by Lundberg
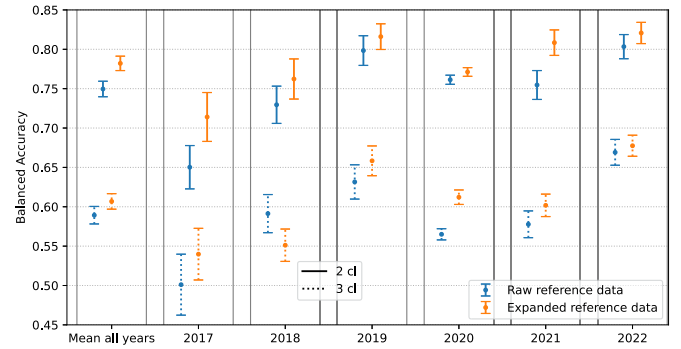

Fig. 9. Balanced accuracy obtained after cross validation with 95% confidence interval (CI). The features used as input are the CRswir and CRre acquired over two years (N years $= 2$). Plain error bars are obtained for a binary classification (healthy/declining) and dashed error bars for three classes.

et al. [70], which leads to similar conclusions. For the sake of simplicity, we do not present detailed results here, but it is interesting to note that both analyses agree.

IV. 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. Second, we focus on explaining these results and provide some insights into the multispectral response of forest dieback.

A. Classification Results

For the sake of brevity, this section will focus on the results obtained using the RF algorithm with the CRswir and CRre acquired over two years (N years $= 2$). Two main framework are compared, namely *raw reference data* and *expanded reference data*. The former is a baseline approach, which trains a model without data expansion and yearly balancing. The latter is the framework proposed in Fig. 4. Additional classification results obtained with other sets of feature or classification algorithms are discussed in Section V, some of them being reported in Appendix A.

1) Standard CV: The results obtained after the standard CV experiment are reported in Fig. 9. These results show the accurate separation between healthy and declining classes (two class scenario), with an average BA equal to 78.4% when using the proposed framework (the average OA obtained is equal to 79%). One can notice that using data expansion w.r.t. using the raw reference data significantly increase the BA in almost all configurations. The results are also given for three classes and highlight the difficulty in separating declining from very declining plots, hence the lower accuracy in the three class scenario. Looking at the F1 values of each class (see Appendix A-A), we can see that the middle class (declining) is the most difficult to classify, while the very declining class is more accurately detected.
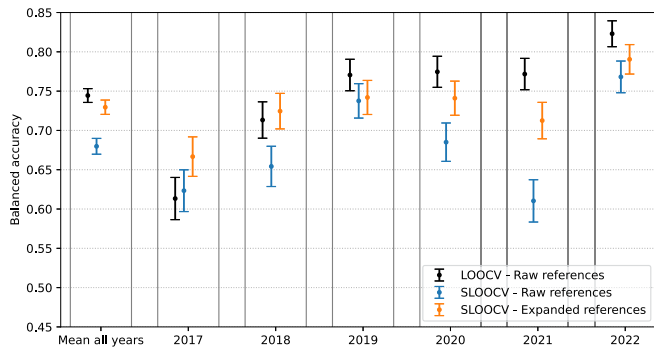
Fig. 10. Balanced accuracy (2 classes) obtained after leave one out cross validation (LOOCV) with 95% confidence interval (CI). SLOOCV means that a buffer has been applied to remove training data around the tested plots. The features used as input are the CRswir and CRre acquired over two years (N years = 2).

B. SLOOCV

To evaluate the robustness of the mapping in areas without reference data, Fig. 10 reports the results obtained after SLOOCV with and without data expansion. For comparison, the results obtained without spatial buffer using the raw reference data are shown in black. Overall, the dieback mapping frameworks are robust in areas without reference data, with a decrease in BA of about 5% compared to the standard CV (for three classes, the decrease in BA is slightly higher, almost 7%). Similarly, the decrease in OA is about 5%. Finally, the use of expanded references leads again to better BA overall, especially with results more consistent over the different years.

C. Temporal Stability

To evaluate the temporal stability of the oak dieback predictions over time, Fig. 11(a) measures the percentage of classified declining pixel that are classified healthy the next year (see definition in Section III-F4) and Fig. 11(b) provides the percentage of classified declining pixels. We compare the classification obtained using the raw and expanded reference frameworks with the CRswir and CRre indices acquired over 2 years before labeling.

The results displayed in Fig. 11(a) show that using only the raw reference data 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, using data expansion and balancing leads to much more stable results, with a percentage of changes in the range 10–15%. Looking at Fig. 11(b), we can see that the percentage of declining pixels oscillates when using only the raw reference data. 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 full proposed framework, the percentage of declining pixels constantly increases over time, which is more consistent with the successive droughts that began in 2018. These results (which
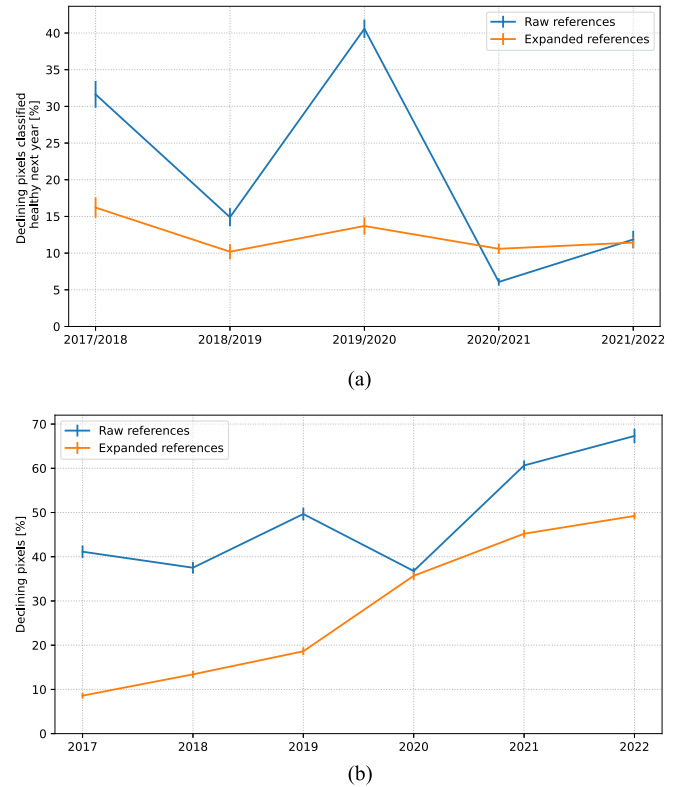

(a)

(b)

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

can also be appreciated visually, see Section V) show that using data expansion and balancing leads to a model with significantly better temporal stability. Moreover, they also show that the use of standard classifications metrics obtained via standard CV may not be sufficient to have a complete overview of the model's behavior.

Finally, note that in Fig. 11(b), the number of pixels in decline (close to 50% in 2022) is not representative of the whole Centre-Val de Loire region, since some areas have more reference points. Looking only at the pixels coming from the systematic road sampling done in 2020 (see Section II-B2), the percentage of pixels classified as declining in 2022 is close to 25%.

D. Feature Analysis

The feature importance obtained using the CR indices (CRswir + CRre) or S2 bands are shown in Fig. 12(a) and (b), respectively (yellow color indicates higher importance). The features are displayed in chronological order, where Y_1 is the year before labeling and Y_2 the year of labeling (i.e., first and second year of acquisition). 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.
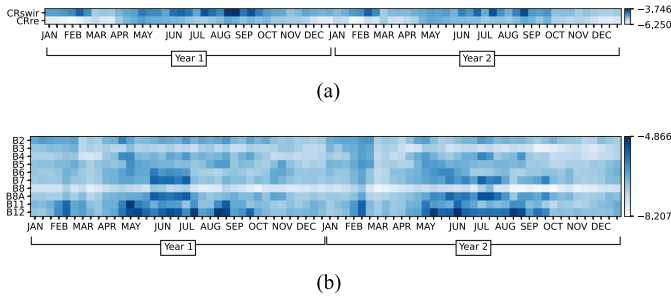
Fig. 12. Log-feature importance obtained when training the model using (a) CRswir and CRre indices and (b) S2 bands, acquired over 2 years before labeling. The features are ordered in temporal order, Year 2 being the year of labeling. More blue indicates higher feature importance.

E. Mapping of the Study Area

Fig. 13 provides a map example for the year 2020. The OSO land cover map [36]) is used to select deciduous trees (from an operational point of view, since no accurate oak mask was available, it was decided to use a deciduous tree to avoid masking too much area). It can be 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).

V. DISCUSSION

A. Visual Map Comparison and Interest of Data Expansion

Fig. 14 provides a visual comparison of mapping obtained using raw and expanded references. Without data expansion and balancing, it can be observed that some areas in the produced maps can vary significantly between years (especially those circled in pink). On the other hand, proper balancing of the training data can mitigate these variations and lead to much more stable maps. We observed that using only the raw reference data, mapping of the years 2019 and 2021 were very pessimistic when compared to 2020, i.e., as seen in Fig. 14(c) but extended to the entire maps. 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).

To the best of our knowledge, the influence of the budding period (April/May) explains most of the differences in the temporal variability of the model predictions obtained with and without data expansion. To highlight this point, Fig. 15 shows the feature importance (in natural scale) obtained using the raw references [see Fig. 15(a)] and expanded and balanced training data [see Fig. 15(b)]. One can clearly see the importance of the April/May period before labeling (Y2) when using the raw references. On the other hand, the model trained with balanced expanded data relies less on this period and focuses more on dates between June and September.

To illustrate this point, Fig. 16 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 with many declining references) or healthy (e.g., 2020 with a majority of healthy plots). However, we also found that adding dates other than summer (June/September) improve the classification results, which means that there is interesting information outside the summer period that can be used to detect forest dieback, but that this information can be misleading if not used correctly. Finally, note that the sharp decrease affecting the CRswir time series in February 2018 is caused by undetected clouds (or snow) affecting most of the region, illustrating that the input data may be subject to noise or disturbances.

B. 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 [32], [71], and [72], while Sapes et al. [73] found that adding SWIR information increased the accuracy for oak wilt detection. Since drought-induced dieback is associated with increased plant water stress [10], the importance of SWIR was also expected in our case [74].

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. [75] for *Phytophthora*-induced symptoms in oak decline). In the case of bark beetle attacks, Zabihi et al. [76] highlighted 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) [77], which combine bands 8 A (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 (see Appendix A-B), which is consistent with the fact that oak dieback is influenced by previous consecutive years of drought [78]. Nevertheless, the fact that acceptable results can be obtained with only 1 year of data is interesting. Furthermore, no significant gain was observed
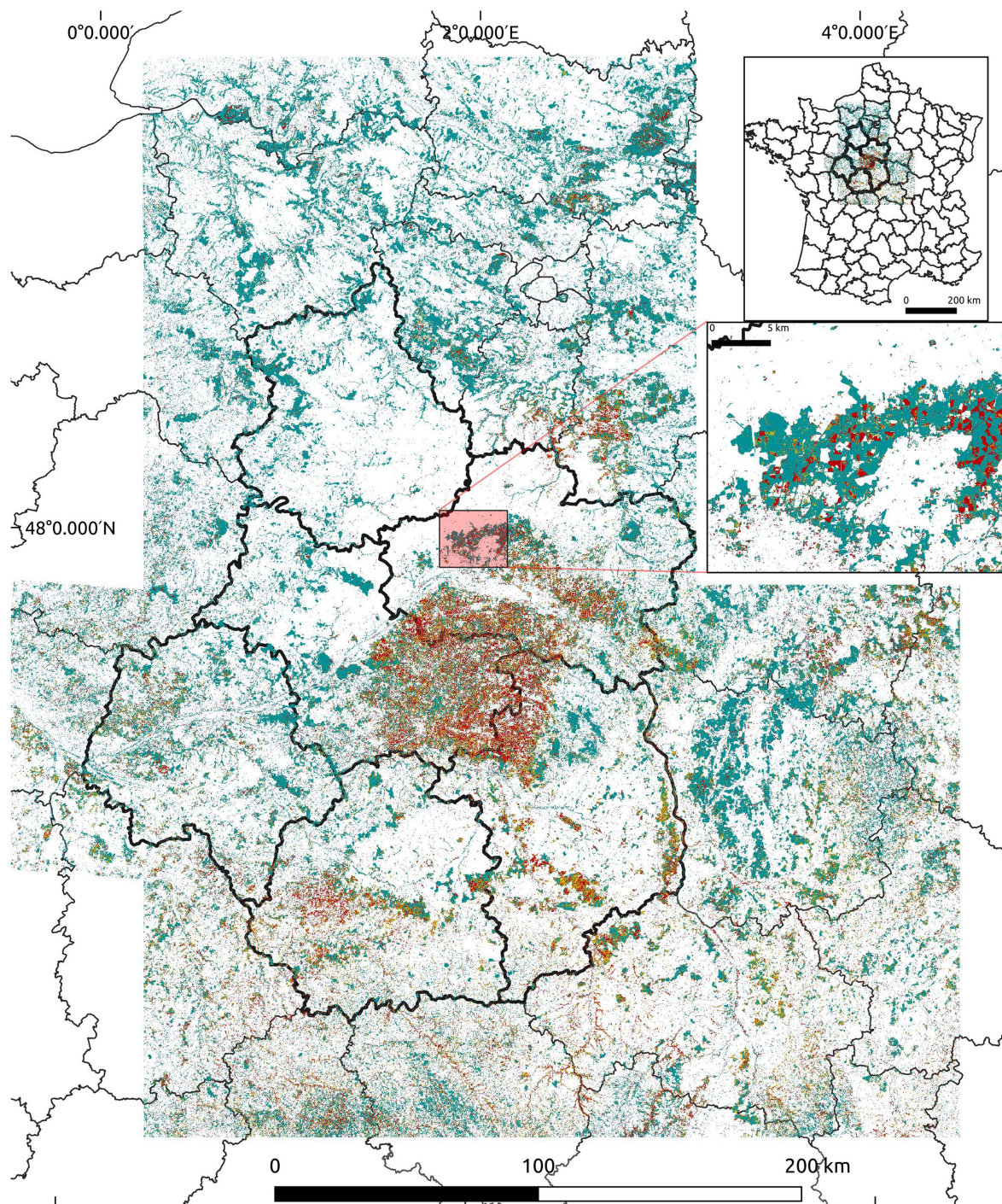
Fig. 13. Final map production for the year 2020. Healthy, declining and very declining pixels are in cyan, orange, and red, respectively. The deciduous trees OSO land cover map is used. A dezoomed version of the map is shown at the top right. Below is a zoomed version of the pink area, in Orléans forest.

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.

C. Classification and Preprocessing 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 (note that during our preliminary analysis, the SVM algorithm performed similarly than RF, but its scalability to large datasets limits its usefulness for operational services). The fact that all models tested so far converged to similar results is encouraging (see Appendix A-B). Since interesting results have been obtained using standard deep learning approaches, further research on that point will be an interesting perspective. Indeed, an advantage of using deep learning approach could also be the potential
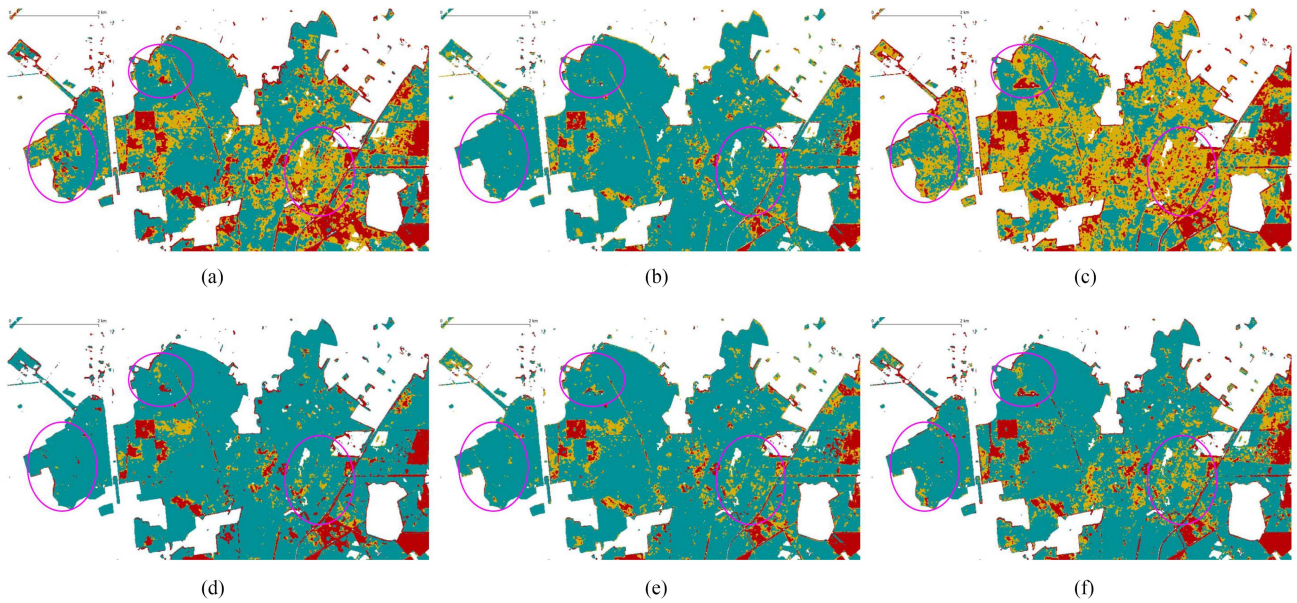
Fig. 14.   Maps produces in the Orléans forest with raw reference data (a), (b), (c) and with data expansion and yearly balancing (d), (e), (f) using the RF algorithm and CRswir and CRre indices (acquired over 2 years). Interesting areas are circled in pink. Healthy, declining, and very declining pixels are in cyan, orange, and red, respectively. (a) Raw references—2019. (b) Raw references—2020. (c) Raw references—2021. (d) Expanded references—2019. (e) Expanded references—2020. (f) Expanded references—2021.
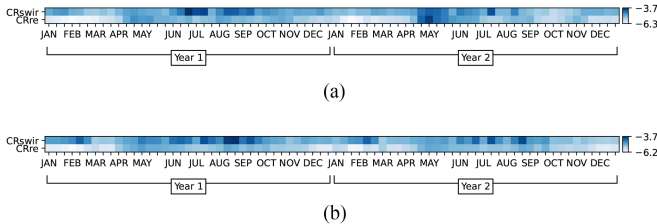

Fig. 15.   Feature importance (in log scale) obtained when training the model using CRswir and CRre indices using (a) the raw reference data and (b) augmented references and balancing. The features are ordered in temporal order, Year 1 is the year before labeling, and Year 2 is the year of labeling. More blue indicates higher feature importance.
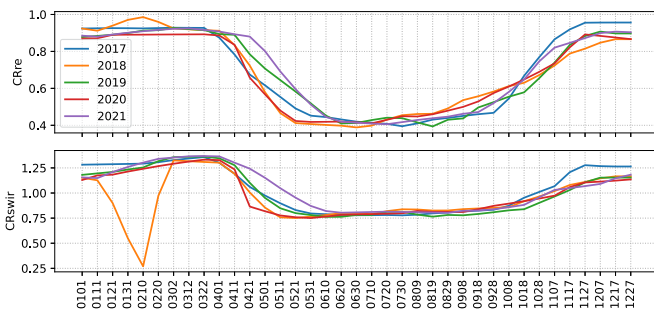

Fig. 16.   For the different labeling years, median of the CRre and CRswir for the healthy plots in the Fontainebleau forest. X axis labels are days of years in month/days (MMDD).

application to other types of trees via transfer learning [79], e.g., *Pinus sylvestris* is another key essence of the studied area. Other state-of-the-art models could be tested, such as sparse Gaussian processes [80] or deep learning models based on

transformers such as lightweight temporal self-attention [81]. The study of unsupervised methods such as FORDEAD [48] 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).

As discussed in the previous section, our results highlight the interest in having data expansion to increase the accuracy and stability of the mapping over time. While our data expansion procedure has proven to be efficient, it is based on simple rules that may not be correct for some specific plots. Moreover, our rules do not cover the case where a plot healthy year Y is still healthy year $Y + 1$ and, conversely, a plot declining year Y was already declining year $Y - 1$. Improvement may be possible by using other techniques, such as label spreading [82] or semisupervised approaches [83], to automatically add new unlabeled examples. However, our preliminary tests are inconclusive; a major drawback of such an approach is the potential addition of noisy or misleading new examples. This highlights the interest of the relatively simple but robust data expansion procedure presented in our study, which uses field knowledge to confidently select additional training samples.

Regarding the oversampling technique, it was observed that using the standard SMOTE algorithm lead to the best results. Using variants (e.g., Borderline SMOTE [84] or ADASYN [59]) lead to a deterioration of the results. One explanation for this deterioration could be related to potentially blurred class boundaries: While SMOTE oversamples 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, other methods than gapfilling have been tested, e.g., multiple imputation by chained equations (MICE) [85] or *k*-nearest neighbors (KNN) [86], 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 possible improvements had been reported in various applications related to remote sensing [45], [87], [88].

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 II-B1). 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 versus 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. An interesting perspective is then to work on noisy labels to mitigate the fact that some pixels of a plot do not necessarily share the label of the plot, e.g., healthy pixels in a declining plot.

VI. CONCLUSION

This article investigates the applicability of satellite imagery for monitoring oak dieback through a case study conducted in the Centre-Val region of France. Our findings are relevant to a range of stakeholders, including forest management authorities, environmental policy makers, and forest health researchers. In particular, our results provide insights into the feasibility of using remote sensing techniques to detect oak dieback, thereby facilitating proactive interventions with a timely mapping of the forest health status. Our work could also facilitate a deeper exploration of the links between environmental factors and forest dieback, for example, by combining the high-resolution maps produced with geolocated environmental data.

Our analysis shows that the use of multispectral data (here Sentinel-2 images) is adapted to accurately detect forest dieback. In the 3-class scenario (healthy, declining, very declining), the separation of declining and very declining classes remains challenging. The proposed approach involves training of a random forest algorithm on S2 time series slices acquired over two years prior to labeling. We propose to use time series from two vegetation indices based on 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. Furthermore, we have shown that it is important to perform stability analysis, especially in areas that are not always covered by reference data every year. Our results also show that two additional steps can improve the performance and temporal stability of the prediction model across the different

TABLE IV
OA AND BA AVERAGED OVER THE DIFFERENT LABELING YEARS FOR A
CLASSIFICATION WITH THREE CLASSES (3 CL.) AND TWO CLASSES (2 CL.)
USING THE RF ALGORITHM

Fram.	Feat.	N	OA 3 cl.	BA 3 cl.	OA 2 cl.	BA 2 cl.
Raw	CRswir, CRre	2	0.64	0.59	**0.79**	0.75
Raw	S2 bands	2	0.62	0.58	0.77	0.72
Raw	CRswir	2	0.62	0.57	0.78	0.74
Exp.	CRswir, CRre	2	**0.65**	**0.61**	**0.79**	**0.78**
Exp.	CRswir, CRre	1	0.64	0.60	0.78	0.77
Exp.	S2 bands	2	**0.65**	**0.61**	0.78	0.77
Exp.	CRswir	2	**0.65**	**0.61**	**0.79**	**0.78**

Frameworks (Fram.) using raw or expanded (Exp.) references are compared with various features (Feat.) and years of data in input ($N = N_{\text{year}}$). All confidence intervals are lower than 0.010.
Best results in bold.

TABLE V
F1 SCORES AVERAGED OVER THE DIFFERENT LABELING YEARS USING THE
RF ALGORITHM FOR THE 3 CLASSES, HEALTHY (HEALTH.), DECLINING
(DECL.) AND VERY DECLINING (V. DECL.)

Fram.	Feat.	N	F1 - h.	F1 - d.	F1 - v.d.
Raw	CRswir, CRre	2	0.70	0.50	0.56
Raw	S2 bands	2	0.66	0.49	**0.58**
Raw	Crswir	2	0.68	0.49	0.55
Exp.	CRswir, CRre	2	**0.74**	0.50	0.57
Exp.	CRswir, CRre	1	0.73	0.47	**0.58**
Exp.	S2 bands	2	**0.74**	0.46	**0.58**
Exp.	CRswir	2	**0.74**	**0.51**	**0.58**

Frameworks using raw and expanded references are compared with different feature sets. All confidence intervals are lower than 0.015.
The bold values are the best results.

years. First, data expansion can be performed using time slices of unlabeled time series, exploiting the fact that healthy pixels were likely healthy in the years prior to labeling, and declining pixels are likely to continue declining in the coming years. Second, the minority class(es) can be oversampled to balance the final dataset. This highlights the importance of collecting samples over multiple years to capture the diversity in phenology.

These results open up interesting perspectives, some of which are discussed in Section V. First of all, specific work on the classification model could improve dieback detection, especially using deep learning models. Since the reference expansion and balancing steps were found to be important for better generalization, further work on this part could increase model stability and performance. Unsupervised detection is another interesting perspective, especially since such approaches do not rely on labeled data. Given the relatively subtle response of oak trees to drought-induced decline, 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 various remote sensing applications [89], [90].

APPENDIX
ADDITIONAL RESULTS

A. OA, BA, and F1 Scores

To complement the results provided in the main document, Tables IV and V provides the OA, BA, and F1 scores for each

TABLE VI
OA AND BA AVERAGED OVER THE DIFFERENT LABELING YEARS FOR A
CLASSIFICATION WITH THREE CLASSES (3 CL.) AND TWO CLASSES (2 CL.)

Classifier	OA 3 cl.	BA 3 cl.	OA 2 cl.	BA 2 cl.
RF	0.653	**0.610**	**0.792**	**0.779**
XGBoost	0.623	0.597	0.777	0.760
FCN	**0.660**	**0.610**	0.783	0.757

Features are the CRswir acquired over 2 years before labeling. All confidence
intervals are lower than 0.011 (0.010 for RF).
The bold values are the best results.
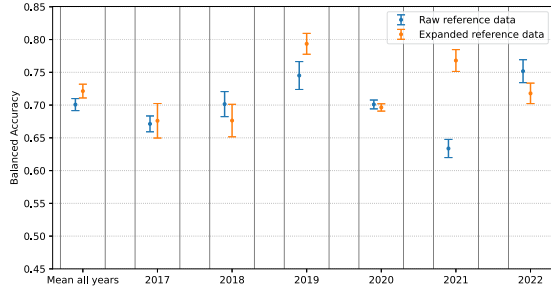
Fig. 17. Balanced accuracy obtained when using two classes of forest dieback with and without expanded references. The features used as input are the CRswir and CRre acquired over two years ($N_{\text{years}} = 2$).

class and for different sets of features. We can see that using data expansion 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. The lower F1 score for the declining classes indicates a confusion between these two classes and not between healthy and declining classes.

### B. Results Obtained With Other Classifiers

Table VI 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 [63] and a fully convolutional network (FCN) [64], which are adapted to time series. A minimal tuning was conducted by grid-search, i.e., for XG-Boost 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 [64], i.e., a multi-channel convolutional network of size (128, 256, 128) with learning rate 0.001 and 30 epochs [64, Fig. 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.

C. Forecasting Study

To analyze the potential interest of the proposed model for forecasting (i.e., predicting dieback in future years) and to test its robustness to lack of data, an experiment was conducted by testing the model on a given year after training on the other years. Subset of reference data (for a given year) were selected by repeating ten times a stratified 5-fold CV. Average BA (2 classes) are reported in Fig. 17 with and without expanded references.
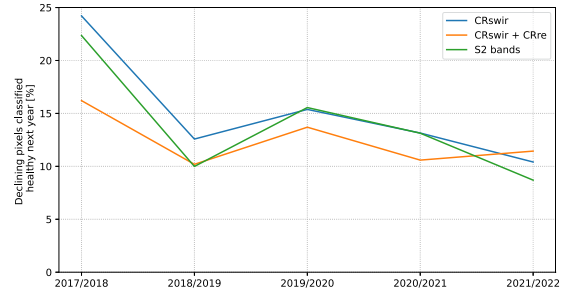
(a)
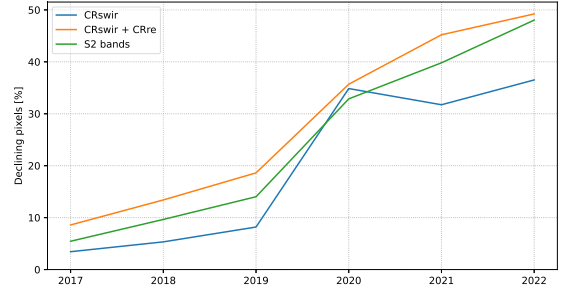
(b)

Fig. 18. Percentage of (a) declining pixels classified healthy next year and (b) pixels classified as declining (average after fivefold CV repeated ten 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.

One can see a decrease in the accuracy, i.e., average BA over years is equal to 0.721 instead of 0.792 (see Table IV). This decrease was expected because the model is not calibrated to the year to be predicted and because a large number of samples are removed from the training dataset (this is especially true for 2020, where the number of training samples is divided by two).

D. Temporal Stability Using Other Features

Fig. 18 is similar to Fig. 11 but focuses only on the framework with expanded references using either CRswir, CRswir, and CRre or all S2 bands as features. Fig. 18(a) shows that using CRswir and CRre together leads to higher overall temporal stability. Moreover, Fig. 18(b) also shows that 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.

ACKNOWLEDGMENT

Centre National de la Propriété Forestiére (CNPF), Unisylva and INRAE EFNO. Finally, they thank J.-B. Féret for its help regarding the CRswir indice, T. Belouard, A. Jolly, and M. Chartier for their scientific advices.

REFERENCES

[1] V. H. Dale et al., "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*, vol. 51, no. 9, pp. 723–734, 2001, doi: 10.1641/0006-3568(2001)051[0723:CCAFD]2.0.CO;2.

[2] R. Seidl et al., "Forest disturbances under climate change," *Nature Climate Change*, vol. 7, pp. 395–402, 2017.

[3] M. G. Turner, "Disturbance and landscape dynamics in a changing world," *Ecology*, vol. 91, no. 10, pp. 2833–2849, 2010, doi: 10.1890/10-0097.1.

[4] M. Zhang and X. Wei, "Deforestation, forestation, and water supply," *Science*, vol. 371, no. 6533, pp. 990–991, 2021, doi: 10.1126/science.abe7821.

[5] FAO, "Global forest resources assessment 2020: Main report," Food Agriculture Org. United Nations, Rome, Italy, 2020. [Online]. Available: http://www.fao.org/3/ca9825en/ca9825en.pdf

[6] D. J. Krieger, "Economic value of forest ecosystem services: A review," The Wilderness Soc., Washington, DC, USA, Tech. Rep., 2001.

[7] T. P. Holmes, R. J. Huggett, and J. M. Pye, "Forest economics, natural disturbances and the new ecology," in *The Economics of Forest Disturbances*, vol. 9, Berlin, Germany: Springer, 2008, pp. 15–32, doi: 10.1007/978-1-4020-4370-3_2.

[8] French Ministry for the Ecological and Solidary Transition, "National low carbon strategy," French Ministry for the Ecological and Solidary Transition, Paris, France, Tech. Rep., 2020. [Online]. Available: https://www.ecologie.gouv.fr/sites/default/files/en_SNBC-2_complete.pdf

[9] C. I. Millar and N. L. Stephenson, "Temperate forest health in an era of emerging megadisturbance," *Science*, vol. 349, no. 6250, pp. 823–826, 2015, doi: 10.1126/science.aaa9933.

[10] C. D. Allen, D. D. Breshears, and N. G. McDowell, "On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the anthropocene," *Ecosphere*, vol. 6, no. 8, 2015, Art. no. 129, doi: 10.1890/ES15-00203.1.

[11] P. D. Manion, *Tree Disease Concepts*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1981.

[12] G. Sangüesa-Barreda, J. C. Linares, and J. J. Camarero, "Reduced growth sensitivity to climate in bark-beetle infested aleppo pines: Connecting climatic and biotic drivers of forest dieback," *Forest Ecol. Manage.*, vol. 357, pp. 126–137, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378112715004429

[13] V. Moravec et al., "Europe under multi-year droughts: How severe was the 2014–2018 drought period?," *Environ. Res. Lett.*, vol. 16, no. 3, Mar. 2021, Art. no. 034062, doi: 10.1088/1748-9326/abe828.

[14] V. Blauhut et al., "Lessons from the 2018–2019 European droughts: A collective need for unifying drought risk management," *Natural Hazards Earth System Sci.*, vol. 22, no. 6, pp. 2201–2217, 2022. [Online]. Available: https://nhess.copernicus.org/articles/22/2201/2022/

[15] M. C. Hansen et al., "High-resolution global maps of 21st-century forest cover change," *Science*, vol. 342, no. 6160, pp. 850–853, 2013, doi: 10.1126/science.1244693.

[16] A. R. Huete, "Vegetation indices, remote sensing and forest monitoring," *Geography Compass*, vol. 6, no. 9, pp. 513–532, 2012, doi: 10.1111/j.1749-8198.2012.00507.x.

[17] P. Torres, M. Rodes-Blanco, A. Viana-Soto, H. Nieto, and M. García, "The role of remote sensing for the assessment and monitoring of forest health: A systematic evidence synthesis," *Forests*, vol. 12, no. 8, Aug. 2021, Art. no. 1134, doi: 10.3390/f12081134.

[18] A. T. Stahl, R. Andrus, J. A. Hicke, A. T. Hudak, B. C. Bright, and A. J. Meddens, "Automated attribution of forest disturbance types from remote sensing data: A synthesis," *Remote Sens. Environ.*, vol. 285, 2023, Art. no. 113416. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425723005223

[19] M. A. Wulder et al., "Fifty years of Landsat science and impacts," *Remote Sens. Environ.*, vol. 280, 2022, Art. no. 113195. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425722003054

[20] M. Drusch et al., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425712000636

[21] M. Mohajane, A. Essahlaoui, F. Oudija, M. E. Hafyani, and A. C. Teodoro, "Mapping forest species in the central middle atlas of Morocco (Azrou Forest) through remote sensing techniques," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 9, Sep. 2017, Art. no. 275, doi: 10.3390/ijgi6090275.

[22] H. Abdullah, A. K. Skidmore, R. Darvishzadeh, and M. Heurich, "Sentinel-2 accurately maps green-attack stage of European spruce bark beetle (Ips typographus, L.) compared with Landsat-8," *Remote Sens. Ecol. Conserv.*, vol. 5, no. 1, pp. 87–106, 2019. [Online]. Available: https://zslpublications.onlinelibrary.wiley.com/doi/abs/10.1002/rse2.93

[23] K. A. Barta, M. Hais, and M. Heurich, "Characterizing forest disturbance and recovery with thermal trajectories derived from Landsat time series data," *Remote Sens. Environ.*, vol. 282, 2022, Art. no. 113274. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425722003807

[24] Y. Cai, Q. Shi, X. Xu, and X. Liu, "A novel approach towards continuous monitoring of forest change dynamics in fragmented landscapes using time series Landsat imagery," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 118, 2023, Art. no. 103226. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1569843223000481

[25] I. Housman, R. Chastain, and M. Finco, "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.*, vol. 10, no. 8, Jul. 2018, Art. no. 1184, doi: 10.3390/rs10081184.

[26] A. L. Mitchell, A. Rosenqvist, and B. Mora, "Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD," *Carbon Balance Manage.*, vol. 12, 2017, Art. no. 9, doi: 10.1186/s13021-017-0078-9.

[27] C. Mulverhill, N. C. Coops, and A. Achim, "Continuous monitoring and sub-annual change detection in high-latitude forests using harmonized Landsat Sentinel-2 data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 197, pp. 309–319, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271623000424

[28] Y. Yang, M. C. Anderson, F. Gao, J. D. Wood, L. Gu, and C. Hain, "Studying drought-induced forest mortality using high spatiotemporal resolution evapotranspiration data from thermal satellite imaging," *Remote Sens. Environ.*, vol. 265, 2021, Art. no. 112640. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425721003606

[29] C. C. Aggarwal, *Outlier Analysis*, 2nd ed. Cham Switzerland: Springer, 2017.

[30] J. Verbesselt, R. Hyndman, G. Newnham, and D. Culvenor, "Detecting trend and seasonal changes in satellite image time series," *Remote Sens. Environ.*, vol. 114, no. 1, pp. 106–115, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003442570900265X

[31] R. E. Kennedy, Z. Yang, and W. B. Cohen, "Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—temporal segmentation algorithms," *Remote Sens. Environ.*, vol. 114, no. 12, pp. 2897–2910, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425710002245

[32] R. Dutrieux, J.-B. Feret, K. Ose, and F. D. Boissieu, "Package fordead," Recherche Data Gouv, Software, 2021, doi: 10.15454/4TEO6H.

[33] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[34] L. Huang, X. Zhao, Y. Liu, and P. Yang, "Analysis of the atmospheric duct existence factors in tropical cyclones based on the SHAP interpretation of extreme gradient boosting predictions," *Remote Sens.*, vol. 14, no. 16, Aug. 2022, Art. no. 3952, doi: 10.3390/rs14163952.

[35] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.

[36] V. Thierion, A. Vincent, and S. Valero, "Theia OSO land cover map 2021," CESBIO, Dataset, Mar. 2022.

[37] J. Cordier, R. Dupré, S. Bellenfant, and S. Gautier, "Atlas de la flore du Centre-Val de Loire," (Series Inventaires & Biodiversité), 19. Muséum National D'histoire Naturelle, Paris, France, 2021.

[38] M. Goudet, F. Saintonge, and L. Nageleisen, "Quantifier l'état de santé de la forêt, méthode simplifiée d'évaluation," Département de la santé des forêts, Tech. Rep. 90879, 2018. [Online]. Available: https://agriculture.gouv.fr/telecharger/90879

[39] DGAL, "Surveillance des dépérissements en forêts," Direction Générale de l'Alimentation (DGAL), Note de service, 2018. [Online]. Available: https://info.agriculture.gouv.fr/gedei/site/bo-agri/instruction-2018-433/telechargement

[40] F. Saintonge and M. Goudet, "Une enquête pour évaluer la santé de 85 massifs de chêne," Département de la santé des forêts, Tech. Rep. 115173, 2020. [Online]. Available: https://agriculture.gouv.fr/telecharger/115173

[41] O. Hagolle, M. Huc, D. V. Pascual, and G. Dedieu, "A multi-temporal and multi-spectral method to estimate aerosol optical thickness over land, for the atmospheric correction of Formosat-2, Landsat, VEN μ S and Sentinel-2 images," *Remote Sens.*, vol. 7, no. 3, pp. 2668–2691, 2015. [Online]. Available: https://www.mdpi.com/2072-4292/7/3/2668

[42] J. Inglada, A. Vincent, M. Arias, and B. Tardy, "iota2-a25386," CESBIO, Software, Jul. 2016.

[43] J. Inglada et al., "Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery," *Remote Sens.*, vol. 7, no. 9, pp. 12356–12379, Sep. 2015, doi: 10.3390/rs70912356.

[44] M. Fauvel et al., "Prediction of plant diversity in grasslands using Sentinel-1 and -2 satellite image time series," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111536. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425719305553

[45] F. Vuolo, W.-T. Ng, and C. Atzberger, "Smoothing and gap-filling of high resolution multi-spectral time series: Example of Landsat data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 57, pp. 202–213, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0303243416302100

[46] A. Bannari, D. Morin, F. Bonn, and A. R. Huete, "A review of vegetation indices," *Remote Sens. Rev.*, vol. 13, no. 1/2, pp. 95–120, 1995.

[47] J. Rouse, R. Haas, J. Schell, and D. Deering, *Monitoring Vegetation Systems in the Great Plains With ERTS*, vol. 351, Washington, DC, USA: NASA Special Pub., 1974, p. 309.

[48] R. Dutrieux, J.-B. Féret, and K. Ose, "Mise au point d'une méthode reproductible pour le suivi généralisé des dégats de scolytes par télédétection satellitaire," *ONF Rendez-Vous Techn.*, vol. 69–70, pp. 37–44, 2021. [Online]. Available: https://www.onf.fr/onf//cec::les-rendez-vous-techniques-de-lonf-no69-70.html

[49] R. N. Clark and T. L. Roush, "Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications," *J. Geophysical Res. Solid Earth*, vol. 89, no. B7, pp. 6329–6340, 1984, doi: 10.1029/JB089iB07p06329.

[50] Z. Huang, B. J. Turner, S. J. Dury, I. R. Wallis, and W. J. Foley, "Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis," *Remote Sens. Environ.*, vol. 93, no. 1, pp. 18–29, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425704001920

[51] E. Grabska, P. Hawryło, and J. Socha, "Continuous detection of small-scale changes in scots pine dominated stands using dense Sentinel-2 time series," *Remote Sens.*, vol. 12, no. 8, Apr. 2020, Art. no. 1298, doi: 10.3390/rs12081298.

[52] J. Olsen et al., "Relation between seasonally detrended shortwave infrared reflectance data and land surface moisture in semi-arid sahel," *Remote Sens.*, vol. 5, no. 6, pp. 2898–2927, Jun. 2013, doi: 10.3390/rs5062898.

[53] J. Bramich, C. J. Bolch, and A. Fischer, "Improved red-edge chlorophyll-a detection for Sentinel 2," *Ecological Indicators*, vol. 120, 2021, Art. no. 106876. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1470160X20308141

[54] J. Delegido, J. Verrelst, L. Alonso, and J. Moreno, "Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content," *Sensors*, vol. 11, no. 7, pp. 7063–7081, Jul. 2011, doi: 10.3390/s110707063.

[55] P. Zarco-Tejada, A. Hornero, R. Hernández-Clemente, and P. Beck, "Understanding the temporal dimension of the red-edge spectral region for forest decline detection using high-resolution hyperspectral and Sentinel-2A imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 137, pp. 134–148, Mar. 2018, doi: 10.1016/j.isprsjprs.2018.01.017.

[56] P. Zarco-Tejada, J. Pushnik, S. Dobrowski, and S. Ustin, "Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects," *Remote Sens. Environ.*, vol. 84, no. 2, pp. 283–294, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003442570200113X

[57] A. Jain and B. Chandrasekaran, "Dimensionality and sample size considerations in pattern recognition practice," in *Classification Pattern Recognition and Reduction of Dimensionality*, (Series Handbook of Statistics), vol. 2. Amsterdam, The Netherlands: Elsevier, 1982, pp. 835–855.

[58] A. K. Bose et al., "Climate sensitivity and drought seasonality determine post-drought growth recovery of quercus petraea and quercus robur in Europe," *Sci. Total Environ.*, vol. 784, 2021, Art. no. 147222.

[59] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Conf. Neural Netw.*, 2008, pp. 1322–1328.

[60] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/tkde.2008.239.

[61] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning," *J. Mach. Learn. Res.*, vol. 18, no. 17, pp. 1–5, 2017. [Online]. Available: http://jmlr.org/papers/v18/16-365.html

[62] C. Pelletier, S. Valero, J. Inglada, N. Champion, and G. Dedieu, "Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas," *Remote Sens. Environ.*, vol. 187, pp. 156–168, 2016.

[63] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[64] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *Proc. Int. Joint Conf. Neural Netw.*, 2017, pp. 1578–1585, doi: 10.1109/ijcnn.2017.7966039.

[65] C. Igel, V. Heidrich-Meisner, and T. Glasmachers, "Shark," *J. Mach. Learn. Res.*, vol. 9, pp. 993–996, 2008.

[66] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[67] N. Karasiak, J.-F. Dejoux, C. Monteil, and D. Sheeren, "Spatial dependence between training and test sets: Another pitfall of classification accuracy assessment in remote sensing," *Mach. Learn.*, vol. 111, no. 7, pp. 2715–2740, Apr. 2021.

[68] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 3121–3124.

[69] A. M.-C. Wadoux, G. B. Heuvelink, S. d. Bruin, and D. J. Brus, "Spatial cross-validation is not the right way to evaluate map accuracy," *Ecol. Model.*, vol. 457, 2021, Art. no. 109692.

[70] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, pp. 56–67, 2020.

[71] V. Bárta, P. Lukeš, and L. Homolová, "Early detection of bark beetle infestation in Norway spruce forests of central Europe using Sentinel-2," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 100, 2021, Art. no. 102335. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0303243421000428

[72] L. Huo, H. J. Persson, and E. Lindberg, "Early detection of forest stress from European spruce bark beetle attack, and a new vegetation index: Normalized distance red & SWIR (NDRS)," *Remote Sens. Environ.*, vol. 255, 2021, Art. no. 112240. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425720306131

[73] G. Sapes et al., "Canopy spectral reflectance detects oak wilt at the landscape scale using phylogenetic discrimination," *Remote Sens. Environ.*, vol. 273, 2022, Art. no. 112961. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003442572200075X

[74] C. y. Huang, W. R. Anderegg, and G. P. Asner, "Remote sensing of forest die-off in the anthropocene: From plant ecophysiology to canopy structure," *Remote Sens. Environ.*, vol. 231, 2019, Art. no. 111233. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425719302524

[75] A. Hornero et al., "Modelling hyperspectral- and thermal-based plant traits for the early detection of phytophthora-induced symptoms in oak decline," *Remote Sens. Environ.*, vol. 263, 2021, Art. no. 112570. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003442572100290X

[76] K. Zabihi, P. Surovy, A. Trubin, V. V. Singh, and R. Jakuš, "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.*, vol. 24, 2021, Art. no. 100638. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352938521001749

[77] B. Gao, "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sens. Environ.*, vol. 58, no. 3, pp. 257–266, 1996.

[78] J. Rodréguez-Calcerrada, D. Sancho-Knapik, N. K. Martin-StPaul, J.-M. Limousin, N. G. McDowell, and E. Gil-Pelegrén, "Drought-induced oak declinefactors involved, physiological dysfunctions, and potential attenuation by forestry practices," in *Oaks Physiological Ecology. Exploring the Functional Diversity of Genus Quercus L.* Berlin, Germany: Springer, 2017, pp. 419–451, doi: 10.1007/978-3-319-69099-5_13.

[79] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, 2016, Art. no. 9, doi: 10.1186/s40537-016-0043-6.

[80] V. Bellet, M. Fauvel, and J. Inglada, "Land cover classification with Gaussian processes using spatio-spectro-temporal features," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–21, 2023.

[81] V. S. F. Garnot and L. Landrieu, "Lightweight temporal self-attention for classifying satellite images time series," in *Proc. Workshop Adv. Analytics Learn. Temporal Data*, V. Lemaire, S. Malinowski, A. Bagnall, T. Guyet, R. Tavenard, and G. Ifrim, Eds. Ghent, Belgium: Springer, 2020, pp. 171–181.

[82] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Advances in Neural Information Processing Systems*, S. Thrun, L. Saul, and B. Schölkopf, Eds., vol. 16. Cambridge, MA, USA: MIT Press, 2003. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2003/file/87682805257e619d49b8e0dfdc14affa-Paper.pdf

[83] J. E. v. Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Mach. Learn.*, vol. 109, no. 2, pp. 373–440, Nov. 2019, doi: 10.1007/s10994-019-05855-6.

[84] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning," in *Advances in Intelligent Computing*, D.-S. Huang, X.-P. Zhang, and G.-B. Huang Eds., Berlin, Heidelberg: Springer, 2005, pp. 878–887.

[85] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," *J. Statist. Softw.*, vol. 45, no. 3, pp. 1–67, 2011. [Online]. Available: https://www.jstatsoft.org/v045/i03

[86] O. Troyanskaya et al., "Missing value estimation methods for DNA microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.

[87] C. K. Turlej, M. Ozdogan, and V. C. Radeloff, "Mapping forest types over large areas with Landsat imagery partially affected by clouds and SLC gaps," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 107, 2022, Art. no. 102689. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0303243422000150

[88] F. Mouret, M. Albughdadi, S. Duthoit, D. Kouamé, G. Rieu, and J.-Y. Tourneret, "Reconstruction of Sentinel-2 derived time series using robust gaussian mixture models—Application to the detection of anomalous crop development," *Comput. Electron. Agriculture*, vol. 198, 2022, Art. no. 106983.

[89] J. Inglada, A. Vincent, M. Arias, and C. Marais-Sicre, "Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series," *Remote Sens.*, vol. 8, no. 5, Apr. 2016, Art. no. 362, doi: 10.3390/rs8050362.

[90] D. Morin et al., "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.*, vol. 11, no. 11, May 2019, Art. no. 1275, doi: 10.3390/rs11111275.

David Morin received the M.Sc. degree in remote sensing and environmental mapping from the University of Toulouse, Toulouse, France, in 2012. He received the Ph.D. degree in environmental sciences from the University of Toulouse (SDU2E Doctoral School), Toulouse, France, in 2020.

From 2013 to 2016, he worked as an Engineer with the CESBIO Laboratory, Toulouse, France. From 2020 to 2023, he was a Postdoctoral Fellow with CESBIO, Toulouse, France. He is currently a Postdoctoral Researcher with the University of Orléans, Orléans, France. His research interests include radar and optical signal processing, machine learning and statistical analysis with remote sensing data for land use classification and vegetation monitoring.

Hilaire Martin received the M.Sc. degree in earth science from the École Pratique des Hautes Etudes section science et vie de la terre from Paris, France, in 2009.

He is currently an Assistant Engineer with INRAE EFNO, Nogent-sur-Vernisson, France, working with GIS and remote sensing. He also carries out floristic and lichenologic surveys and soil descriptions. He is also a tree climber.

Milena Planells received the double master degree in electronical telecommunication and signal processing from Télécom Paris, Palaiseau, France and Télécom BCN, Barcelona, Spain, in 2000, and the Ph.D. degree in electronics and telecommunication from Télécom Paris, Palaiseau, France, in 2003.

She joined the French Space Agency in 2004 and has since then been working on research and development of different space missions related to Communications and Earth Observation. She is currently a Remote Sensing Engineer with the CESBIO Laboratory (UMR 5126). Her research interest include crops and forest parameters retrieval and monitoring, using multimodal satellite data (lidar, optical, and radar).

Florian Mouret received the ingénieur degree in electrical engineering and signal processing from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications de Toulouse, Toulouse, France, in 2016 and the Ph.D. degree in computer science from the National Polytechnic Institute of Toulouse, Toulouse, in 2022.

His Ph.D. thesis was carried out at the IRIT Laboratory In collaboration with the company TerraNIS in Toulouse, France. He is currently a Postdoctoral Researcher with the University of Orléans in collaboration with the CESBIO Laboratory. His research interests include outlier detection, missing data reconstruction and machine learning applied to remote sensing for vegetation monitoring (crops, forests).

Cécile Vincent-Barbaroux received the Ph.D. degree in biology and ecology from the University of Paris XI, Orsay, France, in 2002.

From 2003 to 2004, she was a Postdoctoral Fellow with the Instituto Superior de Agronomia from the Lisbon University, Portugal. In 2004, she joined the University of Orleans, France, as an Associate Professor in ecology and ecophysiology. Since then, she has been working on the response of trees to environmental constraints. Her research work has gradually shifted toward remote sensing as a tool for the analysis of forest dieback using optical satellite data. She is currently coordinating the Reconfort Project on the assessment of forest health in the Centre Val de Loire region.