https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Comparing geometric, spectral and temporal information for forest disturbance classification in the southeastern United States

Franziska Müller[1,2], Gustau Camps-Valls[3], and Ana Bastos[1]

[1]Institute for Earth System Science and Remote Sensing, University of Leipzig, Talstr. 35, 04103 Leipzig, Germany
[2]Max Planck Institute for Biogeochemistry, Dept. of Biogeochemical Integration, Hans-Knöll-Straße 10, 07745 Jena, Germany
[3]Image Processing Laboratory (IPL), Universitat de València, C/ Cat. Agustín Escardino Benlloch 9, 46980 Paterna (València), Spain

Correspondence: Franziska Müller (franziska.mueller@uni-leipzig.de)

Abstract. Changing frequency and severity of forest disturbances under climate change can threaten forest stability and resilience. Since disturbances vary in their responses to climate change and in impacts, accurately monitoring individual forest disturbance types is thus crucial to anticipate climate-related risks and support forest management. Satellite imagery allows for large-scale disturbance mapping, but attributing disturbances to specific agents remains challenging. In particular, wind, bark beetles, and defoliating insects often produce similar spatial and spectral signals, complicating their differentiation using remote sensing.

In this study, we address this challenge by investigating the skill of various classification algorithms and remotely-sensed information in separating between wind, bark beetles and defoliator disturbances, for a region in southeastern United States. Specifically, we evaluate the skill of various machine learning models under nested spatial cross-validation combined with different predictor sets derived from Sentinel-1 and Sentinel-2 in classifying these three disturbances. Predictors include commonly used metrics such as patch geometry and static spectral and radar backscatter information, but also temporally variable spectral and radar backscatter information.

Models using only geometric ($F1_{macro}$ = 0.57) or static predictors ($F1_{macro}$ = 0.65) are outperformed by models that exploit temporal signals. Our best model, MiniRocket, applied to multivariate spectral and radar time-series, reaches an $F1_{macro}$ score of 0.72. Feature ablation identifies red-edge and shortwave-infrared bands, water-related indices, and radar polarizations as key variables for distinguishing between these three disturbances. Comparing windows of one to eight years, we find that richness in spectral and radar backscatter information can compensate for shorter timeseries, with a single year of multivariate data attaining close to the skill of a five-year series on the same predictor set.

We demonstrate that temporal spectral and radiometric information are more suitable to distinguish between subtle disturbances, than spatial information alone. Furthermore, we show the added value of sensor combinations, which allow to resolve the diverse impacts of each disturbance type on vegetation especially at shorter the observation windows. Our study offers a route to improved large-scale disturbance classification, although the approach proposed requires evaluation in other regions before generalization can be established.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Keywords. Forest disturbances; Bark beetle; Defoliation; Wind damage; SAR; Machine learning; Sentinel-2; Time-series classification; Spatial cross-validation

1 Introduction

Forest disturbances shape forest dynamics by altering forest structure, species composition, and diversity (Turner, 2010; White and Pickett, 1985; Seidl et al., 2017; Graefe et al., 2020). Disturbances such as fire, insects, and wind cause widespread tree mortality (Patacca et al., 2023; Forzieri et al., 2021; US Forest Service, 2013) and substantial regional carbon losses (Harris et al., 2016; Thom and Seidl, 2016). In the western USA alone, bark beetles caused the death of about 3.8 billion trees between 1997 and 2018 (Hicke et al., 2020), and insect-driven disturbances result in larger annual carbon losses than those from fire (10 $\pm$ 1.3 Tg C vs. 7 $\pm$ 1.0 Tg C, respectively) (Harris et al., 2016). The frequency and severity of disturbances has been reported to increase in recent decades (Seidl et al., 2017; Kautz et al., 2017; IPCC, 2021). Given the heightened societal importance of forest carbon sequestration (Bustamante et al., 2023; Pan et al., 2024) in this context, large-scale and accurate monitoring of these events is critically important (Seidl et al., 2017; Kautz et al., 2017; Kennedy et al., 2014).

Forest disturbance monitoring has traditionally relied on field-based inventories and reporting systems (Kautz et al., 2017), such as national forest inventories in the US FIA (Forest Service U.S. Department of Agriculture, 2023), the Insect and Disease Survey (US Forest Service, 2013), the pan-European ICP Forests network (ICP Forests, 2025), and global assessments such as the FAO Global Forest Resources Assessment (FAO, 2020). These datasets provide valuable long-term records on disturbance occurrence and impacts. However, they are spatially limited by national monitoring efforts, temporally constrained by infrequent revisits, often contain substantial uncertainties (Coleman et al., 2018; Eifler et al., 2026), and are not always publicly accessible (Rodríguez Paulino et al., 2024). These limitations restrict our ability to study disturbance dynamics at regional to continental scales (Kautz et al., 2017).

Addressing these gaps requires complementing inventory data with remote sensing, which offers the spatial and temporal consistency and coverage that field-based records alone cannot provide. Satellite-assisted databases such as FORWIND (Forzieri et al., 2020) and DEFID2 (Forzieri et al., 2023) have extended disturbance records across Europe, although they remain limited in space and time. While remote sensing enables continuous, multi-decadal monitoring of forest disturbances at scales inaccessible to field surveys (Hansen et al., 2013; Senf and Seidl, 2021; Schleeweis et al., 2020; Viana-Soto and Senf, 2025), the specificity of disturbance attribution achievable decreases systematically with increasing spatial extent (Stahl et al., 2023), as rising ecological complexity and spectral variability often necessitate the aggregation of disturbance classes (Schleeweis et al., 2020; Viana-Soto and Senf, 2025). At the global scale, disturbance datasets are largely limited to binary detection of disturbances (Hansen et al., 2013). At the continental scale, dedicated disturbance classification datasets, including the North American Forest Dynamics product (NAFD) for the conterminous USA (Schleeweis et al., 2020) and analogous European efforts (Viana-Soto and Senf, 2025; Senf and Seidl, 2021), have progressed beyond binary detection to distinguish high-intensity agents such as fire, wind, harvest and other biotic and abiotic stressors. While harvest and fire tend to be reliably

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

identified at continental scales, often reaching classification accuracies of 80–90% (Stahl et al., 2023; Schroeder et al., 2017; Schleeweis et al., 2020), more subtle natural disturbances, such as wind damage, bark-beetle outbreaks, defoliating insects, and drought stress, remain difficult to separate (McDowell et al., 2015; Viana-Soto and Senf, 2025) or are systematically underrepresented. In NAFD, subtle agents (insects, drought, abiotic) account for less than 5% of attributed area (Schleeweis et al., 2020), with partial and low-intensity disturbances misclassified as no-change (Schleeweis et al., 2020; Schroeder et al., 2017), while in the European Forest Disturbance Atlas (EFDA) wind and bark beetle disturbances are grouped into a combined class since they are difficult to separate (Viana-Soto and Senf, 2025). Local and regional studies have attempted finer classification, targeting individual agents such as bark beetle outbreaks (Senf et al., 2015; Candotti et al., 2022), defoliating insects (Wittich et al., 2022; Senf et al., 2015), or wind (Rüetschi et al., 2019). Yet, in such local studies, per-class accuracies range from above 90% for high-intensity agents such as fire, to as low as 22–23% for subtle natural disturbances such as insect damage, even at local scales (Stahl et al., 2023; Sebald et al., 2021).

This limitation arises because the "fingerprint" of a disturbance that a satellite can detect, namely its spatial pattern, spectral reflectance, and temporal evolution, is often less distinct for slowly evolving or subtle disturbances than for abrupt ones (McDowell et al., 2015). While harvesting or fire typically produce abrupt, large-scale, and clearly structured signals, biotic disturbances often leave subtler and more heterogeneous patterns. For example, bark-beetle outbreaks may develop gradually over several months or years and occur in small, spatially fragmented patches (Müller et al., 2026), whereas defoliator events are short-lived, with canopy recovery often occurring within a few weeks. Given these distinct ecological impacts, we expect each disturbance type to leave a characteristic pattern across at least some of the spatial, spectral, or temporal dimensions, rather than being entirely indistinguishable from background variability or from one another (Figure 1). Accurately classifying these disturbance types therefore requires data with sufficient spatial resolution to capture small-scale disturbance patches, temporal resolution to detect short-lived events such as defoliation, and sensitivity to vegetation stress signals, such as red-edge (RE), near-infrared (NIR), and short-wave infrared (SWIR), or water content, such as Synthetic Aperture Radar (SAR) (Abdullah et al., 2019; Kautz et al., 2024; Bozzini et al., 2025; Marvasti-Zadeh et al., 2024; Dalponte et al., 2020, 2023; Einzmann et al., 2017). Together, these constraints have limited the development of robust and transferable models for biotic and abiotic disturbance classification.

Recent developments in Earth observation provide new opportunities to distinguish biotic and abiotic disturbance agents. Existing disturbance classification studies rely primarily on Landsat (Stahl et al., 2023; Viana-Soto and Senf, 2025; Senf and Seidl, 2021), which, while offering long and consistent time-series, has moderate (16-day revisit) temporal resolution and does not resolve the red-edge band that has proven critical for detecting biotic disturbances (Abdullah et al., 2019; Kautz et al., 2024). Since 2016, the Sentinel-1 and Sentinel-2 satellite missions have enabled the acquisition of multi-sensor data at high spatial (10–20 m) and temporal (5-12 days) resolution. Sentinel-2 captures key spectral ranges, including the red-edge, near-infrared (NIR), and shortwave infrared (SWIR), which are sensitive to changes in vegetation structure, water content, and physiological stress. These spectral ranges are particularly relevant for detecting biotic disturbances such as bark-beetle infestations and defoliating insects (Abdullah et al., 2019; Kautz et al., 2024; Marvasti-Zadeh et al., 2024; Dalponte et al., 2020; Einzmann et al., 2017). In addition, Sentinel-1 provides C-band SAR backscatter (VV, VH), which is sensitive to canopy

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

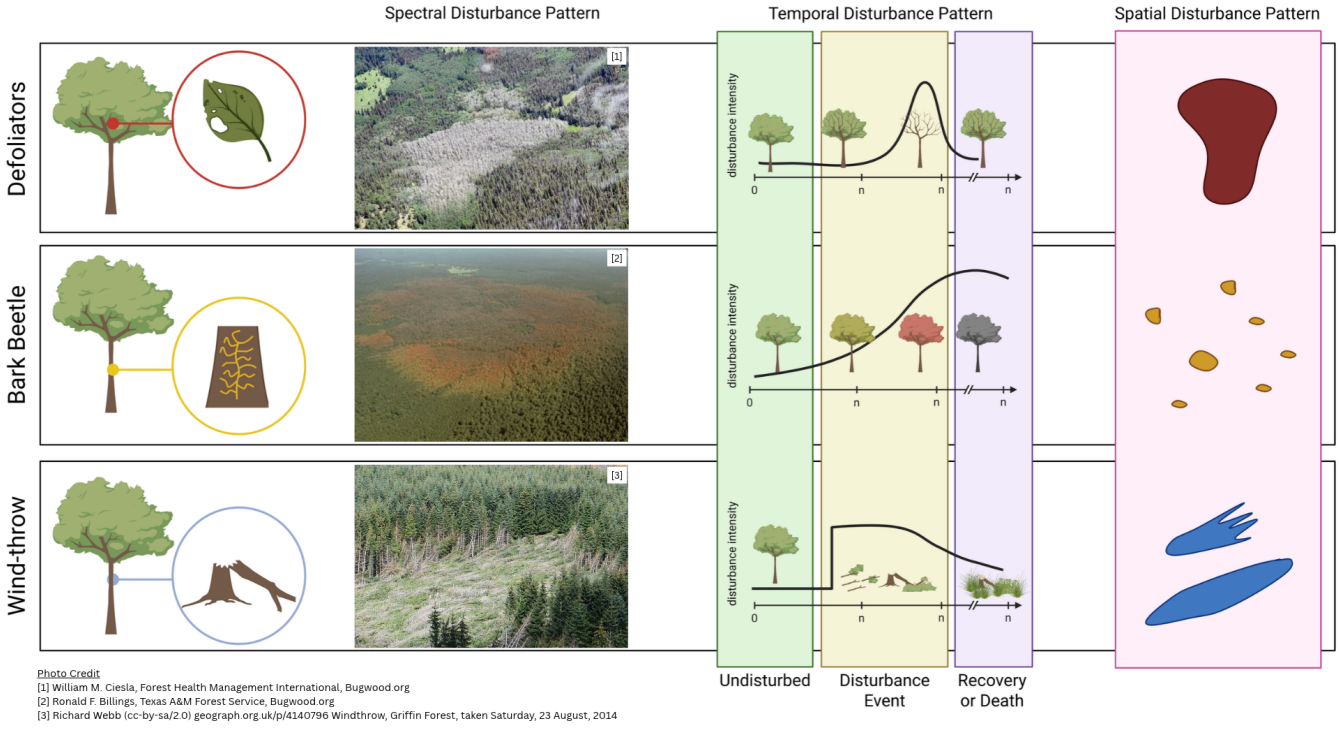


Figure 1. Visual abstract illustrating the study hypothesis. Rows show the three forest disturbance types (wind, bark beetle, defoliators), and columns summarize their expected spectral, temporal and spatial fingerprints. All elements are schematic and highlight the main differences that underlie our classification framework. The photos used in this figure are created by [1] William M. Ciesla, Forest Health Management International, Bugwood.org, [2] Ronald F. Billings, Texas A&M Forest Service, Bugwood.org and [3] Richard Webb (cc-by-sa/2.0) geograph.org.uk/p/4140796 Windthrow, Griffin Forest, taken Saturday, 23 August, 2014.

structure and moisture conditions and allows for improved disturbance mapping (Dalponte et al., 2020, 2023; Einzmann et al., 2017; Müller et al., 2026) at the same high spatial and temporal resolution of Sentinel-2.

Adding to these new sensors, recent advances in machine learning (ML) and deep learning (DL) expand the range of approaches to analyze high-dimensional Earth observation data, offering powerful tools for identifying complex patterns across spatial, spectral, and temporal dimensions. Despite this progress, most forest disturbance classification studies continue to rely on a relatively narrow set of models. According to the review by Stahl et al. (2023), 59% of disturbance studies (20 out of 34) primarily used Random Forest classifiers, 27% employed decision-tree-based models, and only 15% tested other approaches such as Support Vector Machines or specialized rule-based algorithms. Nearly all studies considered used spectral predictors (e.g., Normalized Burn Ratio, Tasseled Cap indices), and around 79% incorporated temporal information, but only about the timing or year of disturbance onset.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

This suggests a substantial opportunity to improve disturbance classification by exploring advanced ML and DL algorithms that can better capture the subtle spatial, spectral, and temporal fingerprints of natural disturbances. However, to our knowledge no study has systematically examined which specific predictors are actually needed to differentiate these natural disturbance types from one another, nor tested which combinations of features and model architectures best achieve this separation. We therefore ask a focused question: which spatial, spectral, and temporal predictors, and which model architectures, allow wind, bark-beetle, and defoliator disturbances to be separated in a way that generalizes transferability within the study region? Building on recent advances in high-resolution satellite data and machine learning, we provide a systematic evaluation of spatial, spectral/radiometric, and temporal predictors and machine learning architectures for distinguishing among natural forest disturbances that have been historically hard to separate, specifically wind, bark beetle outbreaks, and defoliator events.

To address this gap, we use a new dataset for the southeastern USA, the Sentinel-1 Disturbance Mapping product (S1DM, Müller et al., 2026), which provides disturbed patch information at 20 m spatial resolution from 2016–2021. Drawing on spatial, spectral, and temporal information derived from Sentinel-1 and Sentinel-2 data, we evaluate a range of established machine learning models, e.g. random forest, support vector machines, as well as more recently developed architectures, to determine what combination of predictors, type of information and model architectures is best suited to distinguish between these tree disturbance types.

2 Data

2.1 Sentinel-1 Disturbance Map

The Sentinel-1 Disturbance Map (S1DM) dataset (Müller et al., 2026) provides spatially explicit disturbance polygons covering USDA Forest Service Region 8 in the southeastern United States (U.S. Department of Agriculture, Forest Service, n.d.), spanning 30°N–37°N and 96°W–76°W from eastern Texas to the North Carolina coast (see Fig. A1). S1DM is a hybrid product combining spatial delineation of disturbed patches based on Sentinel-1 SAR change detection (2016–2021; Cremer et al., 2020) with disturbance agent information based on the U.S. Forest Service's Insect and Disease Survey (IDS, 2017–2020; US Forest Service, 2013). The disturbance agents are specifically fire, drought, wind, bark beetle, and defoliator agents, though our study focuses on the latter three. The data is provided with unique identifiers (`IDX_D`) at 20 m resolution, with a $\pm$500 m spatial and $\pm$2-year temporal buffer to account for IDS positional and timing uncertainty and patches larger than 15 km^2 are excluded. In SE USA and the period considered, bark beetle and wind disturbances have a much higher prevalence than defoliator disturbances. S1DM reflects this imbalance, with 469 *wind*, 899 *bark beetle*, and 144 *defoliator* events (Müller et al., 2026). In the dataset a single event may be linked to multiple Sentinel-1 detection years where structural change was recorded across consecutive years (e.g., gradual dieback or multi-year defoliation). All three classes are distributed across the same geographic extent, with no class spatially isolated from the others. The dataset excludes compound disturbance events, ensuring each record corresponds to a single and clearly defined disturbance agent. This has the advantage of allowing detection of the spatial, reflectance, and temporal fingerprint of the isolated disturbance agent, without other disturbance signals confounding the signal. A detailed technical description can be found in Müller et al. (2026).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

2.2 Sentinel-1 and Sentinel-2 data

We extracted Sentinel-1 SAR and all 12 L2A Sentinel-2 (S2) spectral reflectance bands (B01–B12) for each disturbance event in S1DM, and over the period 2016–2024. Sentinel-1 (S1) provides C-band SAR data (dual polarization VV/VH) with a native spatial resolution of approximately 20 m, distributed at a 10 m pixel spacing, and a revisit frequency of 6–12 days (Torres et al., 2012), providing information on canopy structure that is not affected by clouds. The S1 product we use here is terrain-flattened γ^0, from GRD IW, in linear power units. Sentinel-2 records multispectral reflectance at native resolutions of 10 m (visible and NIR bands), 20 m (red-edge, additional NIR, and SWIR bands), and 60 m (atmospheric bands), with a 5-day revisit time (Drusch et al., 2012), allowing for analysis of vegetation condition (greenness, water content and pigmentation). The combined use of SAR and optical data ensures sensitivity to both structural changes and canopy discoloration.

To extract Sentinel-1 and Sentinel-2 data, we overlaid a 10 m-resolution Equi7 grid (Bauer-Marschallinger et al., 2014) over North America, subdivided it into non-overlapping 512 × 512 pixel tiles (≈5.12 × 5.12 km), and retained only tiles intersecting disturbance events from S1DM. This tile size was chosen to balance capturing sufficient spatial context for disturbance characterization with maintaining manageable computational and storage requirements for datacube processing. We downloaded the corresponding datacubes as weekly composites from the Copernicus Data Space Ecosystem (CDSE) using the Python package `Sentle` (v2025.12.1; Mosig, 2024), which handles tiling, data retrieval, cloud masking, and snow masking. Cloud detection uses the deep learning-based model of Aybar et al. (2022) rather than the native Sentinel-2 Scene Classification Layer. Residual cloud and cloud-shadow contamination that survives masking is further mitigated by the outlier removal, gap filling, and temporal smoothing described below, and by the 90th-percentile reference phenological envelope used to locate the disturbance timestamp (Sect. 3). We did not quantify per-scene cloud-cover statistics, which adds uncertainty to the timing of the disturbance signal. All bands were resampled to a common 10 m resolution regardless of native resolution and for each tile, we aggregated the bands into weekly time composites. The resulting data cubes then underwent a three-step pre-processing pipeline.

First, temporal outliers were identified per pixel and per band by computing a centered rolling median with a window of 3 time-steps along the temporal dimension to obtain a robust local reference. Observations deviating from this reference by more than 1.9 times the temporal standard deviation of the pixel's full 2016–2024 time series were flagged and removed, which removes short-lived artifacts such as residual clouds or speckle while preserving persistent changes such as disturbances.

Second, missing values resulting from outlier removal or data gaps were filled by linear interpolation or forward/backward filling along the temporal dimension. From 2018 onward, when both Sentinel missions operated with regular systematic coverage, gaps were short. The median gap was 2 time steps (∼14 days) long and 95% of all gaps spanned fewer than 7 time steps (∼49 days), so we applied linear interpolation to these gaps. In the early mission period (2016 to 2017) and after the failure of Sentinel-1B in December 2021, the data coverage is less consistent, with median data availability per weekly timestep before mid-2017 was below 20%, meaning that for the majority of weeks in this period no valid observations existed at all. These gaps at the leading or trailing edge of the time-series, where no valid observations exist on both sides, were filled by forward- and backward-fill, propagating the nearest valid observation. This method produces a flat, constant signal rather than a meaningful

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

spectral trajectory. Since no disturbance events in our dataset took place before 2017, we retain these interpolation artifacts but acknowledge that for longer temporal feature windows (e.g., 5-year and 8-year time-series), the pre-disturbance baseline before 2017 is largely interpolated from the first available real observation rather than from observed phenology.

Finally, we applied a Savitzky–Golay filter (Savitzky and Golay, 1964) (window length = 15, polynomial order = 3) to each time-series to reduce high-frequency noise while preserving seasonal dynamics and disturbance-related signals. This filtering step also partially smooths artifacts introduced by linear interpolation across longer gaps. We note that this window length, while suitable for capturing gradual disturbance dynamics such as bark-beetle outbreaks, may attenuate the short-lived signal characteristic of defoliator events, and we account for this limitation when interpreting model performance for this disturbance class.

To better characterize vegetation dynamics associated with disturbances, we derived a set of vegetation indices (VIs) from Sentinel-2 spectral bands that are commonly used in remote sensing to detect vegetation stress and disturbance. These indices capture key properties of vegetation, including chlorophyll content and photosynthetic activity, moisture status, and stress. Table A1 summarizes all indices used, including their abbreviations, full names, calculation, the properties of vegetation they relate to, and original references.

Sentinel-1 acquires data in ascending and descending orbital passes that follow different ground tracks. Due to geographic coverage constraints, descending-orbit observations (VV_{desc}, VH_{desc}) were absent for approximately 22% of disturbance polygons and 52% of individual disturbance pixels; ascending-orbit observations were missing for less than 1% of polygons and 2% of pixels. For sites where only one orbit was available, the missing features were substituted from the companion orbit ($VV_{desc} \leftarrow VV_{asc}$ and vice versa). Following this cross-fill, disturbance samples lacking valid observations in both orbits were excluded, corresponding to 0.1% of polygons (2 of 1,512) and 2.2% of pixels (78,297 of 3,585,968).

3 Methods

Our goal is to identify suitable data-driven models based on geometric, remote sensing, and temporal predictors that accurately differentiate between forest disturbance agents, namely bark beetle, defoliators, and wind. Here, we describe the pre-processing of predictors, the selection, training, cross-validation and hyperparameter optimization of machine and deep learning models, and the interpretation of model results to identify the relevant features (Fig. 2).

3.1 Model types and predictor sets

We evaluate two categories of models, distinguished by their input data requirements. The first category are feature-based models, which take a fixed-length vector of geometric and spectro-radiometric remote sensing features as input. The second category consists of time-series classification models, which operate directly on uni- and multivariate weekly time-series input and derive temporal features automatically through random convolution kernels, shapelets or symbolic representations without requiring pre-computed feature engineering. Predictor sets and model architectures are described together within each category.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

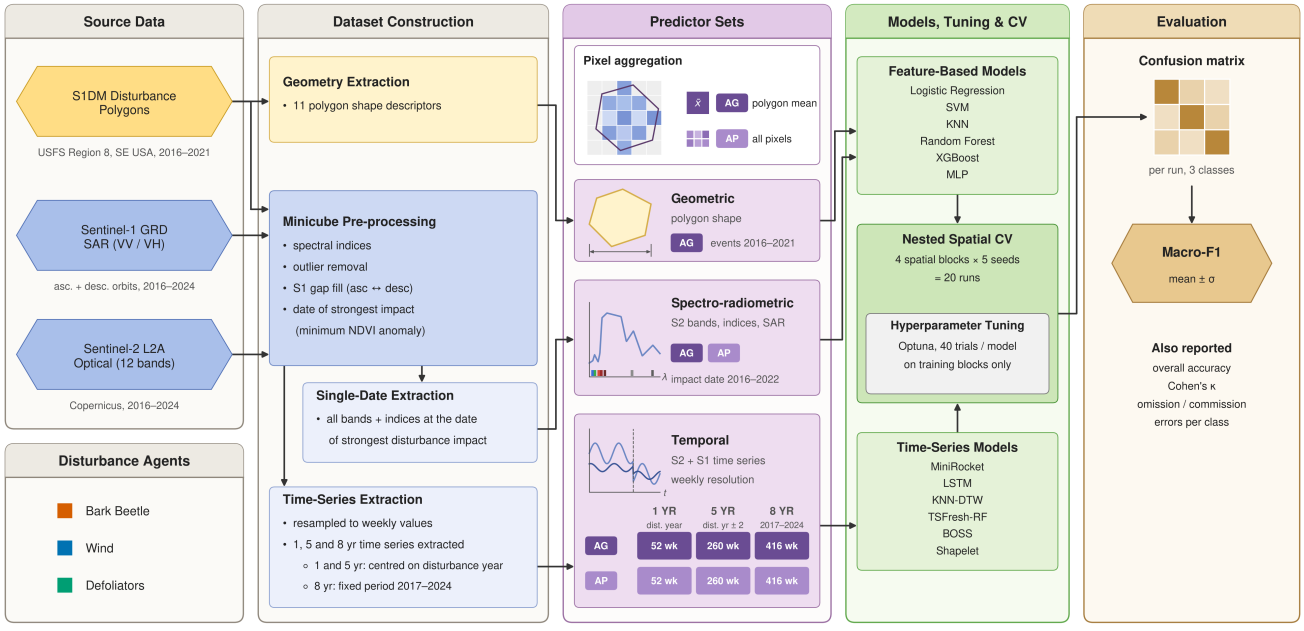


Figure 2. The methodological workflow shows the three sources of data: S1DM disturbance polygons (yellow), Sentinel-1 SAR backscatter and Sentinel-2 optical reflectance (blue). In the dataset construction, eleven shape descriptors are derived from the polygon geometries. To construct the spectro-radiometric and temporal predictors, Sentinel-1 and -2 minicubes are pre-processed and for each disturbance polygon, the date of strongest disturbance impact is defined as the date of the minimum NDVI anomaly. All bands and indices are extracted at this date (single-date extraction), and weekly time series are extracted for three time windows, the 1 and 5 years centered on the disturbance year, and a fixed 8-year period (2017–2024). The resulting predictor sets (spectro-radiometric and temporal) are used either aggregated to the polygon mean (AG) or for all pixels within a polygon (AP). Geometric predictors are, per definition, AG. Feature-based models are trained on the geometric and spectro-radiometric predictors, and time-series models on the temporal predictors. All models are evaluated with nested spatial cross-validation, with hyperparameters tuned on the training blocks only. Performance is reported as $F1_{macro}$ (mean $\pm$ σ), together with overall accuracy, Cohen's κ, and per-class omission and commission errors.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.



3.1.1 Feature-based models

The first predictor set describes the geometry of each disturbance patch derived from S1DM polygons. Geometric features are commonly used as patch-level predictors in disturbance classification (Viana-Soto and Senf, 2025; Schroeder et al., 2017; Schleeweis et al., 2020; Sebald et al., 2021; Senf and Seidl, 2021). We computed 11 metrics per patch using the `Shapely` library (v2.1.2), covering patch size (area, convex hull area), bounding box dimensions (width, height, extent ratio), shape descriptors (Polsby-Popper circularity, elongation), boundary complexity (perimeter, convexity), and internal fragmentation (solidity, number of holes), as detailed in Table B1.

The second predictor set consists of single-date spectro-radiometric features, corresponding to the Sentinel-1 and Sentinel-2 signals at the moment of peak disturbance impact. To identify this timestamp, we computed a patch-level reference phenological envelope (RPE), defined as the mean of the normalized difference vegetation index (NDVI) for each week of the year over the full study period (2016–2024), adapted from the Gorgeous Year concept of Gouveia et al. (2010). Using the full study period rather than only pre-disturbance years is necessary because disturbance events from 2016–2017 have insufficient pre-disturbance observations in the Sentinel-1/2 archive to construct a meaningful phenological baseline. We define the weekly NDVI anomaly as $a_t = \mathrm{NDVI}_{w(t)} - \mathrm{RPE}_{w(t)}$, which is negative when the canopy falls below its reference envelope. The extraction timestamp was then defined as the date of the most negative anomaly (the strongest drop below the envelope, $\arg\min_t a_t$) within a one-year window centered on the reported disturbance year (Figure B1). At this timestamp, we extracted spectro-radiometric variables under different spatial settings. For each disturbance polygon, we tested two spatial approaches. In the aggregated approach (AG), pixel values within the polygon are averaged to produce a single value per event, reducing noise at the cost of discarding within-patch spatial variability. In the all-pixel approach (AP), each pixel within the polygon is treated as an independent sample, retaining spatial variability but introducing autocorrelation between samples from the same event. Models were trained on individual pixels (at most 50 per polygon), but the performance evaluation was calculated by assigning the majority vote of its pixel predictions to each polygon, so that every disturbance event contributes once to the score and is comparable to the evaluation in the aggregated (AG) approach.

The extracted variables consisted of the Sentinel-1 Synthetic Aperture Radar (SAR) backscatter (VV and VH polarizations in ascending and descending orbits) and the Sentinel-2 optical reflectance bands (B02, B03, B04, B05, B06, B07, B08, B8A, B11, and B12), together with twelve spectral indices. The twelve indices (listed in Table A1) were the Normalized Difference Vegetation Index (NDVI), Normalized Burn Ratio (NBR), Normalized Difference Moisture Index (NDMI), Normalized Difference Red Edge Index (NDREI), Green Normalized Difference Vegetation Index (GNDVI), Near-Infrared Reflectance of Vegetation (NIRv), Tasseled Cap Brightness (TCB), Tasseled Cap Greenness (TCG), Tasseled Cap Wetness (TCW), Distance Red & NIR (DRN) and its normalized form (NDRN), adapted from Huo et al. (2021) with NIR in place of SWIR, and the Kernel Normalized Difference Vegetation Index (kNDVI). We tested our models on the SAR backscatter, optical bands, and spectral indices, both as separate groups and in combination.

For the geometric and spectro-radiometric predictor sets described above, we tested models commonly used in forest disturbance classification (Stahl et al., 2023), namely logistic regression (LR; Hosmer et al., 2013), k-nearest neighbors (KNN; Cover

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

and Hart, 1967), random forest (RF; Breiman, 2001b), and support vector machine (SVM; Cortes and Vapnik, 1995), as well as more recently developed methods such as extreme gradient boosting (XGBoost; Chen and Guestrin, 2016) and multilayer perceptron (MLP; Rumelhart et al., 1986). Together, these models span a continuum from linear to highly nonlinear approaches. Margin-based models (SVMs) use high-dimensional boundaries to separate entangled classes, while neural networks (MLPs) learn nonlinear combinations across all features, making both approaches particularly suitable for detecting subtle spectral patterns indicative of specific disturbance types. For each model, we evaluated each predictor set separately (geometry, SAR, optical bands, vegetation indices individually) and in combination, to assess how different models handle overlapping signals across domains.

3.1.2 Time-series based models

For time-series-based models, we derived predictor sets from weekly Sentinel-1 and Sentinel-2 time-series. We tested 12 configurations that differ in three aspects: the feature set, the temporal window, and the spatial aggregation over the disturbance polygon. We tested two feature configurations, providing either a univariate time-series of NDVI only (NDVI), or a multivariate time-series (MV) of ten variables. The variables, namely NDVI, NDMI, GNDVI, NDRN, the Sentinel-2 bands B12, B11, B06, and B05, and the SAR polarizations VV and VH were selected based on a prior feature importance analysis (Section 4.3). Counting ascending and descending VV and VH separately gives 12 input features.

We further considered three temporal windows to evaluate how much pre- and post-disturbance context the models require. For the first, we extracted a one-year time-series centered on the disturbance year (1YR, 52 time steps). For the second, we extracted a five-year time-series centered on the disturbance year, covering two years before and two years after (5YR, 260 time steps), thereby capturing both pre-disturbance conditions and post-disturbance recovery dynamics. For the third, we extracted the complete eight-year weekly time-series from 2017 to 2024 (8YR, 416 time steps), providing maximum temporal context. As the 8Y window is not centered on the disturbance event, pre- and post-disturbance periods are of unequal length.

For each disturbance polygon, we applied the same two spatial approaches (AG and AP) as for the spectro-radiometric data, here producing either a single averaged time-series per event (AG) or one time-series per pixel (AP). The spatial approach determines the cross-validation strategy, as detailed in Section 3.2. Predictor configurations are referred to by combining all three aspects into a single nomenclature, for example NDVI-1YR-AG for the one-year NDVI-only aggregated configuration, or MV-5Y-AP for the five-year multivariate all-pixel configuration.

To classify these time-series, we tested six time-series classifiers to evaluate whether temporal signatures characteristic of each disturbance type can be learned from the data (Figure 1), and whether this improves classification skill over static predictors.

We included a distance-based approach, the K-Nearest Neighbor (KNN; Cover and Hart, 1967) with Dynamic Time Warping (KNN-DTW; Sakoe and Chiba, 1978), which compares entire time-series for shape similarity while accounting for temporal misalignment, thereby effectively assessing whether one temporal trajectory resembles another. We further include the Time-series feature extraction (TSFresh; Christ et al., 2018b) combined with Random Forest (RF; Breiman, 2001b). TSFresh summarizes the temporal behavior into descriptive statistics such as mean, maximum, variance, number of peaks, or slope of the series,

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

rather than using the raw time-series. These features are then used in RF classification. However, simple summary statistics may fail to capture ecologically meaningful events, such as a short decline followed by rapid recovery as observed in defoliator outbreaks. Therefore, we further test two pattern-based approaches, namely the Bag-of-SFA-Symbols (BOSS; Schäfer, 2014) and the Shapelet Transform Classifier (STC; Hills et al., 2014) models, which identify recurring motifs or discriminative subsequences which could be associated with disturbances. SFA (Symbolic Fourier Approximation) discretizes time-series into symbolic representations, and BOSS identifies recurring symbolic motifs associated with each class. STC instead searches for discriminative subsequences, called shapelets, whose presence or absence is predictive of the disturbance type. Finally, we incorporated two approaches that learn relevant features directly from the data rather than relying on manually engineered statistics. Long Short-Term Memory (LSTM; Hochreiter and Schmidhuber, 1997a) networks process the time-series sequentially, maintaining a memory of past observations to capture temporal dependencies across the full series length. The RandOm Convolutional KErnel Transform (ROCKET; Dempster et al., 2020) instead applies a large number of random convolutional kernels to transform the time-series into a high-dimensional feature space, from which a linear classifier is trained. We use its computationally lighter MiniRocket variant (Dempster et al., 2021) in all experiments and refer to it as MiniRocket throughout. Therefore, our model selection allows us to test how different representations of temporal information, ranging from global similarity to localized patterns and learned feature representations, can contribute to differentiate between disturbance types. Additional details on time-series based models are provided in Table B3. BOSS and the Shapelet Transform Classifier were evaluated solely on the univariate NDVI series, since the the BOSS implementation in sktime supports only univariate input, and the shapelet search was computationally to costly for the 12-feature timeseries.

3.2 Model training and evaluation

3.2.1 Spatial block definition

Given the small size of our dataset, the imbalance between classes and likely spatial autocorrelation between samples, we implement a strict cross validation using a nested approach in order to avoid overfitting.

First, we defined four spatial cross-validation blocks by grouping all polygon centroids into 25 clusters by k-means clustering of their coordinates (Fig. 3a) and then grouping them into four groups that balanced the proportions of the three disturbance agents and their sum across blocks. For this, we first distribute the clusters $m \in [1,25]$, depending on the rarest disturbance class (defoliators $<$ wind $<$ bark beetle) present in cluster m. Each cluster m is assigned to the block B which holds the fewest polygons of its rarest class at the moment, except for bark beetle, the dominant class, whose clusters are assigned to the block with the fewest polygons overall. After the initial assignment, we then shuffle other clusters between blocks in order to minimize the objective function

$$J = \sum_{c}^{3} w_c \, \sigma_k(n_{k,c}) + 2 \sum_{k=1}^{4} \left(\sigma_k^{\text{lat}} + \sigma_k^{\text{lon}}\right) \tag{1}$$

$n_{k,c}$ is the number of disturbance events of class c in block k (referred to as B_k below), with $c \in \{\text{bark beetle}, \text{wind}, \text{all events}\}$, as defoliators are balanced in the beginning. w_c corresponds to the weight of class c, with $w_c = 1.5$ for bark beetle, 1.0 for

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

wind and 0.5 for all events. The weights are needed because as moving a cluster to balance one class also changes the other class counts and the block size. Higher weights give a class higher priority for even distribution across blocks. We chose the weights empirically by testing different combinations and selecting the one that gave the most balanced distribution of each disturbance class across the four blocks. In the first term of the objective function, $\sigma_k(\cdot)$ is the standard deviation across the four blocks and in the second term σ_k^{lat} and σ_k^{lon} are the standard deviations of the latitudes and longitudes of the cluster centers in block B_k. The spatial distribution of the four blocks is shown in Fig. 3a. Block 1 has 399 events (of which, 73% bark beetle, 20% wind and 6% defoliator), block 2 318 events (61% bark beetle, 27% wind and 12% defoliator), block 3 405 events (84% bark beetle, 8% wind and 8% defoliator) and block 4 450 events (26% bark beetle, 63% wind and 11% defoliator). Because the assignment prioritizes similar class proportions across blocks, clusters of different blocks can be direct neighbors, so that some spatial autocorrelation might still be present.

3.2.2 Metrics used for model training and evaluation

As our primary metric to evaluate each model's performance during training and evaluation, we use the macro F1 score (F1_{macro}), corresponding to the unweighted mean of the F1 scores of the three disturbance classes:

$$\text{F1}_{macro} = \frac{1}{3}\sum_{c=1}^{3}\text{F1}_c, \qquad c \in \{\text{wind, bark beetle, defoliator}\} \tag{2}$$

and, the per-class is score is given by:

$$\text{F1}_c = 2 \cdot \frac{\text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \tag{3}$$

Precision and recall are defined in Tab. B5. For a 3 class classification problem with our class imbalance, a F1_{macro} of 0.3 would indicate that the model is not better then random guess. A F1_{macro} of 0.25 would indicate that the model always predicts the majority class only (here, bark beetle). For an interpretation of the F1_{macro} score for a 3 class classification with imbalance between classes, refer to Table B6.

We additionally report the macro precision, macro recall, overall accuracy (OA), Cohen's κ for all models. For the best model of each predictor types (see Section 3.1), class-wise F1, omission and commission as defined in Tab. B5.

3.2.3 Nested cross-validation

Cross-validation was then performed for each model type (Tables B2,B3) using a nested approach with k *outer* folds, with $k \in [1, 4]$, corresponding to the number of spatial blocks B, where in each fold k, block B_k is held off once as a test set T_k for independent evaluation of a given model (Fig. 3b). Each outer fold consists of two steps, applied to the remaining three blocks ($D_k = \{B_n\}, n \neq k$). First, a hyperparameter tuning loop is performed in order to optimize the hyperparameters corresponding to each model type (Table B4). Based on these optimized hyperparameters, the model is then trained and evaluated.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Hyperparameter tuning

In order to find the best hyperparameter set within each outer fold k, we perform a *leave-one-out cross validation* with three inner folds j, where $j \in [1,4], \forall j \neq k$. For each inner fold, one of these three blocks serves as a validation set $V_{k,j}$, defined as $V_{k,j} = B_j$, once. The remaining two blocks are used as a training set ($D^*_{k,j} = \{B_n\}, n \neq j,k$) for hyperparameter tuning using the Optuna *python* package (Akiba et al., 2019). Per model type, we predefine a group of hyperparameters to be tuned and the corresponding search space Q (see Table B4). Within this space, Optuna runs 40 trials $\Theta = \{\theta_i\}, i \in [1,40]$ and evaluates each combination on all three cross validation folds. In each trial, the default Optuna sampler proposes one setting θ_i, using the scores of the previous trials to choose promising regions of the search space. The best hyperparameter set θ^*_k of each model in each outer fold k is the set that maximizes the model performance averaged across the three inner folds ($\overline{F1}_{macro,k,\theta}$):

$$\theta^*_k = \underset{\theta \in \Theta_k}{\arg\max}\, \overline{\mathrm{F1}}_{\mathrm{macro},k,\theta} \tag{4}$$

Model training

Per outer fold, and with θ^*_k fixed , each model is then trained on the D_k set and evaluated on the held-out test block T_k five times ($r \in [1,5]$). Per repetition, the models is trained on a randomly perturbed, class-stratified 80% subset of D_k, and the remaining 20% is excluded (see Fig. 3b). These perturbations are conducted with the random seeds $s \in [42, 67, 123, 314, 2024]$. The LSTM additionally holds out 15 % of the training subset for early stopping, trains on the other 85% and stops if there is no improvement in the performance after 20 epochs. During the nested cross validation this leads to a total of 20 training and evaluation runs per model in the *outer loop*.

Class imbalance sensitivity

We finally test how sensitive our results are to the class imbalance present in S1DM by reevaluating the best-performing model of each predictor set with (i) random undersampling of the majority classes, (ii) random oversampling of the minority classes by duplication, and (iii) class weights inversely proportional to class frequency. Undersampling reduces every class to the size of the smallest class and oversampling increased every class to the size of the largest class. We perform this sensitivity test in the training step of each outer fold, resampling only to the training folds D_k.

3.3 Model Interpretability

We perform a *feature ablation* analysis on the the best temporal model-predictor combination in this study. The ascending and descending orbits of each SAR polarization are treated as one feature here, since gaps in one orbit had been previously been filled from the other, which gives in total ten features. During *feature ablation* the trained model (as described in Sec. 3.2) is applied to the test data once with all features and for each feature individually removed. The importance of feature a is defined

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

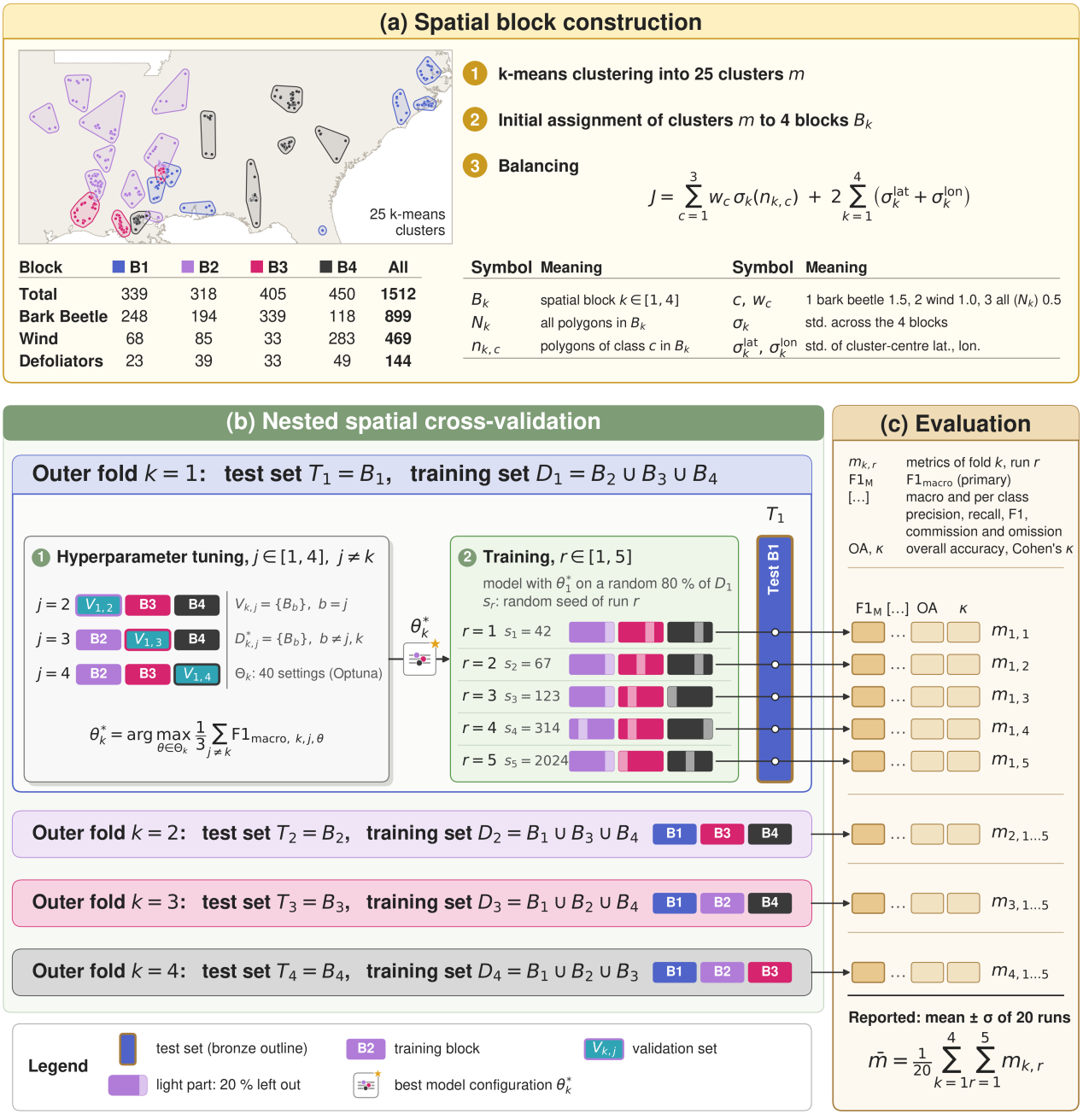


Figure 3. Schematic overview of the nested spatial cross-validation. Panel (a) explains the construction of the four spatial blocks. Panel (b) illustrates the outer loop of the nested cross-validation. Here, in each outer fold, we exclude one spatial block as the test set, while the other three spatial blocks serve as the training set on which the models' hyperparameters are tuned and the final model is fitted. The final model in each outer fold was trained five times on a random, class-stratified 80,% of the training blocks while the the remaining 20,% where excluded. In panel (c), different metrics estimate the overall and per-class performance over the cross-validation runs.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

as the reduction in $\mathrm{F1}_{\mathrm{macro}}$ when a was set to zero,

$$\Delta\mathrm{F1}_{\mathrm{macro},a} = \mathrm{F1}_{\mathrm{macro}} - \mathrm{F1}_{\mathrm{macro},a} \tag{5}$$

where $\mathrm{F1}_{\mathrm{macro}}$ is the macro-F1 of the predictions with all features and $\mathrm{F1}_{\mathrm{macro},a}$ the macro-F1 with feature a ablated.

We conduct this analysis with the hyperparameters selected in the nested spatial cross-validation, as described in Sec. 3.2. With these, we predict the polygons of each test block for each of the five repetitions once with the full feature set and once for each of the ten features set ablated. As a result, every one of our 1512 polygons in the dataset had one true class and 55 predicted classes (5 repetitions $\times$ (10 ablated features $+1$ full feature set)). From this we compute $\mathrm{F1}_{\mathrm{macro}}$ from all 1512 polygons for each repetition and each ablated feature, and $\Delta\mathrm{F1}_{\mathrm{macro},a}$ averaged over the five repetitions. This produces a ranked importance profile that quantifies the contribution of each feature to the model's prediction. A large $\Delta\mathrm{F1}_a$ means that the model depends strongly on the a feature.

We estimate the uncertainty of $\Delta\mathrm{F1}_{\mathrm{macro},a}$ with a cluster bootstrap with 2000 repetitions. The bootstrap reuses the stored predictions of all four spatial blocks and did not retrain or re-apply the models. In each bootstrap repetition, a new test set is drawn from the existing polygons and $\Delta\mathrm{F1}_{\mathrm{macro},a}$ is recalculated from their predictions. Normally bootstrap would resample individual polygons, so that some are drawn several times and others not at all. However, neighboring polygons often share forest type, weather, and even the same disturbance event, and are therefore not independent. We instead resample whole groups of neighboring polygons, using the 25 spatial clusters from which the four spatial blocks were built (Fig. 3). In each repetition, the clusters of every block are drawn with replacement. We recompute $\Delta\mathrm{F1}_{\mathrm{macro},a}$ for all features on the polygons of these selected clusters. The 2.5th and 97.5th percentiles over the 2000 repetitions give the 95 % confidence interval. We consider the importance of a feature consistent when the lower end of its interval is above zero.

We conduct a temporal permutation test on the best temporal model in our study, to estimate weather the best model relies on the temporal trajectory or simply gains it predictive advantage from the larger number of observations compared to the static predictor sets. The sample size between the static predictors (4 to 28 data points per sample) and temporal predictors (12 variables $\times$ 52, 260 or 416 time steps for the 1YR, 5YR and 8YR windows, amount to 624, 3120 or 4992 values in total respectively) varies greatly. We permute the time axis of the best temporal model before training, either identically for all polygons or independently for each polygon. In the shared permutation, one random reordering of the weekly time steps is applied to all polygons (for example week 37 becomes the first time step and week 112 the second for all polygons), which destroys the shape of the trajectory but keeps the observations aligned across polygons.In the per-polygon permutation, each polygon receives its own random reordering of the weekly time steps (e.g. the first time step is week 37 for one polygon and week 205 for the next, even if both belong to the same disturbance agent), which removes all temporal order and alignment between polygons even within the same class. We test 20 random permutations per test and evaluated each with the same nested spatial cross-validation and hyperparameters as the intact series. A positive $\Delta\mathrm{F1}_k$ for the permuted timeseries indicates that the temporal order is informative, while no significant difference indicates that simply the number of multi-temporal observations is important.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

4 Results

Bark beetle, defoliator and wind disturbances are likely to differ in space, spectral or backscatter signal, or time, as proposed in Fig. 1. Figure 4 exemplifies these differences for three events in S1DM, showing that bark beetle, defoliators, and wind indeed leave distinct signatures in space, spectral reflectance, and SAR backscatter. All three disturbances correspond to an NDVI drop within the disturbance year followed by recovery, but the individual signals differ in depth, timing, and speed. Wind causes an abrupt, deep drop with slow recovery well aligned with the disturbance year. Bark beetle shows a similar NDVI and VV pattern to wind but recovers faster and negative anomalies are already present prior to the disturbance year. Defoliators produce a more short-lived and sharper NDVI decline, with an early signal already visible in the VV backscatter signal in the months to a year before the optical disturbance timestamp. Spatial or geometric differentiation is less clear, but both wind and bark beetles show larger patches, with wind having a more compact geometry. We note that spatial characteristics are, however, strongly influenced by terrain and structure of forest stands.

4.1 Overall classification performance

To identify the most suitable approach for classifying wind, bark beetle and defoliator disturbances, we evaluate static predictor types (geometric, radar, optical) and univariate and multivariate time-series of varying lengths (see Sec. 3).

4.1.1 Static feature models

Static features, regardless of which sensors or how many predictors are included, cannot reliably generalize disturbance classification across geographic regions. Single-predictor sets, whether derived from disturbance patch geometry (shape, area, compactness) or from a single sensor at one timestamp, all perform comparably under spatial CV (best model $F1_{macro}$ = 0.45–0.57; Tab. 1). The results reported in Tab. 1 are for the AG setting, as geometric predictors are inherently polygon-level; we present the spectro-radiometric results aggregated the same way for comparability. These $F1_{macro}$ scores indicate that the models learn patterns ($F1_{macro}$= 0.53-0.61), but with a high error rate. Disturbance type is neither recoverable from polygon shape alone nor from a single spectral or SAR backscatter snapshot, showing that static, single-source information is insufficient for spatially generalizing classification. Combining sensors adds a small performance improvement ($F1_{macro}$ = 0.58–0.65). Adding SAR to the vegetation indices predictors leads most models to achieve their highest score of all spectro-radiometric predictor sets ($F1_{macro}$ = 0.65), whereas other combinations, such as all predictors together or the combination subset, do not improve further and plateau around near $F1_{macro}$ of 0.60. This ceiling likely reflects an inherent limitation of the single-timestamp approach: a spectral snapshot captures the state of the forest at one point in time, but not the pre- and post-disturbance trajectory that distinguishes the agents. MLP achieves the highest scores for several predictor sets (geometry, the combination subset, and, tied with other models, the two larger sensor combinations), while SVM performs best on the optical predictor sets and on their combination with SAR.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

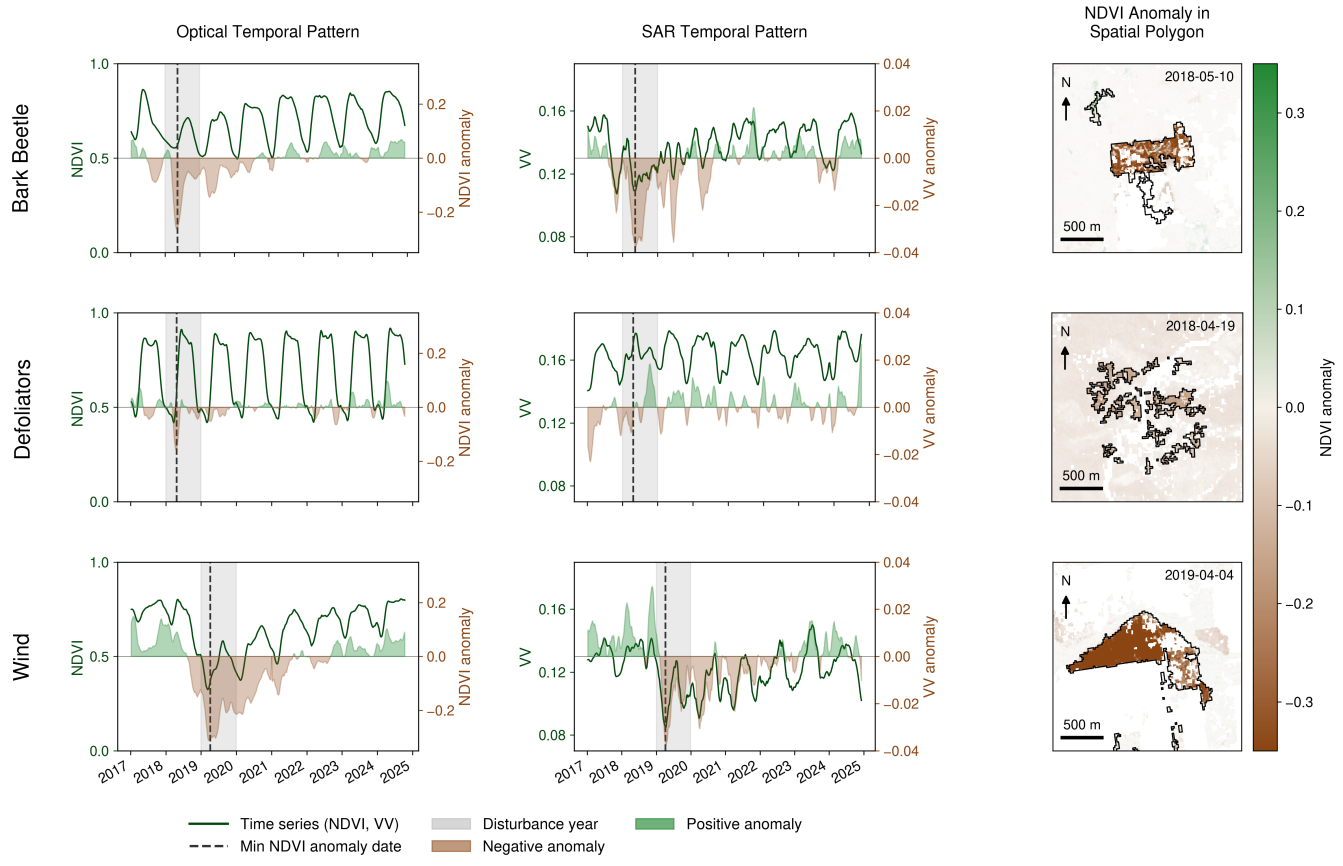


Figure 4. Examples of spatial and temporal signatures of bark beetle, defoliator and wind disturbances (from top to bottom). The first column shows the optical temporal pattern over an eight year period for NDVI on the left y axis and the NDVI anomaly on the right y axis. Positive anomaly values shaded in green and negative values shaded in brown. The second column shows the same temporal pattern for SAR (Synthetic Aperture Radar) VV backscatter over the same eight year period. The last column shows the spatial delineation of the disturbance event in S1DM (black line), overlaid on the NDVI anomaly at the time of strongest disturbance impact.

4.1.2 Time-series based models

The six temporal models evaluated (Tab. 2) were trained on both univariate and multivariate time-series with varying lengths for the AG set (see Sec. 3). The addition of temporal information improves the performance over static features. With performance advances depending primarily in an interplay of radiometric and spectral richness and the length of the observation window (Tab. 2; Fig. 5).

For univariate inputs (NDVI), classification remains modest and window-length dependent. MiniRocket achieves the highest scores among the six models, with $F1_{macro}$ scores between 0.49 (1Y) and 0.62 (5Y). Performance peaks at 5 years for all models and degrades at 8 years, suggesting that a time-series not centered on the disturbance cannot overcome the temporal

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 1. Overview of the $F1_{macro}$ score [0–1] for six machine learning models evaluated on the S1DM dataset across differing geometric and spectro-radiometric predictor sets for the Aggregated (AG) setting. Models evaluated are Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP). The model columns are ordered from linear to increasingly nonlinear and complex models. The values indicate the mean (± standard deviation) performance across the 20 repetitions of nested spatial cross-validation. The feature column indicates the number of predictors used in each dataset. The highest value for each predictor set is highlighted in bold.

Predictors	No. of features	Models					
		LR	SVM	KNN	RF	XGB	MLP
Geometric predictors							
Geometry	11	0.47 ± 0.06	0.48 ± 0.08	0.45 ± 0.09	0.48 ± 0.08	0.50 ± 0.08	$\mathbf{0.57 \pm 0.09}$
SAR predictors (Sentinel-1)							
VV/VH	4	0.46 ± 0.08	0.49 ± 0.09	$\mathbf{0.53 \pm 0.10}$	0.52 ± 0.10	0.52 ± 0.10	0.51 ± 0.10
Optical predictors (Sentinel-2)							
S2 bands	10	0.50 ± 0.06	$\mathbf{0.53 \pm 0.05}$	0.45 ± 0.07	0.51 ± 0.05	0.51 ± 0.04	0.52 ± 0.06
Vegetation indices	12	0.60 ± 0.08	$\mathbf{0.61 \pm 0.08}$	0.45 ± 0.06	0.50 ± 0.07	0.49 ± 0.08	0.55 ± 0.05
Sensor combinations							
Combination subset[a]	12	0.59 ± 0.08	0.56 ± 0.12	0.56 ± 0.12	0.56 ± 0.10	0.58 ± 0.09	$\mathbf{0.60 \pm 0.07}$
VV/VH + S2 bands	14	0.55 ± 0.08	0.51 ± 0.09	0.52 ± 0.11	$\mathbf{0.58 \pm 0.09}$	0.57 ± 0.08	$\mathbf{0.58 \pm 0.09}$
VV/VH + Vegetation indices	16	0.62 ± 0.07	$\mathbf{0.65 \pm 0.06}$	0.56 ± 0.12	0.58 ± 0.11	0.59 ± 0.07	0.63 ± 0.09
VV/VH + S2 bands + Vegetation indices	26	$\mathbf{0.60 \pm 0.06}$	0.58 ± 0.06	0.56 ± 0.12	0.59 ± 0.10	0.59 ± 0.09	$\mathbf{0.60 \pm 0.12}$

[a] Features included: NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

label ambiguity. Multivariate inputs shift this picture substantially, as the multivariate predictor sets outperform the univariate sets across most models. Good classification performance is already achievable with a single year of observations (MiniRocket and TSFresh RF: $F1_{macro}$ = 0.68 and 0.67 at 1 yr), with small gains from extending to five years (0.71-0.72 at 5 yr). LSTM and KNN-DTW performs worst among the models at every multivariate window, with scores ($F1_{macro}$ = 0.56–0.64) remain up to 0.1 below the other models.

The 5-year window centered on the disturbance year performs best overall, but with small gains compared to 1YR, and performance declines at 8 years across all models and predictor sets; the gap between 1- and 5-year multivariate results (0.04 $F1_{macro}$) is far smaller than for univariate inputs (0.09 $F1_{macro}$). Several of these differences are small relative to the reported standard deviations (for example MiniRocket 0.72 ± 0.13 vs. TSFresh RF 0.71 ± 0.08 at 5 yr). We therefore emphasize patterns that are consistent across windows and models rather than differences between individual configurations, which would require a paired significance test across folds to establish. When multiple spectral and SAR backscatter features are available, temporal constraints appear to not be the primary bottleneck (analyzed in detail below).

Balancing the classes during training did not improve the models (Fig. C2). For the best model of each predictor family, class weights, undersampling and oversampling changed $F1_{macro}$ by at most 0.03 compared with training without balancing (e.g.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 2. $F1_{macro}$ score [0–1] for temporal classification models evaluated on different time-series inputs for the Aggregated (AG) setting. Models are ordered by methodological group, from dictionary-based (BOSS), distance-based (KNN-DTW), feature-based (TSFresh RF), pattern-based (Shapelet) to deep learning (LSTM), and transformation-based (MiniRocket). Values are the mean (± standard deviation) across 20 repetitions of nested spatial cross-validation. Time-series inputs include the three temporal aggregation windows of 1Y-AG (1 year), 5Y-AG (5 years), and 8Y-AG (8 years). Univariate inputs are NDVI timeseries only, whereas multivariate input includes multiple spectral indices and bands. The highest value in each row is highlighted in bold.

Time Window	Models					
	BOSS	KNN-DTW	TSFreshRF	Shapelet	LSTM	MiniRocket
Univariate (NDVI only, 1 predictor)						
1Y-AG	0.41 ± 0.08	0.48 ± 0.18	0.38 ± 0.15	0.45 ± 0.13	0.45 ± 0.13	$\mathbf{0.49 \pm 0.14}$
5Y-AG	0.50 ± 0.10	0.57 ± 0.11	0.38 ± 0.05	0.53 ± 0.08	0.51 ± 0.10	$\mathbf{0.62 \pm 0.12}$
8Y-AG	0.47 ± 0.19	0.52 ± 0.14	0.37 ± 0.10	0.48 ± 0.10	0.37 ± 0.09	$\mathbf{0.53 \pm 0.14}$
Multivariate (Multiple indices/bands, 12 predictors) [a]						
1Y-AG	–	0.56 ± 0.16	0.67 ± 0.08	–	0.59 ± 0.07	$\mathbf{0.68 \pm 0.16}$
5Y-AG	–	0.64 ± 0.12	0.71 ± 0.08	–	0.61 ± 0.07	$\mathbf{0.72 \pm 0.13}$
8Y-AG	–	0.58 ± 0.13	0.61 ± 0.13	–	0.56 ± 0.07	$\mathbf{0.63 \pm 0.14}$

[a] Features included: NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

5YR MiniRocket 0.72 without balancing and 0.73, 0.71 and 0.71 with class weights, under- and oversampling). Undersampling mainly shifted errors between classes rather than reducing them, lowering the omission error of defoliators (5YR from 0.37 to 0.29) at the cost of a higher commission error (0.25 to 0.40), and it reduced $F1_{macro}$ of the spectro-radiometric MLP from 0.60 to 0.54.

4.2 Drivers of classification performance

We examined to which the spatial aggregation, temporal window length, and the combination of radiometric and temporal information contribute to the classification performance (Fig. 5).

We found that spatial aggregation improved classification skill for both static and temporal predictors (Fig. 5a). For the best model using static radiometric-spectral predictors, increasing the number of observations by working at pixel level (RF, $F1_{macro} = 0.53$) did not improve performance compared to using spatially-aggregated predictors (MLP, $F1_{macro} = 0.60$ AG). Increasing the size of the training data did not allow the model to extrapolate patterns in the radiometric-spectral signal better, likely because pixels of the same polygon are strongly correlated and add little independent information. Once temporal information was introduced, polygon means also produced more stable trajectories, likely because averaging over within-polygon spatial variability reduces noise (MiniRocket MV-TS-1YR: $F1_{macro} = 0.60$ AP vs. 0.68 AG), whereas for NDVI-1Y both levels performed similarly ($F1_{macro} = 0.47$ AP vs. 0.49 AG).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

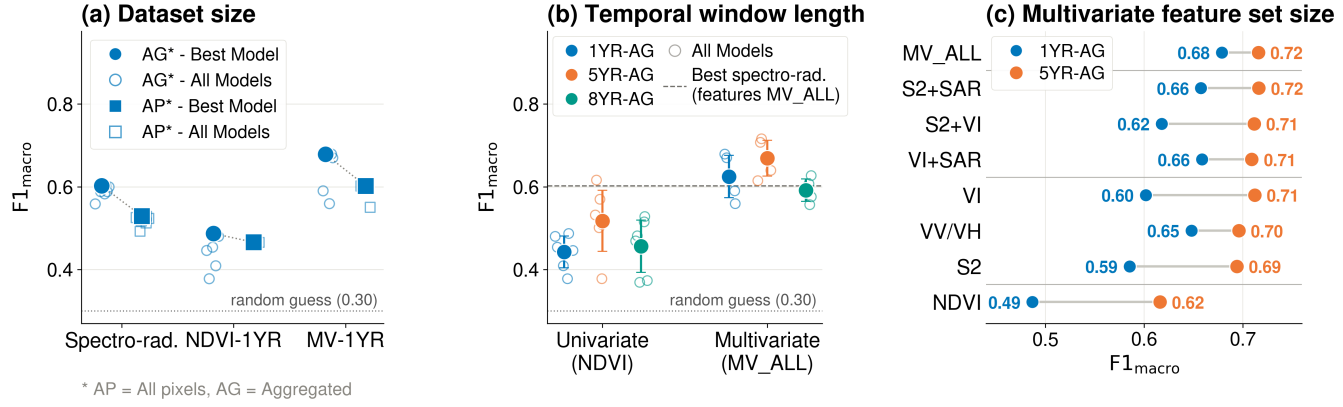


Figure 5. Classification performance ($F1_{macro}$) across three dataset dimensions. Each value is the $F1_{macro}$ of one model on one predictor set, averaged over the 20 runs of the nested spatial cross-validation (4 outer folds × 5 repetitions). Panel (a) shows the difference between the aggregated (AG) and the pixel-wise (AP) classification for the spectro-radiometric, 1YR univariate and 1YR multivariate time series predictors across all tested models. Filled markers indicate the best model, while hollow markers indicate the other models. Panel (b) compares univariate and multivariate time series for the three observation windows (1YR in blue, 5YR in orange and 8YR in green). The hollow circles show the individual models, and the filled circle with whiskers shows the mean ± standard deviation across these models. The dashed horizontal line marks the best spectro-radiometric model on the same 12 features as MV_ALL as a static baseline, and the dotted line indicates the chance level of a uniform random guess ($F1_{macro} = 0.30$). Panel (c) compares eight sensor feature combinations for the 1YR (blue) and 5YR (orange) time windows and shows the best model for each combination. Horizontal lines between the colored dots indicate the temporal improvement gained from extending the observation window.

Extending the observation window in the AG set (Fig. 5b and Tab. 2) improves classification up to five years, after which performance declines, possibly because the 8YR window is not centered around the disturbance year. The benefit of longer temporal windows also depends on predictor type (Fig. 5b). For univariate NDVI, the improvement from 1 YR to 5 YR is substantial (median $F1_{macro}$ across models: $0.45 \rightarrow 0.52$)while for multivariate predictors, the gain is more moderate (1 YR: 0.63; 5 YR: 0.67).

As Tab. 2 and Fig. 5b show, combining longer temporal windows with radiometric and spectral information improves performance. In Fig. 5c we examine how the temporal and the radiometric-spectral information interact to improve classification. The single spatially-aggregated optical predictor tested here (NDVI) at both 1-year and 5-year time windows is not a competitive predictor. Using SAR VV and VH predictors for a single year outperforms 5 years of NDVI values (NDVI-TS-1Y-AG $F1_{macro} = 0.65$; VV-TS-1Y-AG $F1_{macro} = 0.0.62$). The relative value of multivariate SAR backscatter (4 features) compared to optical features (12 features) shifts with the length of the observation window (Fig. 5c). For one year, SAR backscatter outperforms optical features (VV/VH: $F1_{macro} = 0.65$ vs. S2: 0.59, VI: 0.60), likely because SAR backscatter responds more immediately to canopy loss. When using five years, optical features show considerably higher $F1_{macro}$ scores (S2: $0.59 \rightarrow 0.69$; VI: $0.62 \rightarrow 0.70$), while SAR gains only marginally (+0.03). This suggests that pre- and post disturbance trajectories become

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

the primary discriminative signal over longer windows for the optical signals. Combining sensors is most beneficial at one year, where some optical–SAR sensor pairs add up to 0.03 $F1_{macro}$ over the best single sensor predictor set. At five years this advantage disappears, as optical-SAR combinations plateau at around 0.72, and adding the full 12-feature set ($F1_{macro} = 0.72$) offers no improvement over the best eight-feature combination (S2 + VI: 0.72). Classification performance for 5-year windows plateaus at approximately $F1_{macro} = 0.72$ regardless of whether additional features are included.

Finally, we examine how classification performance depends on disturbance type. Bark beetle, the majority class, was the most reliably classified agent for all predictors, with $F1_{bark\ beetle}$ score of 0.67–0.82, and 9–25 % and 20–33% of omission and commission errors (Tabs. C10 and C11). F1 score ranges between 0.36–0.66 for defoliators and 0.25–0.69 for wind, while omission errors range from 0.31–0.68 and 0.33–0.73 for wind. Most misclassifications of the other agents were assigned to bark beetle (Fig. 6).

Furthermore, performance per disturbance type depends considerably on the predictor set (Tabs. C10 and C11). Geometric predictors (Tab. C10) could not separate defoliators (63 % omission, $F1_{defoliators}$ 0.41) and classified only about half of the wind patches correctly (48 % omission), indicating that shape and size alone distinguish bark beetle from the other agents but not defoliators from wind. The spectro-radiometric predictors separate defoliators better (omission 38 % for the best model), but still fail for wind, with omission errors of 45–73 % across the static predictor sets, and 40 % of the wind patches were classified as bark beetle by the best model (Fig. 6). Temporal predictors reach a balanced performance for all three agents (Tab. C11). In Fig. 6 the 5YR MiniRocket model correctly identifies 91 % of the bark beetle, 60 % of the defoliator and 67 % of the wind patches, with commission errors of 23–25 % for all agents and per-class F1 scores of 0.82, 0.64 and 0.69 (see Tab. C11).

4.3 Model Explainability

Understanding which features drive model predictions ensures that classification relies on meaningful physical and ecological signals rather than spurious correlations. To identify the most influential predictors for the best performing model (MiniRocket, MV-5YR-AG), we conduct a feature importance analysis (Sec. 3.3).

The classification performance of the MiniRocket model (MV-5YR-AG) is primarily driven by optical predictors, namely the red-edge bands, SWIR 1 and the chlorophyll-sensitive greenness indices, and only to a lesser extent by SAR backscatter. Ablating Red Edge 2 (B06), Red Edge 1 (B05), SWIR 1 (B11), NDVI and GNDVI reduces the $F1_{macro}$ by 0.26, 0.17, 0.15, 0.12 and 0.10, respectively, with all five features showing a consistent effect across the spatial clusters (bootstrap 95% confidence interval consistently departing from zero). SAR VV and VH are similarly consistent, albeit at a smaller importance (0.09 and 0.07), as was NDRN (0.07). Red Edge 2 (B06) also shows the widest $\Delta F1_k$ interval (0.10–0.42), indicating that its importance varies considerably across spatial clusters. SWIR 2 (B12) and NDMI, in contrast, do not have a consistent effect across the spatial clusters. The model therefore relies on these features only in some regions, which suggests that their contribution depends on local landscape properties.

We have shown that time series outperforms a single-date snapshot of the same 12 variables by a margin (see Sec. 4.1.1, 4.1.2). To test whether the model relies on the temporal order of the observations, we permuted the time axis of the 5YR multivariate series in two ways (Sec. 3.3). The shared permutation increased $F1_{macro}$ slightly from 0.72 to 0.74, while the

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

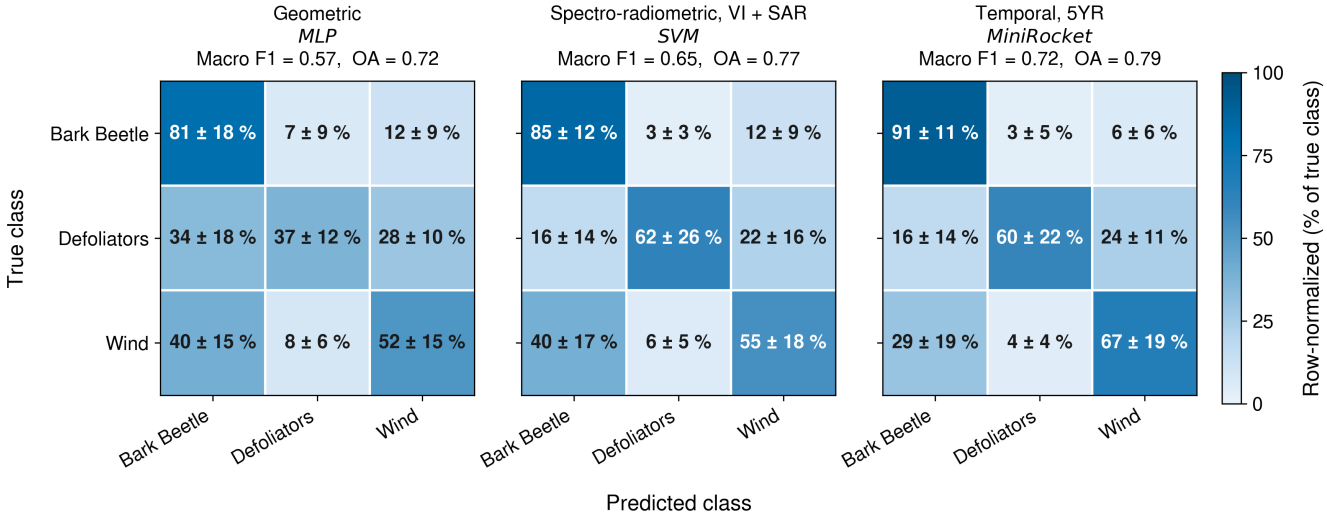


Figure 6. Normalized confusion matrices of the best-performing model for (a) the geometric predictors (MLP, $F1_{macro} = 0.57$), (b) the spectro-radiometric predictors (SVM, $F1_{macro} = 0.65$) and (c) the 5YR multivariate time series (MiniRocket, $F1_{macro} = 0.72$). The $F1_{macro}$ score and overall accuracy are reported above each confusion matrix. The matrices are row-normalized, so each row shows, for one true disturbance type, the proportion of samples assigned to each predicted class; the diagonal therefore gives the proportion of correctly classified samples per disturbance type. True classes are shown on the y-axis and predicted classes on the x-axis. Each cell reports the mean ± standard deviation across the 20 repetitions of the nested spatial cross-validation.

separate permutation lowered $F1_{macro}$ to 0.71. The classification of wind disturbances was most affected ($F1_{wind}$ 0.62 instead of 0.69) by the separate permutations, while the other disturbance classes remained unchanged.

For the 5YR multivariate MiniRocket model, temporal information therefore matters mainly through the number and range of the observations rather than through their sequence. The benefit of multiple observations differs between agents (Table C12). Wind classification improved the most with temporal information ($F1_{wind}$ 0.42 for a single date compared with 0.69 for 5YR) and was the only agent that lost performance when the temporal order was removed (0.62). Bark beetle classification benefited from the number of observations (0.72 to 0.82) but not from their order (0.81 with shuffled dates), and for defoliators a single date performed as well as the time series (Tab. C12b).

5 Discussion

In this study, we aim to determine the most effective modeling approach for classifying and differentiating disturbances caused by wind, bark beetles, and defoliators. These disturbance agents exhibit significantly different phenological dynamics and impact forests through different processes, resulting in distinct geometric, spectral, and temporal signatures in remote sensing data (see Fig. 1; Fig. A2). By evaluating multiple modeling approaches and sets of predictors, we identify model configurations that allow to best distinguish each disturbance type and that show good predictive performance.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

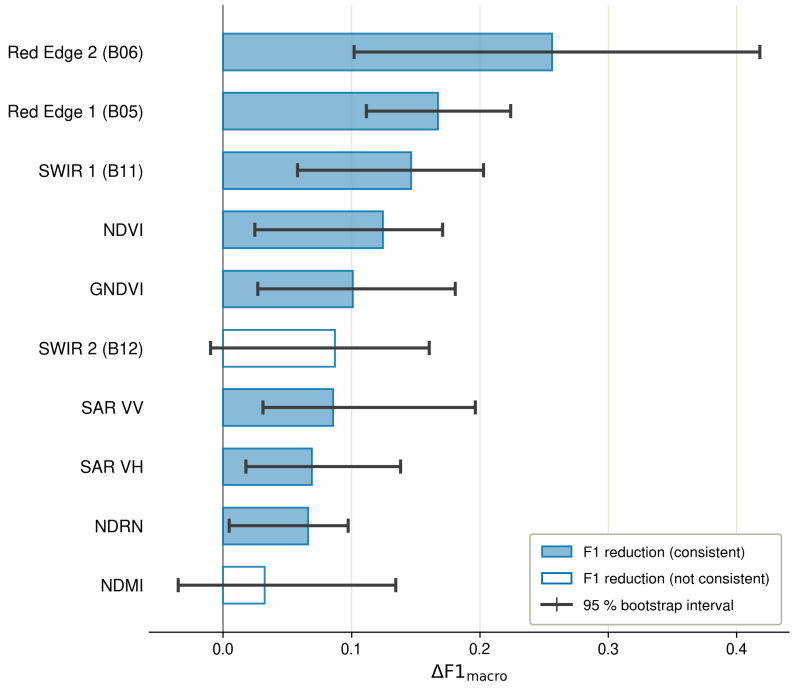


Figure 7. Importance of the ten input variables (spectral indices, Sentinel-2 bands and Sentinel-1 backscatter) for the best model, MiniRocket, trained on the 5YR multivariate time series ($F1_{macro}$ = 0.72). The ascending and descending orbits of each SAR polarization were treated as one feature here, since gaps in one orbit had been previously been filled from the other. The x-axis indicates the reduction ($\Delta F1_k$) in $F1_{macro}$ when the variable's time series on the y-axis is set to zero in the model evaluation. Larger values indicate greater variable importance. Bars show the average $\Delta F1_k$ while error bars show the 95 % interval of the bootstrapping.

5.1 Modeling performance and computational tradeoffs

We showed that the most performant predictors are multivariate time series of Sentinel-2 bands, vegetation indices, and Sentinel-1 polarizations.These predictors outperformed static geometry or single-date spectro-radiometric features at distinguishing wind, bark beetle, and defoliator disturbances. Within the multivariate predictor set, red-edge and shortwave-infrared bands, chlorophyll-sensitive indices, and the VV and VH polarizations emerged as the most informative features for separating disturbance agents (Fig. 7).

To contextualize our results, we compared omission errors of our best performing model with those by other studies (Tab. D1), namely Viana-Soto and Senf (EFDA; 2025), Du et al. (2023), Senf et al. (2015), and Schroeder et al. (2017). Our bark beetle omission error (9.3%) was lower than that reported by Senf et al. (2015) for bark beetle (12.9–15.0%). Omission errors were higher for defoliators (40.1%) and wind (33.2%) than those reported by Senf et al. (2015) for defoliators (11.3–28.9%) and by Du et al. (2023) for wind (10.0%). Unlike our spatial cross-validation, Senf et al. (2015) tested on data from the same study area as the training data, which tends to overestimate predictions. For defoliators, the higher omission

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

additionally reflects their low class prevalence in our dataset (9.5% of polygons), which class balancing could not resolve (Fig. C2). For wind, Du et al. (2023) used a different class setting in which wind is contrasted against fire and harvest rather than bark beetle. In a class setting closer to ours, including defoliators, wind, and bark beetles among other agents, Schroeder et al. (2017) report wind omission of 24–60%, consistent with our results.

We note, however, that comparing classification performances across disturbance agent classification studies is difficult, since they differ in disturbance class definitions, sensor and temporal setups, and validation designs which makes a direct comparison impossible. For example, most of the studies considered have grouped different insects into single classes, whereas we separated bark beetles from defoliators, which limits the comparability with our results. Validation of the model performance is another challenge, since some of the earlier studies (Senf et al., 2015; Schroeder et al., 2017) did not use spatially explicit cross-validation. Non-spatial validation typically overestimates model performance, as both our results (Fig. C1) and previous studies (Meyer and Pebesma, 2021) shows, so their reported accuracies are likely overly optimistic relative to our nested spatial cross-validation. Meaningful quantitative comparison of disturbance agent classifications would require a standardized spatial validation protocol of spatial cross validation and with an assessment of the area of applicability (AOA) (Nowosad et al., 2026; Meyer and Pebesma, 2021).

Beyond accuracy, model applicability also depends on computational cost and data requirements. Whether the higher computational workload of deep learning models translates into better forest disturbance classification, compared with conventional machine learning, remains unclear, with mixed evidence across the literature (Du et al., 2023; Kislov et al., 2021; Zheng, 2025; Aulló-Maestro et al., 2026). Our results show that more data or higher model complexity do not automatically translate into better performance. The transformation- and feature-based models (MiniRocket, TSFresh RF) consistently match or outperform the more complex LSTM architecture for most configurations. This suggests that the discriminative information in disturbance time series is better captured by random convolution patterns and explicit feature extraction than by learned recurrent representations. Our shuffling experiment points in the same direction, as permuting the temporal order of the observations did not, or only marginally, reduce performance (Tab. C12). Additionally it shows that computational efficiency and classification performance need not trade off against each other. MiniRocket requires no GPU (Dempster et al., 2021) and was 3 to 7 times faster than the LSTM on the multivariate AG time series (30 vs. 101 minutes for 1YR, 110 vs. 604 minutes for 8YR) and 12 to 44 times faster on the NDVI AG time series. The same is true for the data volume required for good performance. A single year of SAR and optical information (see Tab. 2) already achieves up to 95% of the performance of the best performing five-year window, at one fifth of the data volume. These results show that scalable, operationally deployable disturbance detection does not require the most complex model nor the richest data configuration.

5.2 Strengths and weaknesses of different predictors

Despite being used effectively in fire and anthropogenic disturbance mapping (Viana-Soto and Senf, 2025; Schleeweis et al., 2020; Senf and Seidl, 2022; Alonso et al., 2022; Schroeder et al., 2017; Marzo et al., 2022), the predictors of disturbance shape and geometry did not allow reliable classification of both insect and wind disturbances in our study (Tab. 1). A possible explanation is that the outline of biotic disturbances is shaped by host species (Grandpré et al., 2018) and forest health

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

(De Grandpré et al., 2022), while wind disturbances are shaped by wind direction and topography (Senf and Seidl, 2018). This natural diversity in patch shape is further amplified by landscape fragmentation (Sebald et al., 2021; Schwartz et al., 2017), artificial shapes from roads and agricultural land, and post disturbance management such as salvage logging, all of which cause a extremely heterogeneous disturbance geometry regardless of the underlying agent and limit differentiation by shape alone.

Previous work has shown that adding spectral information improves both single-date disturbance classification (Cohen et al., 2018, 2020) and time-series disturbance detection (Li et al., 2024). Our results support this, as spectral and radiometric information at a single timestamp (VI + SAR; SVM; $F1_{macro}$ = 0.65; Tab. 1) did not just outperform geometry but also all univariate NDVI time series ($F1_{macro} \leq 0.62$). While previous tree-level studies reliably separated healthy from attacked trees using single-date acquisitions (Minařík et al., 2020, 2021), these relied on much higher resolution imagery (6 cm/pixel UAV-borne multi-spectral data). Our results suggest that this holds at coarser resolution for bark beetle, but less so for disturbances of varying and often moderate severity, such as defoliator (see Tab. C10;C11). These results are consistent with Meddens et al. (2013), who found that single-date classification performed well only at high mortality levels, while multi-date classification was required at intermediate severity.

While a large number of spectral and SAR variables is useful, the best performance is achieved by combining spectral and SAR variables over multiple timestamps (Fig. 5b; Tab. 2). Overall, we find a trade-off between temporal breadth and the richness of spectral and radiometric information, where adding spectral and SAR information can compensate limited temporal information. SAR in particular performs well even with limited temporal coverage (see Fig. 5c): at one year, Sentinel-1 alone ($F1_{macro} = 0.65$) outperformed Sentinel-2 bands ($F1_{macro} = 0.59$) or vegetation indices alone ($F1_{macro} = 0.60$). Extending the multivariate time series from one to five years improved performance, but not necessarily because of the temporal order (see Sec. 4.3)

5.2.1 Importance of red-edge, short-wave infrared, chlorophyll-sensitive indices and SAR predictors

Red-edge bands, which are sensitive to chlorophyll content (Morley et al., 2020), showed the largest $\Delta F1_k$ in the importance analysis (B06: 0.26, B05: 0.17), followed by the SWIR band B11 (0.15), while the red-edge index NDRN contributed less (0.07; Fig. 7). Healthy, physiologically active leaves have high chlorophyll content, while stressed or damaged leaves lose chlorophyll, which shifts the red-edge slope (Velichkova and Krezhova, 2019). Bark beetle attacks disrupt the phloem, which progressively reduces photosynthetic activity and pigment concentration before the canopy visibly discolors (Petit et al., 2026). Since the red-edge position shifts systematically with chlorophyll concentration, it is suited to capture this gradual degradation, allowing physiological stress to be detected before visible browning occurs (Abdullah et al., 2019). Wind damage, in contrast, produces an immediate structural loss of photosynthetically active tissue rather than gradual pigment degradation, so the temporal course of these bands helps to separate bark beetle from wind disturbances. The greenness indices NDVI and GNDVI followed the red-edge and SWIR bands, with consistent $\Delta F1_k$ of 0.12 and 0.10 (Fig. 7). GNDVI, defined here after Gitelson et al. (1996), contrasts green with near-infrared reflectance and is therefore sensitive to chlorophyll content and leaf area (Gitelson et al., 1996, 2003). This is consistent with the canopy changes caused by the biotic agents. Bark beetle attacks lead to progressive needle discoloration, and defoliators remove leaf area, both of which reduce canopy greenness. Together with the red-edge

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

bands, both indices thus track the loss of chlorophyll and photosynthetically active tissue that accompanies bark beetle attack and defoliation.

Canopy water content, captured by shortwave-infrared and NDMI, is important, but not consistently across study area, to the classification performance. All three agents reduce canopy water content, but through different physiological pathways. Bark beetle infestation disrupts water transport as larval galleries girdle the cambium, producing a gradual multi-year decline (Mercado et al., 2021; Hubbard et al., 2013; Junttilaa et al., 2019). Short-term defoliation abruptly reduces leaf area, whereas long-lasting outbreaks cause a persistent and cumulative decline (Balducci et al., 2020) which are reflected as a loss of canopy and therefore also loss of canopy water content. Wind disturbances leave fallen or standing dead wood whose water content declines rapidly (Mitchell, 2013; Rajwa-Kuligiewicz and Bojarczuk, 2024). Because water loss eventually occurs for all agents, it may be more useful for detecting disturbances than for separating their agents.

The importance of SAR backscatter depends on the other information available to the model. Although SAR backscatter appears less important than optical bands such as the red-edge and SWIR bands in the importance analysis (Fig. 7), it still carries information that separates the agents. With a single year of data, SAR alone outperforms the Sentinel-2 bands and the vegetation indices while using fewer variables (Fig. 5c). These results are consistent with the established sensitivity of C-band backscatter to rapid changes in canopy structure and water-content following disturbance (Einzmann et al., 2017; Dalponte et al., 2020, 2023). Despite small in absolute terms (< 0.5 dB at the class mean) and heterogeneous across polygons (Fig. 4), SAR backscatter anomalies relative to the same months before the disturbance differ between agents in both magnitude and timing (Fig. D1; Tab. D2). Figure D1 shows a persistent VV increase for bark beetle during the disturbance year that is consistent with needle loss reducing canopy attenuation while the trunks remain standing. Wind disturbances show VV and VH positive anomalies already at the start of the disturbance year, consistent with the year–round nature of storms in southeastern USA, and the prolonged impacts of windthrow. Defoliators exhibit a distinct SAR response, which show a short-lived peak in late spring, consistent with insect phenology. These class-specific seasonal patterns in the SAR backscatter, provide complementary structural information to optical data and explain the good classification performance with SAR on a single year of data.

We note that while the importance of the remote sensing variables in our study correlates well with current process understanding, our analysis does not provide a causal link. Linking the remote sensing signal to the physiological responses to different disturbance effects would require ground-based measurement campaigns at disturbance sites, tracking variables such as canopy water content, needle or leaf loss, chlorophyll concentration, and tree mortality over time, and pairing these measurements with corresponding satellite observations.

5.3 Transferability to other regions

Our study focuses on the region covered by S1DM, namely the southeastern United States. This region is characterized by specific environmental conditions, including regional tree and insect species composition and meteorological drivers such as hurricanes, resulting in disturbance patterns that may be particular to this region (Müller et al., 2026). A recurring challenge in spatial machine learning for Earth observation is that model performance in one region does not necessarily transfer to other

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

regions, and this needs to be explicitly accounted for during evaluation (Nowosad et al., 2026). This raises the question of whether our results are transferable to other regions.

We addressed this challenge directly by applying spatial cross-validation and by deliberately excluding environmental predictors such as temperature or precipitation, which are themselves spatially structured and could otherwise allow the model to learn region-specific climate values rather than genuine disturbance signals. Doing so tests the model's ability to extrapolate to regions not seen during training, rather than relying on random cross-validation, which can substantially overestimate performance by allowing spatially correlated samples to appear in both training and test sets (Nowosad et al., 2026; Meyer and Pebesma, 2021). Our reported F1 score of 0.72 therefore reflects performance under this more conservative, spatially exclusive evaluation setting, which gives us a more honest picture of how performant the model actually is (Fig. C1).

Comparing the spatially exclusive setting to a random cross-validation setting, where neighboring disturbance events from the same region can appear in both the training and test sets, reveals substantial overestimation of 31% up to 75% relative to spatial cross-validation (Fig. C1). This gap points to substantial spatial autocorrelation in our dataset, suggesting that disturbance signatures for a given agent differ considerably between regions. For bark beetle outbreaks, this might reflect differences in insect species, host tree species, or local environmental conditions, while for wind disturbances, it could relate to differences in topographic exposure and terrain (Senf and Seidl, 2018). Identifying which of these factors drives the regional differences in the USA most strongly would require further investigation, but the size of the gap itself demonstrates why spatially exclusive validation is essential for obtaining a realistic estimate of model transferability.

The high regional variability in natural disturbances is substantial within the United States. One example would be the diversity of bark beetle species, as there are Mountain pine beetles that primarily affects lodgepole, ponderosa, and other pine species across the Western United States, while southern pine beetle predominantly affects southern pines in the southeastern region, though it has increasingly expanded into the Northeast in recent years. A similar pattern holds for defoliators with Spongy moth defoliating primarily oak-dominated hardwood forests in the Northeast and Midwest, whereas western spruce budworm instead targets Douglas fir and true fir stands across the Southwest, Intermountain, and Northern Rocky Mountain regions (Karel and Man, 2017; Stephens et al., 2023). Extrapolating our approach to the rest of the United States would therefore require training data spanning this full range of spatially distinct bark beetle and defoliator species and their respective host forests. This underscores that the spatial variability of natural disturbances must itself be represented in the training data, since a model can only be expected to generalize to disturbance types and host-species combinations it has actually seen during training.

Given that host-agent combinations vary systematically with region, we argue that adding climate variables such as temperature or precipitation as predictors, which are themselves strongly correlated with geography, would risk letting the model learn region-specific climate and forest characteristics rather than the disturbance-specific spectral and structural signals the classification should actually depend on. Instead, the training set should span a sufficiently diverse range of climates, topographies, and management regimes, allowing the underlying spectral and structural signal to be learned directly.

High-quality datasets available to support such analysis are limited. In the United States, datasets such as the Insect and Disease Survey (Forest Service U.S. Department of Agriculture, 2024), the Forest Inventory and Analysis program (Forest

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Service U.S. Department of Agriculture, 2023), the Hansen Global Forest Change dataset (Hansen et al., 2024), and NAFD (Schleeweis et al., 2020) exist, but these come with known spatial and temporal uncertainties (Eifler et al., 2026; Coleman et al., 2018). In Europe, spatially biased datasets such as FORWIND (Forzieri et al., 2020) and DEFID2 (Forzieri et al., 2023) are available, though new, more systematic ground truth datasets are in preparation (Viehweger et al., 2026, DISFOR). Availability of data on these disturbances, however, is heavily skewed towards the Northern Hemisphere. FAO (2020) reports that, between 2002 and 2016, insect disturbance data was available for over 85% of forested areas in North, Central America and Europe, but only 45% in Asia, while severe weather disturbance reports covered 50% - 85% of North America and European forests, but just 8% of Asia's forests. This gap persists even in recent global remote-sensing disturbance products that do cover the Southern Hemisphere the newly published Global Forest Disturbance dataset (GFD; Wang et al., 2026). GFD is a 30 m resolution map of 11 disturbance types for 2000 to 2020 that explicitly excludes biotic (insect and disease) and wind disturbances, citing a lack of adequate reference data at the global scale. Data specifically on wind disturbance in the Southern Hemisphere remains particularly rare; a notable exception is the Windthrow Inventory Database (WInD V.1 Urquiza-Muñoz et al., 2024a) for the Amazon basin, an openly available catalog of large-scale windthrow events from 1985 to 2020 (Urquiza-Muñoz et al., 2024b), though no comparable structured dataset appears to exist for biotic or abiotic disturbances elsewhere in the Southern Hemisphere. These disparities highlight the uneven spatial coverage and inconsistencies in the types of natural disturbances recorded in inventory data. Closing this gap will require dedicated investment in high-quality, field-validated disturbance datasets in currently underrepresented regions, since without such data, model transferability to these areas cannot be reliably achieved.

5.4 Transferability to other sensors

The importance of especially Red-Edge and, to a lesser extent, SWIR and C-Band SAR information for differentiating (rather than merely detecting) forest disturbances, has implications for which sensors are useful for classifying biotic and abiotic disturbances. For one, our results show that no single sensor is sufficient; instead, successful discrimination relies on combining complementary measurements of vegetation structure, water content, and pigmentation. In this context, the combination of Sentinel-1 and Sentinel-2 is particularly effective, as they provide C-Band, Red-Edge, and SWIR reflectance at high temporal and spatial resolution. While C-Band SAR provides the critical advantage of weather independence, it is not without challenges, as backscatter can be altered by freezing conditions, flooding, and complex terrain, potentially confounding disturbance-specific responses. These drawbacks need to be considered when applying this sensor to other regions.

Other sensors, such as Landsat and MODIS (Irons et al., 2012; Roy et al., 2014; Justice et al., 1998), have considerably longer archives, which would allow disturbance records to be extended much further back in time than the Sentinel constellation permits. While these sensors have proven effective for general disturbance mapping (Senf et al., 2015; Viana-Soto and Senf, 2025; Senf and Seidl, 2021), they exhibit limitations for the more nuanced differentiation addressed here. Both Landsat 8 and MODIS provide several of the predictors we identify as important, including NDMI, GNDVI, the SWIR bands, and NBR, but neither offers dedicated Red-Edge bands. Beyond this limitation, each sensor carries its own resolution trade-off. Landsat 8 has a considerably lower temporal resolution of approximately 16 days, while MODIS offers very high temporal resolution (1–2

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

days) but at coarse spatial resolution, ranging from 250 m for the red and near-infrared bands to 500 m for its other bands, including SWIR. This coarse resolution limits MODIS's applicability for stand-level analysis and leads to mixed pixels that obscure the structural and moisture contrasts between disturbance types. Differences in spatial, temporal and spectral resolution between the Sentinel constellation and these long-running sensors would therefore likely affect classification performance.

Our results show that spatial coarseness is not inherently detrimental to classification performance, as spatial aggregation improved predictions (Fig. 5a). However, this finding does not translate automatically into the conclusion that coarser satellites such as MODIS (daily revisit, 250 to 1000 m) could be applied to this problem without concern. While MODIS's high temporal resolution (daily revisit) would be expected to provide classification benefits, especially for defoliator disturbances, which have a short disturbance period and are the least well classified of the three agents (see Fig. 6). However, because disturbance patches are often small and spatially scattered (Fig. 4), a coarser pixel would frequently include healthy forest or other land cover alongside the disturbance signal, producing an already-mixed observation rather than a clean average over disturbance-specific pixels (Müller et al., 2026). For this reason, MODIS's spatial resolution (Justice et al., 1998) is likely too coarse to be applied to this classification problem, even though it may remain useful for coarser detection of disturbance timing.

To extend temporal coverage while maintaining high spatial and spectral resolution, fusion approaches combining Sentinel-2 and Landsat imagery are promising but not yet mature. Model-based methods have been shown to improve spatial resolution of Landsat to 10 m (Shao et al., 2019; Sigurdsson et al., 2022), and Landsat-based Red-Edge reflectance can be derived from Sentinel-2 observations through machine learning models (Chen and Zhang, 2023). However, both approaches are trained and validated on the overlapping period between the two sensors (since 2015/16) when concurrent Sentinel-2 and Landsat observations are available. The extension of these methods to the pre-2016 Landsat-only archive, remains untested and is so far only discussed as future potential in the studies. Moreover, since these reconstructed bands are statistical predictions learned from the overlap period rather than independent physical measurements, it remains unclear how much genuinely new information they might add, and whether they can reproduce the fine spectral and structural anomalies that agent-level disturbance classification depends on. While such approaches could create continuous, high-resolution time-series that include Red-Edge information, which is otherwise absent in Landsat, and would be highly valuable for differentiating forest disturbances at a global scale, this potential remains largely hypothetical for now.

Building on our finding that Red-Edge bands are among the most important predictors for disturbance differentiation, a more detailed examination of these specific spectral regions could deepen our understanding of the physiological processes driving disturbance-specific signatures. Emerging hyperspectral missions offer promising extensions to this framework. For example, EnMAP (Storch et al., 2023) provides continuous spectral information across the visible to shortwave infrared range, at spectral resolutions far finer than the discrete bands available from Sentinel-2 (Chabrillat et al., 2024), though at a coarser spatial resolution (30 m) and a substantially longer revisit interval (27 days). The higher spectral resolution enables more detailed retrieval of biochemical canopy properties such as water content, chlorophyll, and dry matter constituents including lignin and cellulose, since these compounds have distinct absorption features that broadband multispectral sensors cannot resolve (Tagliabue et al., 2022; Juola et al., 2024). This finer biochemical information could enhance the separability of disturbances that produce similar structural change but distinct physiological signatures, for example distinguishing gradual water stress

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

from bark beetle infestation from the more abrupt biochemical changes associated with wind damage. Similarly, the recently launched FLEX mission (Coppo et al., 2017) measures solar-induced chlorophyll fluorescence, a signal more directly linked to photosynthetic activity. While some studies report that SIF can detect physiological stress earlier than reflectance-based indices (Mohammed et al., 2019; Cannière et al., 2022), this advantage is not universal as under extreme heat stress, the SIF-photosynthesis coupling itself can break down or invert (Martini et al., 2022). Flying in tandem with Sentinel-3 at 300 m spatial resolution (Coppo et al., 2017), FLEX can not replace fine-scale structural or spectral monitoring, but might offer a complementary, functional early-warning signal for biotic disturbances such as bark beetle infestation, where physiological decline precedes visible canopy change by weeks to months, potentially extending the detection window beyond what red-edge or SWIR based indices alone currently provide.

Our study relies on Sentinel-1 C-band data (Torres et al., 2012), which offers a combination of spatial resolution, revisit frequency, and free global availability that no other SAR mission currently matches. Other radar wavelengths could nonetheless complement the classification presented here (Flores et al., 2019). X-band sensors such as TerraSAR-X and COSMO-SkyMed offer very high spatial resolution but penetrate only the upper canopy, and their data are commercial rather than open. P-band offers the deepest canopy penetration, but the only spaceborne P-band mission, BIOMASS (Quegan et al., 2019), is subject to coverage restrictions over Europe and North America and delivers annual, coarse-resolution products rather than dense time series. L-band therefore represents the most promising middle ground, penetrating further into the canopy than C-band while remaining compatible with frequent, moderate-resolution acquisition. NISAR (launched July 2025) provides free global L- and S-band observations at 3–48 m resolution with a 12-day repeat cycle (Kellogg et al., 2020), and the planned Copernicus ROSE-L mission (launch foreseen 2028) will add L-band capability at 5–10 m resolution with a three- to six-day revisit, explicitly targeting forest applications (Pierdicca et al., 2019). Whether the deeper canopy penetration of L-band improves agent-level classification remains to be tested, but these missions are currently too young to provide the multi-year archive that our approach requires.

5.5 Limitations and Challenges

Our results are based on the S1DM dataset, which provides temporally and spatially explicit disturbance information for wind, bark beetle, and defoliators in the southeastern USA from 2016–2021 (see Sec. 2 Müller et al., 2026). An important aspect to consider is that S1DM includes exclusively moderate to severe disturbances (severity above 30 %), due to the requirements of the trend detection algorithm (Müller et al., 2026; Cremer et al., 2020). Therefore, the classification skill obtained here (Tab. 2) might partly reflect a strict selection of high-impact disturbance events.

Furthermore, to obtain clear disturbance signatures, the S1DM focuses on single disturbance events that do not overlap with other disturbances within ± 5 years (Müller et al., 2026). In reality, these disturbances frequently co-occur in overlapping spatiotemporal patterns in natural ecosystems (Seidl et al., 2017). For instance, insects often target forests previously affected by drought or windstorms (Zhai et al., 2022; Anderegg et al., 2015). Disentangling these compound events is extremely challenging. The absence of compound events in our study, combined with their importance and changing patterns due to climate change (Dudney et al., 2025), makes it important to expand classification and disentangling studies targeting specifically these

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

compounding disturbances. Some of our architectures are in principle capable of identifying multiple, temporally distinct sub-patterns within a single time-series. MiniRocket applies a large bank of convolutional kernels of varying length and dilation, summarizing each response through pooling, which allows features occurring at different times and temporal scales within the series to be captured simultaneously (Dempster et al., 2020, 2021). Similarly, LSTM networks maintain a memory of past observations across the full sequence and could, in principle, encode a sequence of successive disturbance events (Hochreiter and Schmidhuber, 1997a). This contrasts with feature-based approaches such as TSFresh, which reduce the series to summary statistics like mean, variance, or slope (Christ et al., 2018b), and with shapelet-based methods such as STC, which search for the presence of individual discriminative subsequences (Hills et al., 2014), both of which are less suited to representing several overlapping causes at once. However, this representational capacity alone is not sufficient. Without labeled examples of compound events in the training data, the model has no supervision signal from which to learn such compound signatures, regardless of architecture. Expanding training data to include labeled compound disturbances therefore remains a prerequisite, and would be a natural next step for testing whether architectures like MiniRocket can meaningfully disentangle overlapping causes once given the opportunity to learn them.

Anthropogenic disturbances, such as harvesting and land use change, are explicitly excluded from the S1DM dataset (Müller et al., 2026). However, in managed forests, natural disturbances are often followed by salvage logging after wind storms or bark beetle outbreaks (van Eupen et al., 2025; Matisone et al., 2026), meaning that even disturbance events labeled as purely natural may carry a mixed signal, combining the original disturbance with subsequent human intervention. This risk is compounded by the fact that our study area itself is shaped by decades of forest management, visible in forest edges, fragmentation from agricultural land and roads, and settlement patterns. As a result, some of the temporal and spectral signatures our model learned as characteristic of a given natural disturbance type may already be partially confounded by post-disturbance management, which could limit the model's applicability to regions where salvage logging practices differ substantially from those in our study area. This problem could be mitigated with additional labels and model training to learn to distinguish between additional signatures of human disturbance.

Our dataset has a significant class imbalance, which results in consistently lower and more variable classification accuracy for defoliators (Fig. 6). This is consistent with known effects of class imbalance in machine learning, where minority classes tend to produce unstable decision boundaries (He and Garcia, 2009). We tested standard approaches for addressing class imbalance (He and Garcia, 2009), but none substantially improved classification (Fig. C2). This indicates that the difficulty does not stem from the class proportions alone, but from the limited information the defoliator samples contain. Resolving this issue would therefore require more and more diverse defoliator examples, rather than a reweighting of the existing ones.

Beyond class imbalance, two further factors, spatial and temporal resolution of our satellite products, could plausibly limit correct classification, especially of defoliators. Our results oppose the notion that spatial resolution is to coarse, as pixel-level (AP) classification performs worse than polygon-mean (AG) classification (Fig. 5a). This argues against insufficient spatial resolution, since finer-grained data should improve, not worsen, performance. Instead, coarse temporal resolution remains a more plausible, but currently untested, limitation. Our weekly composites of Sentinel-1 and Sentinel-2, combined with temporal smoothing during preprocessing, may suppress short-duration defoliation signals. This is precarious since defoliation

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

events can unfold and partially recover within only a few weeks, particularly where leaf reflush begins later in the same growing season. Testing this would require finer temporal resolution at similar or higher spatial resolution to that of Sentinel-2. Commercial sensors such as PlanetScope (Frazier and Hemingway, 2021) offer near-daily revisit at 3 to 5 m resolution, but are limited to RGB and near-infrared bands and require a commercial license outside of Planet's limited free-access programs. To reliably classify defoliators, a larger and more balanced training sample, or finer temporal resolution capable of capturing their fast dynamics, may be necessary.

6 Conclusion

This study evaluated the relative discriminatory power of geometric, spectral-radiometric, and temporal information for classifying three disturbance types that are particularly difficult to disentangle: wind, bark beetle, and defoliator disturbances. We find that multi-temporal observations of multispectral and SAR backscatter variables provide the strongest separability, while geometry alone, often used in classification studies, is largely insufficient. Notably, simpler, computationally efficient models such as MiniRocket match or outperform deep learning architectures across nearly all configurations, at a fraction of the computational cost, indicating that distinguishing these disturbance types does not require the most complex available modeling approach.

Feature importance analyses indicate that disturbance-specific spectral and structural responses drive this performance. Greenness- and chlorophyll-sensitive predictors, in particular the red-edge bands, contributed the most to distinguish between the three disturbance types, while water-sensitive SWIR information was largely redundant. SAR backscatter (VV and VH polarizations) reveals distinct structural responses for bark beetle and wind, whereas defoliators show only weak backscatter changes, consistent with their more subtle structural impact. Observation windows centered on the disturbance year were most effective, and spectral or radiometric richness in the predictor set can partly compensate for longer observation windows, with a single year of multivariate data reaching about 94 % of the performance of the best five-year window.

Several open questions remain. The low prevalence of defoliators and the episodic nature of defoliation remain unresolved in this study, as class balancing did not improve their classification. Larger, purpose-built training datasets or finer temporal compositing are likely required to determine whether defoliators are fundamentally harder to classify or simply underrepresented in current data. The scarcity of ground-truth disturbance training data in regions such as Asia represents a genuine barrier to global transferability, and addressing it is a prerequisite for extending this approach beyond well-monitored regions like the United States and Europe. Finally, emerging hyperspectral missions, such as EnMAP and FLEX, could extend this framework further by capturing physiological stress before visible structural change occurs, potentially widening the detection window for slow-onset biotic disturbances.

Code availability. The code used to generate the results of this study will be made publicly available on reviewer request or after publication on GitHub.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Data availability. The Sentinel-1 Disturbance Map (S1DM) is publicly available on Zenodo Müller et al. (2026). Sentinel-1 and Sentinel-2 data are publicly available through the Copernicus Open Access Hub and associated Copernicus data services.

Author contributions. AB and FM designed the study. FM conducted the data management, preprocessing, model development, and coding, with scientific and technical input from AB and GCV. FM wrote the first draft of the manuscript. All authors contributed to the interpretation of the results and the final manuscript.

Competing interests. The authors declare that they have no competing interests.

Disclaimer. The views expressed in this article are those of the authors and do not necessarily reflect the views of their affiliated institutions or funding agencies.

Acknowledgements. FM and AB acknowledge funding from the European Research Council (ERC StG, ForExD, grant agreement No. 101039567). FM and AB thank support from the CALIPSO (Carbon Loss in Plant Soils and Oceans) project, funded through the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program. GC-V thanks funding from the European Research Council (ERC) under the ERC Synergy Grant USMILE (grant agreement 855187) and the H2020 project XAIDA (grant agreement 101003469). This work uses Copernicus Sentinel data 2016-2022. The views expressed are purely those of the writers and may not in any circumstances be regarded as stating an official position of the European Commission.
FM thanks Dr. Sebastian Wieneke for his advice on connecting spectral bands to plant physiology, and Eleanor Butler for her support with mathematical descriptions.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Appendix A: Data

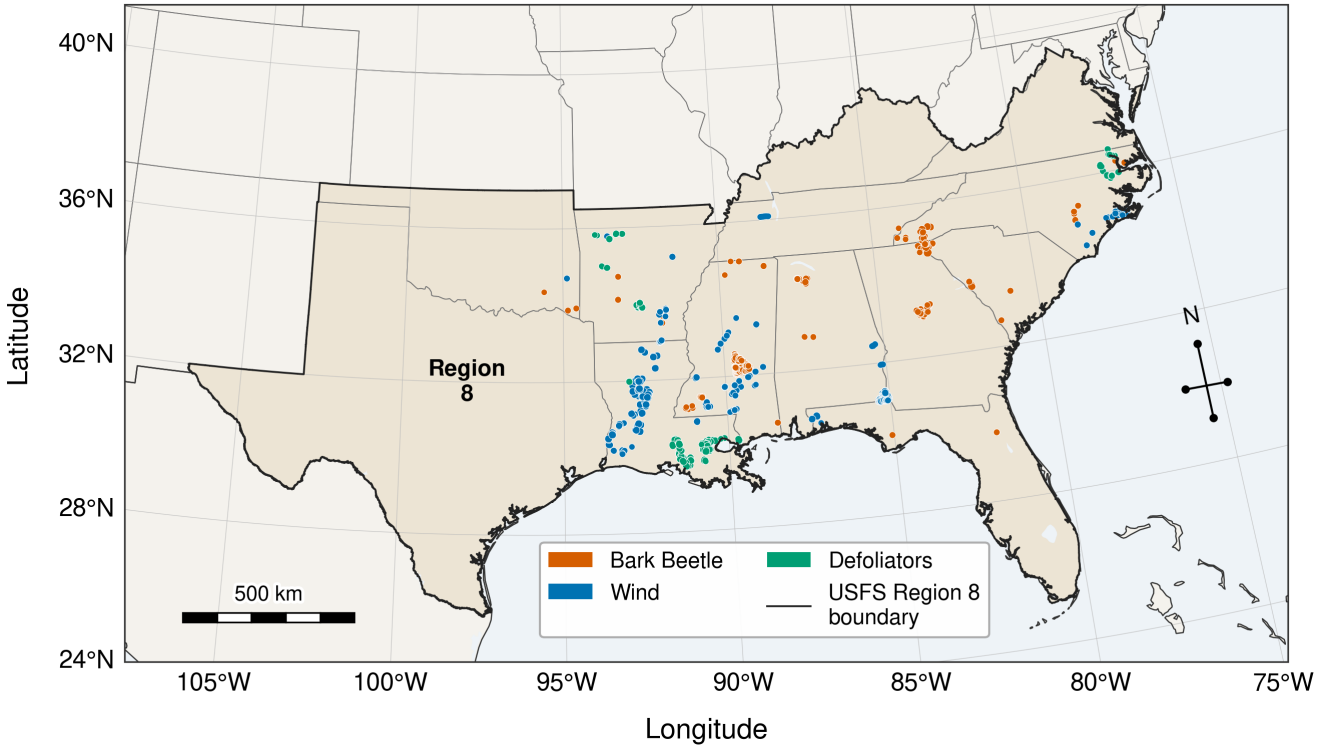


Figure A1. Study sites and spatial distribution of the S1DM dataset within US Forest Service Region 8 in the southeastern United States. Disturbance polygons are colored by disturbance agent (bark beetle, defoliators, wind) and the black outline marks the Region 8 boundary. Disturbances span 30°N–37°N and 96°W–76°W, from eastern Texas to the North Carolina coast.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

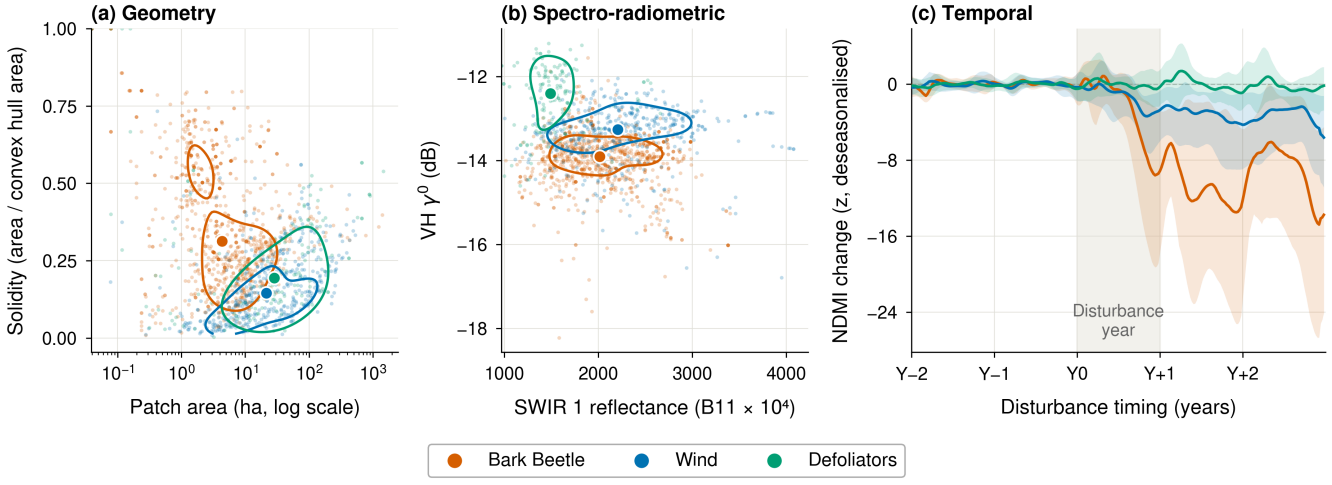


Figure A2. Overview of value distributions of the geometric, spectral and temporal signatures of the three disturbance agents bark beetles (orange), defoliators (green) and wind (blue). Panel (a) shows patch area (log scale) against solidity, and panel (b) the polygon mean SWIR 1 reflectance (B11) against Sentinel-1 VH backscatter (dB) at the single-date snapshot. In both panels, small dots represent individual polygons, the contour line encloses the densest 50 % of polygons of each agent, and the large dot marks the median across all polygons of that agent. Panel (c) shows NDMI over the 5YR window around the S1DM detection year (Y0, grey shading). Each polygon's NDMI is expressed relative to its own pre-disturbance seasonal cycle and in units of its pre-disturbance variability (z-score). The dashed line ($z = 0$) indicates no deviation from the pre-disturbance years. Lines show the weekly median per agent, and the shading shows the interquartile range.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table A1. Vegetation indices that are derived from Sentinel-2 bands and used as predictor variables in this study. These indices capture different vegetation properties such as greenness, moisture, brightness, and disturbance signals, relevant for assessing forest health, stress, and disturbance events. The index formulas and references were checked against Montero et al. (2023).

No.	Abb.	Name	Formula (Sentinel-2 bands)	Aspect	Reference
1	NDVI	Normalized Difference Vegetation Index	$\frac{B8 - B4}{B8 + B4}$	Vegetation greenness / vigor	Rouse et al. (1973)
2	NBR	Normalized Burn Ratio	$\frac{B8 - B12}{B8 + B12}$	Disturbance / burn severity	Coffelt and Livingston (2002); Key and Benson (2006)
3	NDMI	Normalized Difference Moisture Index	$\frac{B8 - B11}{B8 + B11}$	Moisture / stress	Wilson and Sader (2002)
4	NDREI	Normalized Difference Red Edge Index	$\frac{B8 - B5}{B8 + B5}$	Chlorophyll / stress	Gitelson and Merzlyak (1994); Barnes et al. (2000)
5	GNDVI	Green Normalized Difference Vegetation Index	$\frac{B8 - B3}{B8 + B3}$	Canopy chlorophyll content (greenness)	Gitelson et al. (1996)
6	NIRv	Near-Infrared Reflectance of Vegetation	$B8 \times \frac{B8 - B4}{B8 + B4}$	Greenness / productivity	Badgley et al. (2017)
7	TCB[a]	Tasseled Cap Brightness	$0.3037\,B2 + 0.2793\,B3 + 0.4743\,B4 + 0.5585\,B8 + 0.5082\,B11 + 0.1863\,B12$	Brightness / soil	Crist and Cicone (1984)
8	TCG[a]	Tasseled Cap Greenness	$-0.2848\,B2 - 0.2435\,B3 - 0.5436\,B4 + 0.7243\,B8 + 0.0840\,B11 - 0.1800\,B12$	Greenness / vegetation	Crist and Cicone (1984)
9	TCW[a]	Tasseled Cap Wetness	$0.1509\,B2 + 0.1973\,B3 + 0.3279\,B4 + 0.3406\,B8 - 0.7112\,B11 - 0.4572\,B12$	Moisture / stress	Crist and Cicone (1984)
10	DRN[b]	Distance Red & NIR	$\sqrt{B4^2 + B8^2}$	Brightness / canopy reflectance	adapted from Huo et al. (2021)
11	NDRN[b]	Normalized Distance Red & NIR	$\frac{\mathrm{DRN} - \mathrm{DRN}_{\min}}{\mathrm{DRN}_{\max} - \mathrm{DRN}_{\min}}$	Brightness / canopy reflectance	adapted from Huo et al. (2021)
12	kNDVI[c]	Kernel NDVI	$\tanh\left(\left(\frac{B8 - B4}{2\sigma}\right)^2\right)$	Greenness / vigor	Camps-Valls et al. (2021)

[a] Landsat TM coefficients (digital numbers) applied to Sentinel-2 reflectance ($\times 10^4$); values are indicative, not physically calibrated.

[b] Adapted from DRS/NDRS of Huo et al. (2021), with NIR (B8) instead of SWIR (B12); min/max per Sentinel-2 minicube over all pixels and dates.

[c] $\sigma = 0.5\,(B8 + B4)$ per pixel (Camps-Valls et al., 2021), i.e. kNDVI $= \tanh(\mathrm{NDVI}^2)$.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Appendix B: Methods

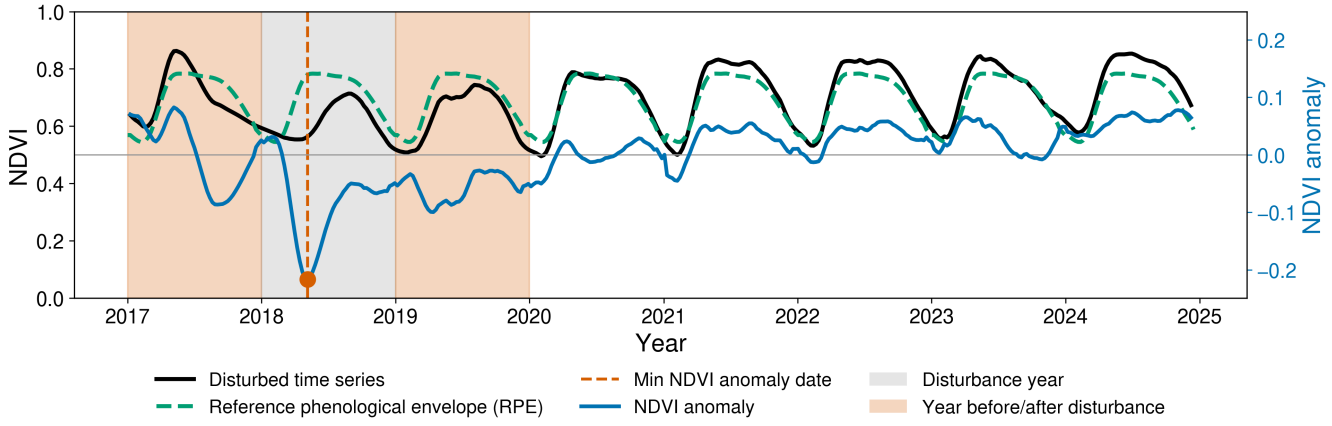


Figure B1. Illustration of the reference phenological envelope (RPE), which is used to identify the time of the strongest disturbance impact, following an approach similar to Gouveia et al. (2010). The black line is the observed NDVI and the green dotted line shows the RPE, defined as the 90th percentile of the weekly NDVI values over the 8 years. The NDVI anomaly (blue line) is calculated as the difference between the observed NDVI and the RPE. A one-year window around the disturbance year is highlighted in orange. The strongest negative NDVI anomaly within this window is the timestamp of the most severe disturbance (red dot).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table B1. Geometric features calculated from the forest disturbance multipolygons in the S1DM dataset. For each feature, the table gives its name and a description of the geometric property it captures, such as size, shape, elongation, compactness or internal complexity.

Feature	Description
Area	Sum of the areas of all polygon parts, the overall size of the disturbance (m^2).
Perimeter	Combined length of all polygon boundaries (m).
Polsby–Popper index	$4\pi \cdot \text{area}/\text{perimeter}^2$, how closely the shape approaches a circle (1 for a circle, towards 0 for irregular or elongated shapes).
Bounding box width	East–west extent of the axis-aligned bounding box (m).
Bounding box height	North–south extent of the axis-aligned bounding box (m).
Extent ratio	Area divided by the area of the bounding box. Metric of how completely the disturbance fills its bounding box.
Elongation	Ratio of the long to the short side of the minimum rotated rectangle, an orientation-independent measure of elongation (≥ 1).
Convex hull area	Area of the smallest convex polygon enclosing all polygon parts, the overall footprint and spatial dispersion of the disturbance (m^2).
Solidity	Area divided by the convex hull area, the proportion of the hull actually occupied by the disturbance.
Convexity	Convex hull perimeter divided by the polygon perimeter, boundary irregularity (1 for a convex shape, lower for more irregular boundaries).
Number of holes	Number of interior gaps within the multipolygon, internal fragmentation.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table B2. Feature-based classification models used for the geometric and spectro-radiometric predictor sets. The table summarizes the six machine-learning models evaluated on non-temporal predictors, including their name, methodological family, core characteristics and the motivation for their inclusion. They range from simple, interpretable baselines to strong conventional classifiers and are used to assess the discriminative power of the static predictor sets.

No.	Name	Family	Description	Rationale
1	Logistic Regression (LR)	Linear model	Learns a linear decision boundary and estimates class probabilities with the logistic (sigmoid) function (Hosmer et al., 2013).	Simple, interpretable baseline, common in ecological modelling
2	Support Vector Machine (SVM)	Margin-based	Finds the separating hyperplane (linear or kernelised) that maximises the margin between classes (Cortes and Vapnik, 1995).	Robust in high-dimensional feature spaces and effective with small datasets
3	K-Nearest Neighbors (KNN)	Distance-based	Assigns each sample the majority class of its k nearest training samples under a distance metric (Cover and Hart, 1967).	Captures local structure without parametric assumptions
4	Random Forest (RF)	Tree-based ensemble	Trains many decision trees on bootstrap samples and random feature subsets and aggregates their votes (Breiman, 2001a).	Robust to non-linear interactions and heterogeneous features, widely used in remote sensing
5	Extreme Gradient Boosting (XGBoost)	Tree-based, gradient boosting	Builds an ensemble of decision trees sequentially, each fitted to the gradient of the loss of the previous ensemble (Chen and Guestrin, 2016).	High predictive performance, widely adopted for tabular data
6	Multilayer Perceptron (MLP)	Neural network	Learns non-linear relationships with one or more fully connected hidden layers trained by backpropagation (Rumelhart et al., 1986).	Evaluates the benefit of a non-linear neural network on the static predictors

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table B3. Time-series classification models that are used to predict on the univariate and multivariate temporal predictor sets. The table summarizes the six models evaluated on the time series, including their methodological family, core characteristics and the motivation for their inclusion. They cover the main families of time-series classification, from distance- and dictionary-based methods to feature extraction, shapelets, convolutional kernels and recurrent neural networks.

No.	Name	Family	Description	Rationale
1	Bag-of-SFA-Symbols (BOSS)	Dictionary-based	Converts sliding windows of the time series into symbolic words with the Symbolic Fourier Approximation (SFA) and classifies the histograms of word counts; an ensemble of such classifiers is used (Schäfer, 2014).	Captures recurring temporal patterns and is robust to noise and small distortions
2	K-Nearest Neighbors with Dynamic Time Warping (KNN-DTW)	Distance-based	Assigns each time series the majority class of its k nearest training series, with similarity measured by Dynamic Time Warping, which aligns sequences that are shifted or stretched in time (Ding et al., 2008).	Strong standard baseline for time-series classification that tolerates phase shifts common in ecological time series
3	TSFresh features + Random Forest (TSFresh RF)	Feature-based	Summarises each variable with the minimal TSFresh set of statistical features (e.g. mean, variance, extremes) and classifies them with a Random Forest (Christ et al., 2018a).	Transforms the time series into interpretable summary statistics and combines them with a robust tree-based classifier
4	Shapelet Transform Classifier (STC)	Shapelet-based	Searches the training data for short subsequences (shapelets) that best separate the classes, describes each series by its distance to these shapelets and classifies them with a rotation forest (Hills et al., 2014).	Yields interpretable temporal patterns that can be linked to disturbance events
5	Long Short-Term Memory (LSTM)	Recurrent neural network	Bidirectional recurrent network whose gated memory cells learn long-term and non-linear temporal dependencies directly from the raw time series (Hochreiter and Schmidhuber, 1997b; Ismail Fawaz et al., 2019).	Represents deep learning approaches that model temporal dependencies without handcrafted features
6	MiniRocket	Convolutional kernel transform	Transforms the time series with a large set of fixed random convolutional kernels into features that are classified by a ridge regression classifier (Dempster et al., 2020, 2021).	Computationally efficient and among the most accurate time-series classifiers in benchmark comparisons

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table B4. The parameters of each model, that can be adjusted during hyperparameter tuning with Optuna in the inner cross-validation. Categorical settings with a single choice are fixed and not counted here.

Model	No. hyperparameters	Tuned hyperparameters
Static models		
LR	3	`C`, `solver`, `max_iter`
SVM	3	`C`, `kernel`, `gamma`
KNN	3	`n_neighbors`, `weights`, `algorithm`
RF	6	`n_estimators`, `max_depth`, `max_features`, `min_samples_split`, `min_samples_leaf`, `bootstrap`
XGB	5	`n_estimators`, `learning_rate`, `max_depth`, `subsample`, `colsample_bytree`
MLP	6	`hidden_layer_sizes`, `activation`, `solver`, `alpha`, `learning_rate_init`, `max_iter`
Time-series models		
BOSS	7	`threshold`, `max_ensemble_size`, `max_win_len_prop`, `min_window`, `alphabet_size`, `use_boss_distance`, `feature_selection`
KNN-DTW	3	`n_neighbors`, `weights`, `leaf_size`
TSFreshRF	4	`n_estimators`, `max_depth`, `min_samples_split`, `min_samples_leaf`
Shapelet	7	`n_shapelet_samples`, `max_shapelets`, `max_shapelet_length`, `transform_limit_in_minutes`, `time_limit_in_minutes`, `contract_max_n_shapelet_samples`, `save_transformed_data`
LSTM	7–8[a]	`hidden_dim`, `num_layers`, `dropout`, `learning_rate`, `batch_size`, `epochs`, `patience`, `max_chunk_len`[a]
MiniRocket	3	`num_kernels`, `max_dilations_per_kernel`, `n_features_per_kernel`

[a] `max_chunk_len` tuned for the 5Y and 8Y windows only; ranges of `batch_size`, `epochs` and `patience` scaled to the window length.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table B5. Evaluation metrics that are used to assess the classification performance of the models, with their formulas and an explanation of what each metric measures. Precision and recall are reported as macro averages over the $C = 3$ disturbance classes and per-class values are used to compute the F1 score and the omission and commission errors. Symbols are defined below the table.

Metric	Formula	Explanation
Overall accuracy (OA)	$\mathrm{OA} = \frac{1}{N}\sum_{c=1}^{C} n_{cc}$	Measures the proportion of correctly classified samples (diagonal of the confusion matrix) out of all samples. Provides an overall sense of model correctness, but can be biased if classes are imbalanced.
Precision	$\mathrm{Precision} = \frac{1}{C}\sum_{c=1}^{C} \frac{TP_c}{TP_c + FP_c}$	Quantifies the fraction of predictions of a class that are correct ($\mathrm{Precision}_c$, user's accuracy), averaged over all classes. High precision indicates that the model makes few false positives.
Recall	$\mathrm{Recall} = \frac{1}{C}\sum_{c=1}^{C} \frac{TP_c}{TP_c + FN_c}$	Measures the proportion of actual samples of a class that are correctly identified (Recall_c, producer's accuracy), averaged over all classes. High recall means the model misses very few positive cases.
F1-score (class c)	$\mathrm{F1}_c = 2 \cdot \frac{\mathrm{Precision}_c \cdot \mathrm{Recall}_c}{\mathrm{Precision}_c + \mathrm{Recall}_c}$	Harmonic mean of precision and recall for class c; balances false positives (FP_c) and false negatives (FN_c).
$\mathrm{F1}_{\mathrm{macro}}$	$\mathrm{F1}_{\mathrm{macro}} = \frac{1}{C}\sum_{c=1}^{C} \mathrm{F1}_c$	Unweighted mean of the per-class F1 scores ($C = 3$: wind, bark beetle, defoliator). Weighting all classes equally makes it sensitive to the rare defoliator class.
Cohen's κ	$\kappa = \frac{\mathrm{OA} - p_e}{1 - p_e}, \quad p_e = \frac{1}{N^2}\sum_c n_{c\cdot}\, n_{\cdot c}$	Agreement between predicted and reference classes corrected for the agreement expected by chance (p_e), computed from the row ($n_{c\cdot}$) and column ($n_{\cdot c}$) totals of the confusion matrix. $\kappa = 0$ corresponds to chance-level agreement, $\kappa = 1$ to perfect agreement.
Omission error (class c)	$\mathrm{OE}_c = 1 - \mathrm{Recall}_c = \frac{FN_c}{TP_c + FN_c}$	Fraction of reference samples of class c that the model assigned to another class (missed disturbances of agent c).
Commission error (class c)	$\mathrm{CE}_c = 1 - \mathrm{Precision}_c = \frac{FP_c}{TP_c + FP_c}$	Fraction of samples predicted as class c that belong to another class (false attributions to agent c).

TP_c, FP_c, FN_c: true positives, false positives and false negatives per class c;
$\mathrm{Precision}_c$, Recall_c: precision and recall of class c;
n_{cc}: diagonal entries of the confusion matrix;
$n_{c\cdot}$, $n_{\cdot c}$: row and column totals;
N: number of samples;
C: number of classes

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table B6. An overview of the qualitative classification performance levels defined in this study for the imbalanced three-class problem (bark beetle 60 %, wind 31 %, defoliators 9 %), to help readers understand how the F1 scores are to be interpreted. The levels range from no learning and random guessing, which is defined as values around the majority-class baseline (always predicting bark beetle, $F1_{macro}$ = 0.25), the uniform random baseline (each class predicted with equal probability, $F1_{macro}$ = 0.30) and the stratified random baseline (classes predicted in proportion to their frequency, $F1_{macro}$ = 0.33), through slight and moderate learning up to excellent performance for $F1_{macro}$ scores above 0.90.

Level	$F1_{macro}$	**Interpretation**
No learning	≤ 0.25	At or below the majority-class baseline; no better than always predicting bark beetle
Random guessing	$0.25 - 0.33$	Between the majority-class and the random baselines; no meaningful discriminative signal
Slight learning	$0.33 - 0.50$	Above chance, but one or two classes near random; defoliator classification likely failing
Moderate	$0.50 - 0.70$	Model distinguishes disturbance types but with substantial errors
Good	$0.70 - 0.80$	All three disturbance types clearly above chance, with remaining errors mainly in the minority class
Very good	$0.80 - 0.90$	Strong performance with all classes well separated
Excellent	> 0.90	Near-ceiling performance; in this study only reached under random cross-validation

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Appendix C: Results

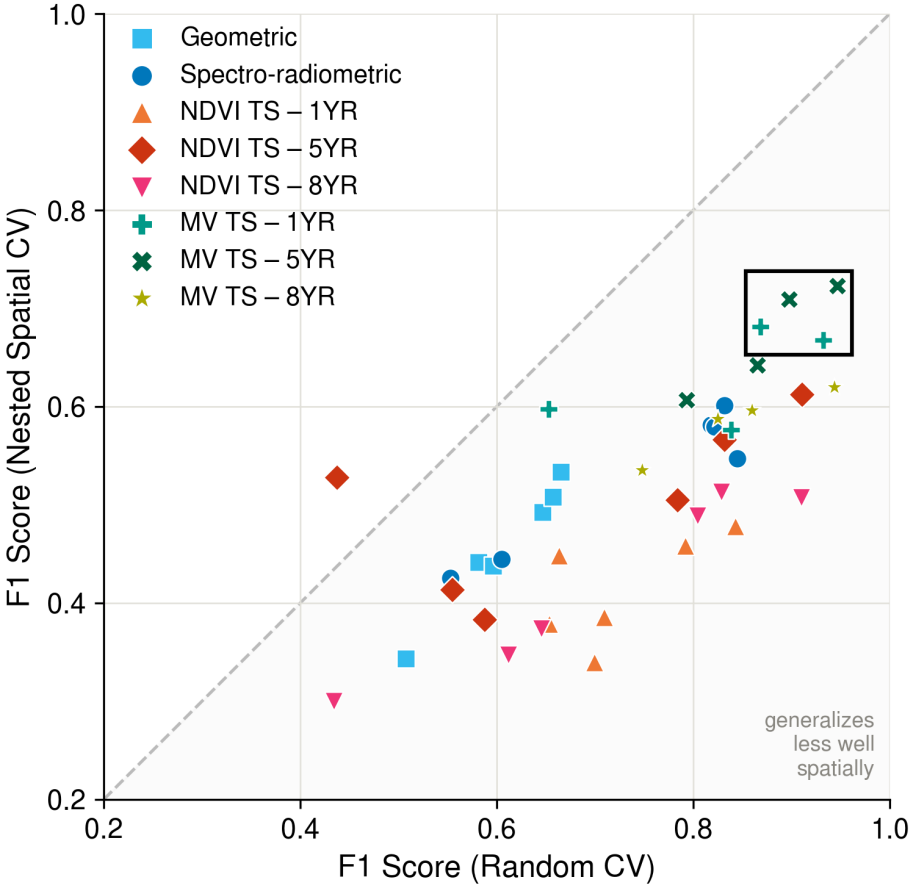


Figure C1. Comparison of the models' performance under random cross-validation (x-axis) versus nested spatial cross-validation (y-axis). All predictor sets, the geometric (blue squares) and spectro-radiometric (blue dots) predictor sets as well as all univariate and multivariate time series over all window lengths (1YR, 5YR and 8YR), and all tested models are evaluated. Points below the diagonal indicate that random cross-validation overestimates the actual performance of the models because, differently than in spatial cross-validation, spatially auto-correlated samples end up in both the training and the test data.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table C1. Macro precision [0–1] for all six machine learning models evaluated on the S1DM dataset across different static predictor sets for the Aggregated (AG) setting. Models include the Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP). Columns are ordered from linear to increasingly nonlinear and complex models. Values are the mean (± standard deviation) across 20 repetitions of nested spatial cross-validation. The feature column indicates the number of predictors used in each dataset. The highest value for each predictor set is highlighted in bold.

Predictors	No. of features	Models					
		LR	SVM	KNN	RF	XGB	MLP
Geometric predictors							
Geometry	11	0.53 ± 0.12	0.57 ± 0.11	0.53 ± 0.14	0.49 ± 0.09	0.57 ± 0.10	$\mathbf{0.61 \pm 0.11}$
SAR predictors (Sentinel-1)							
VV/VH	4	0.49 ± 0.14	0.55 ± 0.08	0.58 ± 0.10	0.59 ± 0.10	$\mathbf{0.60 \pm 0.09}$	0.57 ± 0.11
Optical predictors (Sentinel-2)							
S2 bands	10	$\mathbf{0.58 \pm 0.05}$	0.57 ± 0.06	0.49 ± 0.07	0.57 ± 0.06	0.57 ± 0.05	0.55 ± 0.07
Vegetation indices	12	0.65 ± 0.05	$\mathbf{0.66 \pm 0.05}$	0.49 ± 0.06	0.55 ± 0.06	0.54 ± 0.07	0.59 ± 0.04
Sensor combinations							
Combination subset[a]	12	0.63 ± 0.04	0.60 ± 0.10	0.62 ± 0.12	0.63 ± 0.06	0.63 ± 0.04	$\mathbf{0.64 \pm 0.04}$
VV/VH + S2 bands	14	0.58 ± 0.04	0.56 ± 0.06	0.60 ± 0.11	$\mathbf{0.65 \pm 0.06}$	0.63 ± 0.06	0.62 ± 0.08
VV/VH + Vegetation indices	16	0.66 ± 0.01	$\mathbf{0.68 \pm 0.03}$	0.63 ± 0.10	0.65 ± 0.07	0.64 ± 0.04	0.66 ± 0.05
VV/VH + S2 bands + Vegetation indices	26	0.65 ± 0.01	0.63 ± 0.03	0.64 ± 0.12	$\mathbf{0.66 \pm 0.05}$	0.64 ± 0.05	0.64 ± 0.09

[a] Features included: NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table C2. Macro recall [0–1] for six machine learning models evaluated on the S1DM dataset across different static predictor sets for the Aggregated (AG) setting. Models include Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP). Columns are ordered from linear to increasingly nonlinear and complex models. Values are the mean (± standard deviation) across 20 repetitions of nested spatial cross-validation. The feature column indicates the number of predictors used in each dataset. The highest value for each predictor set is highlighted in bold.

Predictors	No. of features	Models					
		LR	SVM	KNN	RF	XGB	MLP
Geometric predictors							
Geometry	11	0.49 ± 0.07	0.50 ± 0.07	0.48 ± 0.07	0.50 ± 0.06	0.52 ± 0.07	$\mathbf{0.57 \pm 0.08}$
SAR predictors (Sentinel-1)							
VV/VH	4	0.53 ± 0.08	0.53 ± 0.07	$\mathbf{0.56 \pm 0.08}$	$\mathbf{0.56 \pm 0.07}$	$\mathbf{0.56 \pm 0.08}$	0.55 ± 0.08
Optical predictors (Sentinel-2)							
S2 bands	10	0.54 ± 0.07	$\mathbf{0.57 \pm 0.06}$	0.48 ± 0.08	0.54 ± 0.07	0.54 ± 0.06	$\mathbf{0.57 \pm 0.07}$
Vegetation indices	12	0.62 ± 0.09	$\mathbf{0.63 \pm 0.10}$	0.47 ± 0.07	0.52 ± 0.09	0.52 ± 0.09	0.58 ± 0.09
Sensor combinations							
Combination subset[a]	12	$\mathbf{0.64 \pm 0.11}$	0.62 ± 0.12	0.59 ± 0.09	0.61 ± 0.11	0.62 ± 0.11	$\mathbf{0.64 \pm 0.10}$
VV/VH + S2 bands	14	0.59 ± 0.10	0.55 ± 0.09	0.56 ± 0.09	0.62 ± 0.11	0.62 ± 0.11	$\mathbf{0.63 \pm 0.11}$
VV/VH + Vegetation indices	16	0.64 ± 0.09	$\mathbf{0.67 \pm 0.09}$	0.58 ± 0.10	0.62 ± 0.10	0.62 ± 0.09	0.66 ± 0.11
VV/VH + S2 bands + Vegetation indices	26	$\mathbf{0.64 \pm 0.10}$	0.62 ± 0.09	0.58 ± 0.10	0.62 ± 0.11	0.63 ± 0.11	$\mathbf{0.64 \pm 0.14}$

[a] Features included: NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table C3. Overall accuracy [0–1] for six machine learning models evaluated on the S1DM dataset across different static predictor sets for the Aggregated (AG) setting. Models include Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP). Columns are ordered from linear to increasingly nonlinear and complex models. Values are the mean (± standard deviation) across 20 repetitions of nested spatial cross-validation. The feature column indicates the number of predictors used in each dataset. The highest value for each predictor set is highlighted in bold.

Predictors	No. of features	Models					
		LR	SVM	KNN	RF	XGB	MLP
Geometric predictors							
Geometry	11	$\mathbf{0.72 \pm 0.13}$	0.71 ± 0.13	0.66 ± 0.16	0.69 ± 0.13	0.70 ± 0.16	$\mathbf{0.72 \pm 0.13}$
SAR predictors (Sentinel-1)							
VV/VH	4	0.58 ± 0.16	0.59 ± 0.17	0.57 ± 0.16	$\mathbf{0.60 \pm 0.18}$	0.59 ± 0.16	0.59 ± 0.18
Optical predictors (Sentinel-2)							
S2 bands	10	0.64 ± 0.14	$\mathbf{0.66 \pm 0.10}$	0.55 ± 0.13	0.61 ± 0.12	0.61 ± 0.12	0.63 ± 0.09
Vegetation indices	12	$\mathbf{0.71 \pm 0.12}$	$\mathbf{0.71 \pm 0.12}$	0.54 ± 0.12	0.59 ± 0.15	0.59 ± 0.16	0.65 ± 0.14
Sensor combinations							
Combination subset[a]	12	$\mathbf{0.68 \pm 0.10}$	0.63 ± 0.18	0.64 ± 0.19	0.63 ± 0.16	0.63 ± 0.15	0.66 ± 0.13
VV/VH + S2 bands	14	$\mathbf{0.66 \pm 0.11}$	0.63 ± 0.15	0.60 ± 0.17	0.64 ± 0.15	0.63 ± 0.14	$\mathbf{0.66 \pm 0.15}$
VV/VH + Vegetation indices	16	0.73 ± 0.05	$\mathbf{0.77 \pm 0.04}$	0.62 ± 0.18	0.65 ± 0.16	0.65 ± 0.13	0.69 ± 0.15
VV/VH + S2 bands + Vegetation indices	26	$\mathbf{0.69 \pm 0.09}$	0.68 ± 0.13	0.64 ± 0.19	0.65 ± 0.15	0.65 ± 0.14	0.66 ± 0.18

[a] Features included: NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table C4. Macro precision [0–1] for temporal classification models are reported here for three temporal aggregation windows of 1 year (1Y-AG), 5 years (5Y-AG), and 8 years (8Y-AG) length in the Aggregated (AG) setting. Models are ordered by methodological group starting with dictionary-based (BOSS), distance-based (KNN-DTW), feature-based (TSFresh RF), pattern-based (Shapelet), deep learning (LSTM), and transformation-based (MiniRocket). The values reported are the mean (± standart deviation) across 20 repetitions of nested spatial cross-validation. Univariate input refers to NDVI only, whereas multivariate input includes multiple spectral indices and bands. The highest value in each row is highlighted in bold.

Time Window	Models					
	BOSS	KNN-DTW	TSFreshRF	Shapelet	LSTM	MiniRocket
Univariate (NDVI only, 1 predictor)						
1Y-AG	0.45 ± 0.06	0.53 ± 0.17	0.46 ± 0.15	0.50 ± 0.15	0.47 ± 0.15	$\mathbf{0.54 \pm 0.18}$
5Y-AG	0.59 ± 0.12	0.64 ± 0.12	0.39 ± 0.03	0.57 ± 0.13	0.55 ± 0.11	$\mathbf{0.65 \pm 0.12}$
8Y-AG	0.53 ± 0.18	0.54 ± 0.13	0.45 ± 0.06	0.51 ± 0.15	0.42 ± 0.12	$\mathbf{0.58 \pm 0.14}$
Multivariate (Multiple indices/bands, 12 predictors)[a]						
1Y-AG	–	0.60 ± 0.12	$\mathbf{0.73 \pm 0.08}$	–	0.59 ± 0.06	0.72 ± 0.14
5Y-AG	–	0.68 ± 0.08	0.75 ± 0.05	–	0.61 ± 0.05	$\mathbf{0.76 \pm 0.11}$
8Y-AG	–	0.64 ± 0.08	$\mathbf{0.68 \pm 0.06}$	–	0.57 ± 0.07	0.67 ± 0.13

[a] Features included: NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table C5. Macro recall [0–1] for temporal classification models are reported here for three temporal aggregation windows of 1 year (1Y-AG), 5 years (5Y-AG), and 8 years (8Y-AG) length in the Aggregated (AG) setting. Models are ordered by methodological group starting with dictionary-based (BOSS), distance-based (KNN-DTW), feature-based (TSFresh RF), pattern-based (Shapelet), deep learning (LSTM), and transformation-based (MiniRocket). The values reported are the mean (± standart deviation) across 20 repetitions of nested spatial cross-validation. Univariate input refers to NDVI only, whereas multivariate input includes multiple spectral indices and bands. The highest value in each row is highlighted in bold.

Time Window	Models					
	BOSS	KNN-DTW	TSFreshRF	Shapelet	LSTM	MiniRocket
Univariate (NDVI only, 1 predictor)						
1Y-AG	0.44 ± 0.06	0.52 ± 0.13	0.45 ± 0.10	0.49 ± 0.09	0.50 ± 0.11	$\mathbf{0.53 \pm 0.11}$
5Y-AG	0.52 ± 0.10	0.57 ± 0.06	0.44 ± 0.04	0.56 ± 0.09	0.57 ± 0.08	$\mathbf{0.62 \pm 0.12}$
8Y-AG	0.49 ± 0.17	$\mathbf{0.55 \pm 0.14}$	0.40 ± 0.06	0.50 ± 0.10	0.41 ± 0.08	0.53 ± 0.13
Multivariate (Multiple indices/bands, 12 predictors)[a]						
1Y-AG	–	0.62 ± 0.14	$\mathbf{0.71 \pm 0.09}$	–	0.64 ± 0.08	0.70 ± 0.16
5Y-AG	–	0.66 ± 0.12	$\mathbf{0.75 \pm 0.11}$	–	0.66 ± 0.08	0.72 ± 0.13
8Y-AG	–	0.59 ± 0.11	$\mathbf{0.63 \pm 0.14}$	–	0.59 ± 0.09	0.62 ± 0.15

[a] Features included: NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table C6. Overall accuracy [0–1] for temporal classification models are reported here for three temporal aggregation windows of 1 year (1Y-AG), 5 years (5Y-AG), and 8 years (8Y-AG) length in the Aggregated (AG) setting. Models are ordered by methodological group starting with dictionary-based (BOSS), distance-based (KNN-DTW), feature-based (TSFresh RF), pattern-based (Shapelet), deep learning (LSTM), and transformation-based (MiniRocket). The values reported are the mean (± standart deviation) across 20 repetitions of nested spatial cross-validation. Univariate input refers to NDVI only, whereas multivariate input includes multiple spectral indices and bands. The highest value in each row is highlighted in bold.

Time Window	Models					
	BOSS	KNN-DTW	TSFreshRF	Shapelet	LSTM	MiniRocket
Univariate (NDVI only, 1 predictor)						
1Y-AG	0.56 ± 0.16	$\mathbf{0.61 \pm 0.23}$	0.57 ± 0.21	0.59 ± 0.21	0.58 ± 0.19	$\mathbf{0.61 \pm 0.23}$
5Y-AG	0.67 ± 0.20	0.71 ± 0.18	0.59 ± 0.11	0.72 ± 0.15	0.64 ± 0.14	$\mathbf{0.74 \pm 0.17}$
8Y-AG	0.62 ± 0.24	0.65 ± 0.17	0.58 ± 0.18	$\mathbf{0.70 \pm 0.10}$	0.56 ± 0.17	$\mathbf{0.70 \pm 0.15}$
Multivariate (Multiple indices/bands, 12 predictors)[a]						
1Y-AG	–	0.65 ± 0.22	0.73 ± 0.13	–	0.70 ± 0.12	$\mathbf{0.74 \pm 0.22}$
5Y-AG	–	0.73 ± 0.18	0.77 ± 0.10	–	0.72 ± 0.14	$\mathbf{0.79 \pm 0.15}$
8Y-AG	–	0.67 ± 0.20	0.68 ± 0.16	–	0.71 ± 0.11	$\mathbf{0.74 \pm 0.16}$

[a] Features included: NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table C7. Cohen's κ [0–1] for temporal classification models are reported here for three temporal aggregation windows of 1 year (1Y-AG), 5 years (5Y-AG), and 8 years (8Y-AG) length in the Aggregated (AG) setting. Models are ordered by methodological group starting with dictionary-based (BOSS), distance-based (KNN-DTW), feature-based (TSFresh RF), pattern-based (Shapelet), deep learning (LSTM), and transformation-based (MiniRocket). The values reported are the mean ($\pm$ standart deviation) across 20 repetitions of nested spatial cross-validation. Univariate input refers to NDVI only, whereas multivariate input includes multiple spectral indices and bands. The highest value in each row is highlighted in bold.

Predictors	No. of features	Models					
		LR	SVM	KNN	RF	XGB	MLP
Geometric predictors							
Geometry	11	0.36 ± 0.14	0.34 ± 0.13	0.31 ± 0.16	0.35 ± 0.12	0.36 ± 0.15	$\mathbf{0.40 \pm 0.13}$
SAR predictors (Sentinel-1)							
VV/VH	4	0.18 ± 0.06	0.18 ± 0.09	0.21 ± 0.13	$\mathbf{0.24 \pm 0.14}$	0.23 ± 0.11	0.22 ± 0.15
Optical predictors (Sentinel-2)							
S2 bands	10	0.25 ± 0.08	$\mathbf{0.29 \pm 0.06}$	0.14 ± 0.07	0.22 ± 0.07	0.22 ± 0.06	0.27 ± 0.06
Vegetation indices	12	$\mathbf{0.39 \pm 0.12}$	$\mathbf{0.39 \pm 0.12}$	0.11 ± 0.08	0.18 ± 0.11	0.19 ± 0.12	0.29 ± 0.14
Sensor combinations							
Combination subset[a]	12	$\mathbf{0.34 \pm 0.08}$	0.31 ± 0.15	0.29 ± 0.16	0.29 ± 0.15	0.30 ± 0.14	$\mathbf{0.34 \pm 0.12}$
VV/VH + S2 bands	14	0.28 ± 0.10	0.22 ± 0.11	0.22 ± 0.14	0.31 ± 0.14	0.30 ± 0.13	$\mathbf{0.34 \pm 0.15}$
VV/VH + Vegetation indices	16	0.39 ± 0.07	$\mathbf{0.49 \pm 0.06}$	0.26 ± 0.14	0.31 ± 0.14	0.31 ± 0.11	0.40 ± 0.16
VV/VH + S2 bands + Vegetation indices	26	$\mathbf{0.37 \pm 0.08}$	0.36 ± 0.12	0.28 ± 0.16	0.32 ± 0.14	0.32 ± 0.13	0.36 ± 0.19

[a] Features included: NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table C8. Mean Cohen's κ [0–1] ($\pm$ standard deviation) for temporal classification models evaluated on different time-series inputs for the Aggregated (AG) setting. Models are ordered by methodological group: dictionary-based (BOSS), distance-based (KNN-DTW), feature-based (TSFresh RF), pattern-based (Shapelet), deep learning (LSTM), and transformation-based (MiniRocket). Values are the mean across 20 repetitions of nested spatial cross-validation (5 seeds $\times$ 4 spatial blocks; hyperparameters tuned within each outer fold on the training blocks only). Time-series inputs include three temporal aggregation windows: 1Y-AG (1 year), 5Y-AG (5 years), and 8Y-AG (8 years). Univariate input refers to NDVI only, whereas multivariate input includes multiple spectral indices and bands. The highest value in each row is highlighted in bold.

Time Window	Models					
	BOSS	KNN-DTW	TSFreshRF	Shapelet	LSTM	MiniRocket
Univariate (NDVI only, 1 predictor)						
1Y-AG	0.16 ± 0.10	$\mathbf{0.29 \pm 0.21}$	0.22 ± 0.17	0.25 ± 0.18	0.25 ± 0.17	$\mathbf{0.29 \pm 0.20}$
5Y-AG	0.34 ± 0.19	0.41 ± 0.16	0.20 ± 0.09	0.43 ± 0.17	0.36 ± 0.13	$\mathbf{0.49 \pm 0.21}$
8Y-AG	0.27 ± 0.28	0.32 ± 0.19	0.15 ± 0.10	0.32 ± 0.12	0.11 ± 0.13	$\mathbf{0.34 \pm 0.23}$
Multivariate (Multiple indices/bands, 12 predictors)[a]						
1Y-AG	–	0.37 ± 0.22	0.49 ± 0.14	–	0.42 ± 0.13	$\mathbf{0.53 \pm 0.28}$
5Y-AG	–	0.48 ± 0.20	0.55 ± 0.11	–	0.47 ± 0.16	$\mathbf{0.60 \pm 0.22}$
8Y-AG	–	0.35 ± 0.21	0.34 ± 0.19	–	0.38 ± 0.14	$\mathbf{0.46 \pm 0.24}$

[a] Features included: NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table C9. Overall accuracy (OA), Cohen's kappa (κ), and per-class omission and commission errors [0–1] for the best model of each predictor group. Values are the mean ($\pm$ standard deviation) across 20 repetitions of nested spatial cross-validation.

Configuration	OA	Cohen's κ	Omission error			Commission error		
			Bark beetle	Defoliators	Wind	Bark beetle	Defoliators	Wind
Geometric (MLP)	0.72 ± 0.13	0.40 ± 0.13	0.19 ± 0.18	0.63 ± 0.12	0.48 ± 0.15	0.26 ± 0.22	0.48 ± 0.26	0.41 ± 0.17
Spectro-radiometric (MLP)	0.66 ± 0.18	0.36 ± 0.19	0.21 ± 0.13	$\mathbf{0.35 \pm 0.26}$	0.52 ± 0.18	0.29 ± 0.29	0.27 ± 0.13	0.51 ± 0.16
Temporal, 1YR (MiniRocket)	0.74 ± 0.22	0.53 ± 0.28	0.15 ± 0.15	0.39 ± 0.27	0.36 ± 0.25	0.26 ± 0.30	0.28 ± 0.16	0.31 ± 0.08
Temporal, 5YR (MiniRocket)	$\mathbf{0.79 \pm 0.15}$	$\mathbf{0.60 \pm 0.22}$	$\mathbf{0.09 \pm 0.11}$	0.40 ± 0.23	$\mathbf{0.33 \pm 0.20}$	$\mathbf{0.23 \pm 0.26}$	$\mathbf{0.25 \pm 0.16}$	$\mathbf{0.25 \pm 0.07}$
Temporal, 8YR (MiniRocket)	0.74 ± 0.16	0.46 ± 0.24	0.19 ± 0.26	0.50 ± 0.26	0.44 ± 0.21	0.29 ± 0.28	0.34 ± 0.20	0.35 ± 0.08

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table C10. Per-class F1 score, omission and commission errors [0–1] for the static predictor sets in the Aggregated (AG) setting. For each predictor set, the model with the highest $F1_{macro}$ score (in brackets) is shown. Values are the mean (± standard deviation) across 20 repetitions of nested spatial cross-validation.

Predictors	F1 score			Omission error			Commission error		
	Bark beetle	Defoliators	Wind	Bark beetle	Defoliators	Wind	Bark beetle	Defoliators	Wind
Geometric predictors									
Geometry (MLP)	0.77 ± 0.20	0.41 ± 0.17	0.52 ± 0.13	0.19 ± 0.18	0.63 ± 0.12	0.48 ± 0.15	0.26 ± 0.22	0.48 ± 0.26	$\mathbf{0.41 \pm 0.17}$
SAR predictors (Sentinel-1)									
VV/VH (KNN)	0.67 ± 0.19	$\mathbf{0.66 \pm 0.10}$	0.25 ± 0.14	0.25 ± 0.10	0.36 ± 0.18	0.73 ± 0.14	0.33 ± 0.26	$\mathbf{0.26 \pm 0.15}$	0.66 ± 0.18
Optical predictors (Sentinel-2)									
S2 bands (SVM)	0.74 ± 0.15	0.48 ± 0.12	0.38 ± 0.13	0.18 ± 0.05	0.51 ± 0.16	0.61 ± 0.19	0.31 ± 0.21	0.47 ± 0.19	0.51 ± 0.23
Vegetation indices (SVM)	0.76 ± 0.18	0.60 ± 0.16	0.48 ± 0.09	$\mathbf{0.15 \pm 0.09}$	0.44 ± 0.21	0.52 ± 0.17	0.29 ± 0.24	0.29 ± 0.13	0.44 ± 0.17
Sensor combinations									
Combination subset[a](MLP)	0.72 ± 0.18	$\mathbf{0.66 \pm 0.19}$	0.42 ± 0.10	0.21 ± 0.08	$\mathbf{0.31 \pm 0.28}$	0.55 ± 0.12	0.29 ± 0.26	0.29 ± 0.08	0.51 ± 0.21
VV/VH + S2 bands (MLP)	0.72 ± 0.21	0.60 ± 0.15	0.42 ± 0.07	0.22 ± 0.12	0.37 ± 0.25	0.52 ± 0.20	0.29 ± 0.28	0.33 ± 0.11	0.51 ± 0.17
VV/VH + Vegetation indices (SVM)	$\mathbf{0.82 \pm 0.11}$	0.59 ± 0.16	$\mathbf{0.54 \pm 0.18}$	$\mathbf{0.15 \pm 0.12}$	0.38 ± 0.26	$\mathbf{0.45 \pm 0.19}$	$\mathbf{0.20 \pm 0.13}$	0.36 ± 0.09	$\mathbf{0.41 \pm 0.22}$
VV/VH + S2 bands + Vegetation indices (MLP)	0.72 ± 0.23	$\mathbf{0.66 \pm 0.21}$	0.43 ± 0.12	0.21 ± 0.13	0.35 ± 0.26	0.52 ± 0.18	0.29 ± 0.29	0.27 ± 0.13	0.51 ± 0.16

[a] Features included: NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

Table C11. Per-class F1 score, omission and commission errors [0–1] for the temporal predictors in the Aggregated (AG) setting. For each predictor set, the model with the highest $F1_{macro}$ score (in brackets) is shown. Values are the mean (± standard deviation) across 20 repetitions of nested spatial cross-validation.

Time Window	F1 score			Omission error			Commission error		
	Bark beetle	Defoliators	Wind	Bark beetle	Defoliators	Wind	Bark beetle	Defoliators	Wind
Univariate (NDVI only)									
1Y-AG (MiniRocket)	0.71 ± 0.25	0.36 ± 0.09	0.40 ± 0.20	0.23 ± 0.13	0.67 ± 0.16	0.51 ± 0.31	0.32 ± 0.30	0.48 ± 0.28	0.59 ± 0.11
5Y-AG(MiniRocket)	$\mathbf{0.79 \pm 0.22}$	$\mathbf{0.44 \pm 0.17}$	$\mathbf{0.62 \pm 0.09}$	$\mathbf{0.13 \pm 0.13}$	$\mathbf{0.61 \pm 0.16}$	$\mathbf{0.38 \pm 0.19}$	$\mathbf{0.25 \pm 0.26}$	$\mathbf{0.46 \pm 0.22}$	$\mathbf{0.33 \pm 0.07}$
8Y-AG (MiniRocket)	0.74 ± 0.24	0.36 ± 0.22	0.49 ± 0.15	0.19 ± 0.23	0.68 ± 0.20	0.53 ± 0.20	0.31 ± 0.26	0.50 ± 0.28	0.45 ± 0.15
Multivariate (12 predictors)									
1Y-AG (MiniRocket)	0.77 ± 0.26	0.63 ± 0.22	0.63 ± 0.17	0.15 ± 0.15	$\mathbf{0.39 \pm 0.27}$	0.36 ± 0.25	0.26 ± 0.30	0.28 ± 0.16	0.31 ± 0.08
5Y-AG (MiniRocket)	$\mathbf{0.82 \pm 0.21}$	$\mathbf{0.64 \pm 0.20}$	$\mathbf{0.69 \pm 0.09}$	$\mathbf{0.09 \pm 0.11}$	0.40 ± 0.23	$\mathbf{0.33 \pm 0.20}$	$\mathbf{0.23 \pm 0.26}$	$\mathbf{0.25 \pm 0.16}$	$\mathbf{0.25 \pm 0.07}$
8Y-AG (MiniRocket)	0.75 ± 0.27	0.55 ± 0.23	0.58 ± 0.12	0.19 ± 0.26	0.50 ± 0.26	0.44 ± 0.21	0.29 ± 0.28	0.34 ± 0.20	0.35 ± 0.08

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table C12. Effect of temporal order and time-series length on overall accuracy (OA), Cohen's kappa (κ), $F1_{macro}$ and class-wise F1 per disturbance type. In panel (a) the same 12 variables as a single-date spectro-radiometric predictors or as time series of increasing length is shown. In panel (b) the performance under the time-order permutation for the 5YR multivariate MiniRocket model, with the time axis permuted either identically for all polygons or independently for each polygon, is shown. Values are means [0–1] ($\pm$ standard deviation) across the repetitions of the nested spatial cross-validation.

Input	OA	Cohen's κ	$F1_{macro}$	$F1_{bark\ beetle}$	$F1_{defoliators}$	$F1_{wind}$
(a) Unaltered time series vs. single date, same 12 variables[a]						
Single date (MLP)	0.66 ± 0.13	0.34 ± 0.12	0.60 ± 0.07	0.72 ± 0.18	$\mathbf{0.66 \pm 0.19}$	0.42 ± 0.10
1YR time series (MiniRocket)	0.74 ± 0.22	0.53 ± 0.28	0.68 ± 0.16	0.77 ± 0.26	0.63 ± 0.22	0.63 ± 0.17
5YR time series (MiniRocket)	$\mathbf{0.79 \pm 0.15}$	$\mathbf{0.60 \pm 0.22}$	$\mathbf{0.72 \pm 0.13}$	$\mathbf{0.82 \pm 0.21}$	0.64 ± 0.20	$\mathbf{0.69 \pm 0.09}$
8YR time series (MiniRocket)	0.74 ± 0.16	0.46 ± 0.24	0.63 ± 0.14	0.75 ± 0.27	0.55 ± 0.23	0.58 ± 0.12
(b) Shuffled dates, 5YR multivariate (MiniRocket)						
Shared shuffle (3×)	$\mathbf{0.80 \pm 0.15}$	$\mathbf{0.63 \pm 0.21}$	$\mathbf{0.74 \pm 0.13}$	$\mathbf{0.82 \pm 0.20}$	0.68 ± 0.21	$\mathbf{0.72 \pm 0.09}$
Per-polygon shuffle (3×)	0.78 ± 0.14	0.56 ± 0.19	0.71 ± 0.12	0.81 ± 0.18	$\mathbf{0.69 \pm 0.20}$	0.62 ± 0.07

[a] NDVI, NDMI, GNDVI, NDRN, B12, B11, B06, B05, VV, VH (ascending and descending).

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

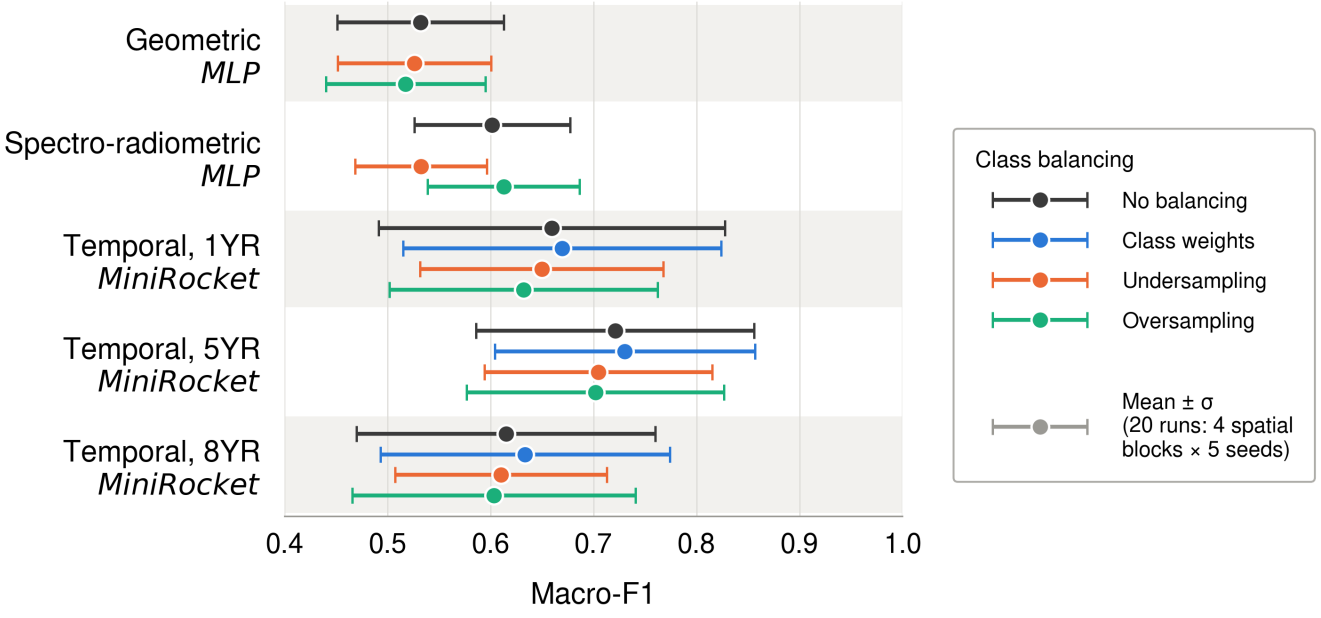


Figure C2. Performance of the best model of each predictor family (geometric, spectro-radiometric and temporal 1YR, 5YR and 8YR) when class balancing is applied to the training data. The balancing strategies are class weights inversely proportional to the class frequencies, random undersampling of all classes to the size of the smallest class, or random oversampling of the smaller classes to the size of the largest class. Dots show the mean and whiskers ± one standard deviation across the 20 repetitions of the nested spatial cross-validation.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Appendix D: Discussion

Table D1. Omission rates by disturbance class across disturbance classification studies. Dashes indicate classes not evaluated separately.

Study	Validation	Bark beetle	Defoliators	Wind
This study	Spatially exclusive CV	9.3%	40.1%	33.2%
Senf et al. (2015)	OOB within study area[a]	12.9–15.0%[b]	11.3–28.9%[b]	–
Viana-Soto and Senf (2025)	Spatial block CV	19.1%[c]	–	19.1%[c]
Du et al. (2023)	Held-out regions	32.7%[d]		10.0%
Schroeder et al. (2017)	Per-scene OOB[a]	29–88%[e]		24–60%

[a] Test sets drawn from the same study region; not spatially exclusive.

[b] Range from pure to mixed forest stands.

[c] Bark beetle and wind not separated; value reflects the combined class.

[d] Reported as “Pests and Diseases” class; bark beetle and defoliators not separated.

[e] Reported as “Stress” class; bark beetle and defoliators not separated.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table D2. Mean Sentinel-1 backscatter (γ^0) anomalies (dB) in the disturbance year relative to each polygon's own mean of the same calendar month in the two preceding years, which removes the normal seasonal cycle. Values are averaged across all polygons per class by season (bark beetle $n = 703$; defoliators $n = 132$; wind $n = 468$; only polygons with a pre-disturbance baseline). The baseline is the mean pre-disturbance γ^0. Spring = March–May; Summer = June–August; Autumn = September–November. Values in bold have a 95 % confidence interval across polygons.

Class	Ch.	Baseline γ^0 (dB)	Jan–Feb Δ	Spring Δ	Summer Δ	Autumn Δ	Dec Δ
Bark beetle	VV	−9.1	**+0.14**	**+0.27**	**+0.18**	**+0.32**	**+0.41**
	VH	−13.9	**−0.04**	+0.08	−0.05	−0.03	−0.02
Defoliators	VV	−7.5	+0.02	+0.01	+0.02	**+0.06**	**+0.09**
	VH	−12.7	−0.07	−0.05	**−0.07**	−0.04	−0.06
Wind	VV	−8.6	**+0.28**	**+0.23**	**+0.22**	**+0.21**	**+0.26**
	VH	−13.7	**+0.12**	**+0.06**	**+0.06**	−0.01	−0.02

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

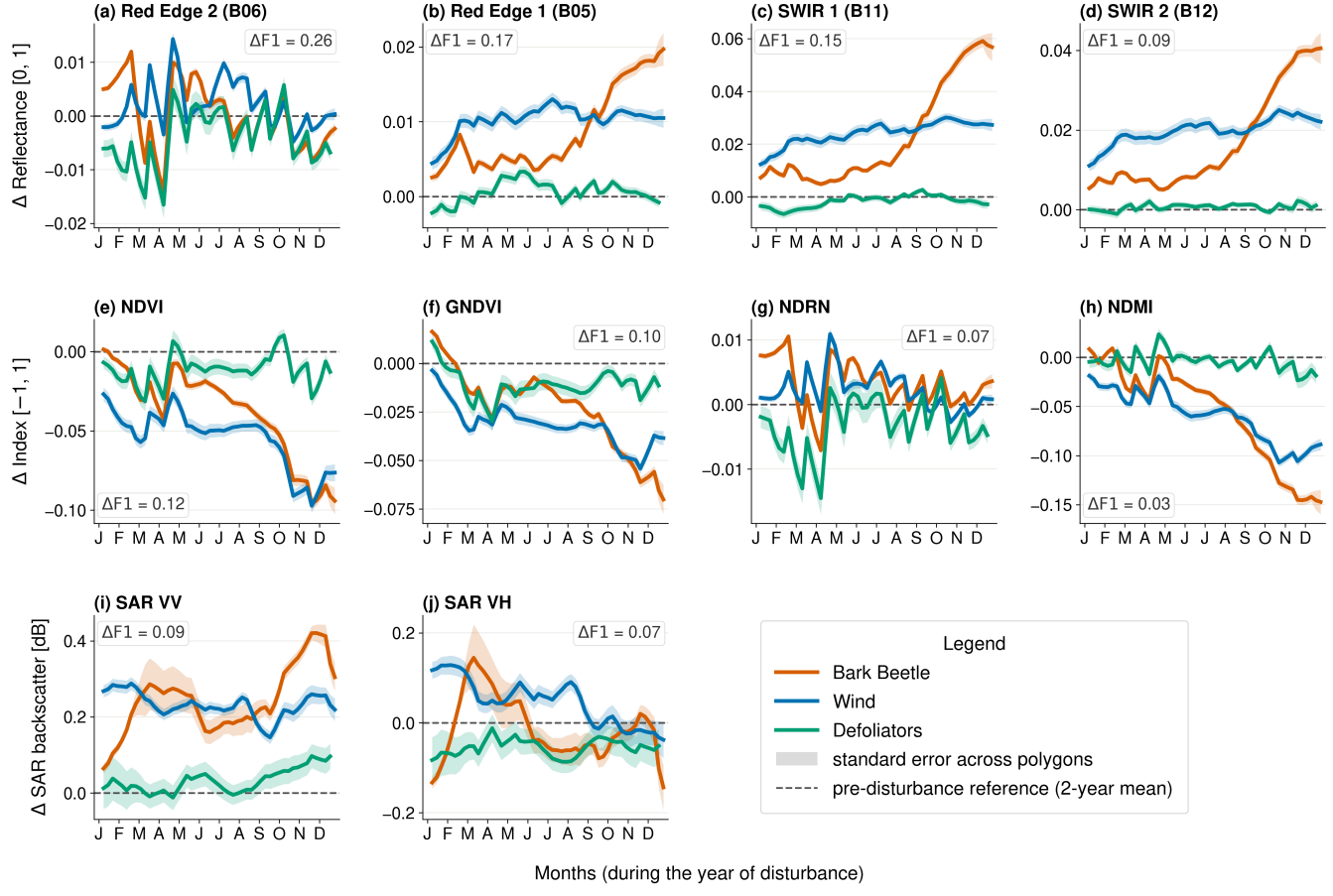


Figure D1. Anomalies during the disturbance year for the features in panels (a)–(d) for surface reflectance of the red-edge and short-wave infrared bands, (e)–(h) spectral indices and (i)–(j) SAR backscatter γ^0 (ascending and descending orbits averaged). Anomalies are calculated as the weekly value minus the mean of the same calendar month in the two years before the disturbance (pre-disturbance reference, dashed line at zero). Colored lines show the mean anomaly across all polygons per disturbance class (bark beetle $n = 899$ (orange); defoliators $n = 144$ (green); wind $n = 469$ (blue)) and shading indicates $\pm$ standard error of the mean. A 3-week rolling average is applied to reduce high-frequency noise. The $\mathrm{F1}_{\mathrm{macro}}$ value in each panel is the reduction of the feature in the feature ablation analysis (Sec. 3.3, Fig. 7)

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

References

Abdullah, H., Skidmore, A. K., Darvishzadeh, R., and Heurich, M.: Timing of red-edge and shortwave infrared reflectance critical for early stress detection induced by bark beetle (*Ips typographus*, L.) attack, International Journal of Applied Earth Observation and Geoinformation, 82, 101 900, https://doi.org/10.1016/j.jag.2019.101900, 2019.

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A Next-generation Hyperparameter Optimization Framework, https://doi.org/10.48550/arXiv.1907.10902, arXiv:1907.10902, 2019.

Alonso, L., Picos, J., and Armesto, J.: Automatic Identification of Forest Disturbance Drivers Based on Their Geometric Pattern in Atlantic Forests, Remote. Sens., 14, 697, https://doi.org/10.3390/rs14030697, 2022.

Anderegg, W. R. L., Hicke, J. A., Fisher, R. A., Allen, C. D., Aukema, J., Bentz, B., Hood, S., Lichstein, J. W., Macalady, A. K., McDowell, N., Pan, Y., Raffa, K., Sala, A., Shaw, J. D., Stephenson, N. L., Tague, C., and Zeppel, M.: Tree mortality from drought, insects, and their interactions in a changing climate, New Phytol., 208, 674–683, 2015.

Aulló-Maestro, I., Francini, S., Chirici, G., Gómez, C., Montes, F., Alberdi, I., and Cañellas, I.: A Two-Step Framework for Mapping, Classification, and Area Estimation of Stand- and Non-Stand-Replacing Forest Disturbances, Remote Sensing, 18, 1038, https://doi.org/10.3390/rs18071038, 2026.

Aybar, C., Ysuhuaylas, L., Loja, J., Gonzales, K., Herrera, F., Bautista, L., Yali, R., Flores, A., Diaz, L., Cuenca, N., Espinoza, W., Prudencio, F., Llactayo, V., Montero, D., Sudmanns, M., Tiede, D., Mateo-García, G., and Gómez-Chova, L.: CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2, Scientific Data, 9, 782, https://doi.org/10.1038/s41597-022-01878-2, 2022.

Badgley, G., Field, C. B., and Berry, J. A.: Canopy near-infrared reflectance and terrestrial photosynthesis, Science Advances, 3, https://doi.org/10.1126/sciadv.1602244, 2017.

Balducci, L., Fierravanti, A., Rossi, S., Delzon, S., De Grandpré, L., Kneeshaw, D. D., and Deslauriers, A.: The paradox of defoliation: Declining tree water status with increasing soil water content, Agricultural and Forest Meteorology, 290, 108 025, https://doi.org/https://doi.org/10.1016/j.agrformet.2020.108025, 2020.

Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R. J., Li, H., and Moran, M. S.: Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data, in: Proceedings of the 5th International Conference on Precision Agriculture, Bloomington, MN, USA, 2000.

Bauer-Marschallinger, B., Sabel, D., and Wagner, W.: Optimisation of global grids for high-resolution remote sensing data, Computers & Geosciences, 72, 84–93, https://doi.org/10.1016/j.cageo.2014.07.005, 2014.

Bozzini, A., Huo, L., Brugnaro, S., Morgante, G., Persson, H. J., Finozzi, V., Battisti, A., and Faccoli, M.: Multispectral drone images for the early detection of bark beetle infestations: assessment over large forest areas in the Italian South-Eastern Alps, Frontiers in Forests and Global Change, 8, https://doi.org/10.3389/ffgc.2025.1532954, 2025.

Breiman, L.: Random forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001a.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001b.

Bustamante, M., Roy, J., Ospina, D., Achakulwisut, P., Aggarwal, A., Bastos, A., Broadgate, W., Canadell, J. G., Carr, E. R., Chen, D., and et al.: Ten new insights in climate science 2023, Global Sustainability, 7, e19, https://doi.org/10.1017/sus.2023.25, 2023.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Camps-Valls, G., Campos-Taberner, M., Álvaro Moreno-Martínez, Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Marí, J., García-Haro, F. J., Guanter, L., Jung, M., Gamon, J. A., Reichstein, M., and Running, S. W.: A unified vegetation index for quantifying the terrestrial biosphere, Science Advances, 7, https://doi.org/10.1126/sciadv.abc7447, 2021.

Candotti, A., De Giglio, M., Dubbini, M., and Tomelleri, E.: A Sentinel-2 Based Multi-Temporal Monitoring Framework for Wind and Bark Beetle Detection and Damage Mapping, Remote Sensing, 14, 6105, https://doi.org/10.3390/rs14236105, 2022.

Cannière, S. D., Vereecken, H., Defourny, P., and Jonard, F.: Remote Sensing of Instantaneous Drought Stress at Canopy Level Using Sun-Induced Chlorophyll Fluorescence and Canopy Reflectance, Remote. Sens., 14, 2642, https://doi.org/10.3390/rs14112642, 2022.

Chabrillat, S., Foerster, S., Segl, K., Beamish, A., Brell, M., Asadzadeh, S., Milewski, R., Ward, K. J., Brosinsky, A., Koch, K., Scheffler, D., Guillaso, S., Kokhanovsky, A., Roessner, S., Guanter, L., Kaufmann, H., Pinnel, N., Carmona, E., Storch, T., Hank, T., Berger, K., Wocher, M., Hostert, P., van der Linden, S., Okujeni, A., Janz, A., Jakimow, B., Bracher, A., Soppa, M. A., Alvarado, L. M. A., Buddenbaum, H., Heim, B., Heiden, U., Moreno, J., Ong, C., Bohn, N., Green, R. O., Bachmann, M., Kokaly, R., Schodlok, M., Painter, T. H., Gascon, F., Buongiorno, F., Mottus, M., Brando, V. E., Feilhauer, H., Betz, M., Baur, S., Feckl, R., Schickling, A., Krieger, V., Bock, M., La Porta, L., and Fischer, S.: The EnMAP spaceborne imaging spectroscopy mission: Initial scientific results two years after launch, Remote Sens. Environ., 315, 114 379, 2024.

Chen, J. and Zhang, Z.: An improved fusion of Landsat-7/8, Sentinel-2, and Sentinel-1 data for monitoring alfalfa: Implications for crop remote sensing, International Journal of Applied Earth Observation and Geoinformation, 124, 103 533, https://doi.org/10.1016/j.jag.2023.103533, 2023.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, https://doi.org/10.1145/2939672.2939785, arXiv:1603.02754 [cs.LG], 2016.

Christ, M., Braun, N., Neuffer, J., and Kempa-Liehr, A. W.: Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package), Neurocomputing, 307, 72–77, https://doi.org/10.1016/j.neucom.2018.03.067, 2018a.

Christ, M., Braun, N., Neuffer, J., and Kempa-Liehr, A. W.: Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package), Neurocomputing, 307, 72–77, https://doi.org/10.1016/j.neucom.2018.03.067, 2018b.

Coffelt, J. and Livingston, R. K.: Second U.S. Geological Survey Wildland Fire Workshop: Los Alamos, New Mexico, October 31-November 3, 2000, Open-File Report, https://doi.org/10.3133/ofr0211, 2002.

Cohen, W. B., Yang, Z., Healey, S. P., Kennedy, R. E., and Gorelick, N.: A LandTrendr multispectral ensemble for forest disturbance detection, Remote Sensing of Environment, 205, 131–140, https://doi.org/10.1016/j.rse.2017.11.015, 2018.

Cohen, W. B., Healey, S. P., Yang, Z., Zhu, Z., and Gorelick, N.: Diversity of Algorithm and Spectral Band Inputs Improves Landsat Monitoring of Forest Disturbance, Remote Sensing, 12, 1673, https://doi.org/10.3390/rs12101673, 2020.

Coleman, T. W., Graves, A. D., Heath, Z., Flowers, R. W., Hanavan, R. P., Cluck, D. R., and Ryerson, D.: Accuracy of aerial detection surveys for mapping insect and disease disturbances in the United States, Forest Ecology and Management, 430, 321–336, https://doi.org/10.1016/j.foreco.2018.08.020, 2018.

Coppo, P., Taiti, A., Pettinato, L., Francois, M., Taccola, M., and Drusch, M.: Fluorescence imaging spectrometer (FLORIS) for ESA FLEX mission, Remote Sens. (Basel), 9, 649, 2017.

Cortes, C. and Vapnik, V.: Support-vector networks, Machine Learning, 20, 273–297, https://doi.org/10.1007/BF00994018, 1995.

Cover, T. M. and Hart, P. E.: Nearest neighbor pattern classification, IEEE Transactions on Information Theory, 13, 21–27, https://doi.org/10.1109/TIT.1967.1053964, 1967.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Cremer, F., Urbazaev, M., Cortés, J., Truckenbrodt, J., Schmullius, C., and Thiel, C.: Potential of Recurrence Metrics from Sentinel-1 Time Series for Deforestation Mapping, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 5233–5240, https://doi.org/10.1109/JSTARS.2020.3019333, 2020.

Crist, E. P. and Cicone, R. C.: A Physically-Based Transformation of Thematic Mapper Data—The TM Tasseled Cap, IEEE Transactions on Geoscience and Remote Sensing, GE-22, https://doi.org/10.1109/tgrs.1984.350619, 1984.

Dalponte, M., Marzini, S., Solano-Correa, Y. T., Tonon, G., Vescovo, L., and Gianelle, D.: Mapping forest windthrows using high spatial resolution multispectral satellite images, International Journal of Applied Earth Observation and Geoinformation, 93, 102 206, https://doi.org/10.1016/j.jag.2020.102206, 2020.

Dalponte, M., Solano-Correa, Y. T., Marinelli, D., Liu, S., Yokoya, N., and Gianelle, D.: Detection of forest windthrows with bitemporal COSMO-SkyMed and Sentinel-1 SAR data, Remote Sensing of Environment, 297, 113 787, https://doi.org/10.1016/j.rse.2023.113787, 2023.

De Grandpré, L., Marchand, M., Kneeshaw, D. D., Paré, D., Boucher, D., Bourassa, S., Gervais, D., Simard, M., Griffin, J. M., and Pureswaran, D. S.: Defoliation-induced changes in foliage quality may trigger broad-scale insect outbreaks, Communications Biology, 5, 463, https://doi.org/10.1038/s42003-022-03407-8, 2022.

Dempster, A., Petitjean, F., and Webb, G. I.: ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels, Data Mining and Knowledge Discovery, 34, 1454–1495, https://doi.org/10.1007/s10618-020-00701-z, 2020.

Dempster, A., Schmidt, D. F., and Webb, G. I.: MiniRocket: A Very Fast (Almost) Deterministic Transform for Time Series Classification, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 248–257, ACM, Virtual Event Singapore, https://doi.org/10.1145/3447548.3467231, 2021.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., and Keogh, E.: Querying and mining of time series data: experimental comparison of representations and distance measures, Proceedings of the VLDB Endowment, 1, 1542–1552, https://doi.org/10.14778/1454159.1454226, 2008.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P.: Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services, Remote Sensing of Environment, 120, 25–36, https://doi.org/10.1016/j.rse.2011.11.026, 2012.

Du, B., Yuan, Z., Bo, Y., and Zhang, Y.: A Combined Deep Learning and Prior Knowledge Constraint Approach for Large-Scale Forest Disturbance Detection Using Time Series Remote Sensing Data, Remote Sensing, 15, 2963, https://doi.org/10.3390/rs15122963, 2023.

Dudney, J., Edwards, J., Harvey, B. J., and Seidl, R.: Climate change effects on interacting disturbances in forest ecosystems, Annu. Rev. Ecol. Evol. Syst., 56, 393–420, 2025.

Eifler, L., Müller, F., and Bastos, A.: Evaluating the consistency of forest disturbance datasets in continental USA, Biogeosciences, 23, 1291–1325, https://doi.org/10.5194/bg-23-1291-2026, 2026.

Einzmann, K., Immitzer, M., Böck, S., Bauer, O., Schmitt, A., and Atzberger, C.: Windthrow Detection in European Forests with Very High-Resolution Optical Data, Forests, 8, 21, https://doi.org/10.3390/f8010021, 2017.

FAO: Global Forest Resources Assessment 2020 (FRA2020), Tech. rep., Food and Agriculture Organization of the United Nations, Rome, https://doi.org/10.4060/ca9825en, 2020.

Flores, A., Herndon, K., Thapa, R., and Cherrington, E., eds.: The SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation, SERVIR Global Science Coordination Office, NASA Marshall Space Flight Center, https://doi.org/10.25966/nr2c-s697, 2019.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Forest Service U.S. Department of Agriculture: FIA DataMart, https://research.fs.usda.gov/products/dataandtools/tools/fia-datamart, data downloaded: 05.06.2023, 2023.

Forest Service U.S. Department of Agriculture: Detection Surveys, https://www.fs.usda.gov/science-technology/data-tools-products/fhp-mapping-reporting/detection-surveys, data downloaded: 12.03.2024, 2024.

Forzieri, G., Pecchi, M., Girardello, M., Mauri, A., Klaus, M., Nikolov, C., Rüetschi, M., Gardiner, B., Tomaštík, J., Small, D., Nistor, C., Jonikavicius, D., Spinoni, J., Feyen, L., Giannetti, F., Comino, R., Wolynski, A., Pirotti, F., Maistrelli, F., Savulescu, I., Wurpillot-Lucas, S., Karlsson, S., Zieba-Kulawik, K., Strejczek-Jazwinska, P., Mokroš, M., Franz, S., Krejci, L., Haidu, I., Nilsson, M., Wezyk, P., Catani, F., Chen, Y.-Y., Luyssaert, S., Chirici, G., Cescatti, A., and Beck, P. S. A.: A spatially explicit database of wind disturbances in European forests over the period 2000–2018, Earth System Science Data, 12, 257–276, https://doi.org/10.5194/essd-12-257-2020, 2020.

Forzieri, G., Girardello, M., Ceccherini, G., Spinoni, J., Feyen, L., Hartmann, H., Beck, P. S. A., Camps-Valls, G., Chirici, G., Mauri, A., and Cescatti, A.: Emergent vulnerability to climate-driven disturbances in European forests, Nature Communications, 12, 1081, https://doi.org/10.1038/s41467-021-21399-7, number: 1, 2021.

Forzieri, G., Dutrieux, L. P., Elia, A., Eckhardt, B., Caudullo, G., Álvarez Taboada, F., Andriolo, A., Bălăcenoiu, F., Bastos, A., Buzatu, A., Castedo Dorado, F., Dobrovolný, L., Duduman, M.-L., Fernandez-Carrillo, A., Cescatti, A., A Beck, P. S., Giovanni Forzieri, C., and Commission, E.: The Database of European Forest Insect and Disease Disturbances: DEFID2, Glob Change Biol, 00, 30, https://doi.org/10.1111/gcb.16912, 2023.

Frazier, A. E. and Hemingway, B. L.: A Technical Review of Planet Smallsat Data: Practical Considerations for Processing and Using PlanetScope Imagery, Remote Sensing, 13, 3930, https://doi.org/10.3390/rs13193930, 2021.

Gitelson, A. and Merzlyak, M. N.: Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves, Journal of Photochemistry and Photobiology B: Biology, 22, https://doi.org/10.1016/1011-1344(93)06963-4, 1994.

Gitelson, A. A., Kaufman, Y. J., and Merzlyak, M. N.: Use of a green channel in remote sensing of global vegetation from EOS-MODIS, Remote Sensing of Environment, 58, https://doi.org/10.1016/s0034-4257(96)00072-7, 1996.

Gitelson, A. A., Gritz †, Y., and Merzlyak, M. N.: Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves, Journal of Plant Physiology, 160, 271–282, https://doi.org/10.1078/0176-1617-00887, 2003.

Gouveia, C., DaCamara, C. C., and Trigo, R. M.: Post-fire vegetation recovery in Portugal based on spot/vegetation data, Natural Hazards and Earth System Sciences, 10, 673–684, https://doi.org/10.5194/nhess-10-673-2010, 2010.

Graefe, S., Rodrigo, R., Cueva, E., Butz, P., Werner, F. A., and Homeier, J.: Impact of disturbance on forest structure and tree species composition in a tropical dry forest of South Ecuador, ECOTROPICA, 22, 202 002–202 002, https://www.ecotropica.eu/index.php/ecotropica/article/view/31, 2020.

Grandpré, L. D., Waldron, K., Bouchard, M., Gauthier, S., Beaudet, M., Ruel, J.-C., Hébert, C., and Kneeshaw, D.: Incorporating Insect and Wind Disturbances in a Natural Disturbance-Based Management Framework for the Boreal Forest, Forests, https://doi.org/10.3390/f9080471, 2018.

Hansen, M., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S., Goetz, S., Loveland, T., Kommareddy, A., Egorov, A., Chini, L., Justice, C., and Townshed, J.: Global Forest Change 2000-2024 Data Download, https://storage.googleapis.com/earthenginepartners-hansen/GFC-2024-v1.12/download.html, 2024.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R.: High-resolution global maps of 21st-century forest cover change, Science, 342, 850–853, https://doi.org/10.1126/SCIENCE.1244693, data downloaded: 09.08.2022, 2013.

Harris, N. L., Hagen, S. C., Saatchi, S. S., Pearson, T. R. H., Woodall, C. W., Domke, G. M., Braswell, B. H., Walters, B. F., Brown, S., Salas, W., Fore, A., and Yu, Y.: Attribution of net carbon change by disturbance type across forest lands of the conterminous United States, Carbon Balance and Management, 11, 24, https://doi.org/10.1186/s13021-016-0066-5, 2016.

He, H. and Garcia, E. A.: Learning from Imbalanced Data, IEEE Transactions on Knowledge and Data Engineering, 21, 1263–1284, https://doi.org/10.1109/TKDE.2008.239, 2009.

Hicke, J. A., Xu, B., Meddens, A. J. H., and Egan, J. M.: Characterizing recent bark beetle-caused tree mortality in the western United States from aerial surveys, Forest Ecology and Management, 475, 118 402, https://doi.org/10.1016/j.foreco.2020.118402, 2020.

Hills, J., Lines, J., Baranauskas, E., Mapp, J., and Bagnall, A.: Classification of time series by shapelet transformation, Data Mining and Knowledge Discovery, 28, 851–881, https://doi.org/10.1007/s10618-013-0322-1, 2014.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Computation, 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997a.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural Computation, 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997b.

Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X.: Applied Logistic Regression, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ, 3 edn., https://doi.org/10.1002/9781118548387, 2013.

Hubbard, R., Rhoades, C., Elder, K., and Negrón, J.: Changes in transpiration and foliage growth in lodgepole pine trees following mountain pine beetle attack and mechanical girdling, Forest Ecology and Management, 289, 312–317, https://doi.org/10.1016/j.foreco.2012.09.028, 2013.

Huo, L., Persson, H. J., and Lindberg, E.: Early detection of forest stress from European spruce bark beetle attack, and a new vegetation index: Normalized distance red & SWIR (NDRS), Remote Sensing of Environment, 255, 112 240, https://doi.org/10.1016/j.rse.2020.112240, 2021.

ICP Forests: Forest Condition in Europe: The 2025 Assessment. ICP Forests Technical Report under the UNECE Convention on Long-range Transboundary Air Pollution (Air Convention), Tech. rep., Thünen Institute, Eberswalde, https://doi.org/10.3220/253-2025-62, 2025.

IPCC: Summary for Policymakers, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., p. 3-32, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896.001, type: Book Section, 2021.

Irons, J. R., Dwyer, J. L., and Barsi, J. A.: The next Landsat satellite: The Landsat Data Continuity Mission, Remote Sensing of Environment, 122, 11–21, https://doi.org/10.1016/j.rse.2011.08.026, 2012.

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A.: Deep learning for time series classification: a review, Data Mining and Knowledge Discovery, 33, 917–963, https://doi.org/10.1007/s10618-019-00619-1, 2019.

Junttilaa, S., Holopainena, M., Vastarantab, M., Lyytikäinen-Saarenmaaa, P., Kaartinend, H., Hyyppäb, J., and Hyyppäb, H.: The potential of dual-wavelength terrestrial lidar in early detection of Ips typographus (L.) infestation – Leaf water content as a proxy, 2019.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Juola, J., Hovi, A., and Rautiainen, M.: Comparison of contemporaneous Sentinel-2 and EnMAP data for vegetation index-based estimation of leaf area index and canopy closure of a boreal forest, Eur. J. Remote Sens., 57, 2432 975, 2024.

Justice, C., Vermote, E., Townshend, J., Defries, R., Roy, D., Hall, D., Salomonson, V., Privette, J., Riggs, G., Strahler, A., Lucht, W., Myneni, R., Knyazikhin, Y., Running, S., Nemani, R., Wan, Z., Huete, A., van Leeuwen, W., Wolfe, R., Giglio, L., Muller, J., Lewis, P., and Barnsley, M.: The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research, IEEE Transactions on Geoscience and Remote Sensing, 36, 1228–1249, https://doi.org/10.1109/36.701075, 1998.

Karel, T. H. and Man, G.: Major Forest Insect and Disease Conditions in the United States: 2015, Tech. Rep. FS-1093, U.S. Department of Agriculture, Forest Service, Washington, DC, 2017.

Kautz, M., Meddens, A. J. H., Hall, R. J., and Arneth, A.: Biotic disturbances in Northern Hemisphere forests-a synthesis of recent data, uncertainties and implications for forest monitoring and modelling, Global Ecology and Biogeography, 26, 533–552, https://doi.org/10.1111/geb.12558, 2017.

Kautz, M., Feurer, J., and Adler, P.: Early detection of bark beetle (*Ips typographus*) infestations by remote sensing – A critical review of recent research, Forest Ecology and Management, 556, 121 595, https://doi.org/10.1016/j.foreco.2023.121595, 2024.

Kellogg, K., Hoffman, P., Standley, S., Shaffer, S., Rosen, P., Edelstein, W., Dunn, C., Baker, C., Barela, P., Shen, Y., Shah, A. M., Xaypraseuth, P., Sagi, V. R., Sreekantha, C. V., Harinath, N., Kumar, R., Bhan, R., and Sarma, C. V. H. S.: NASA-ISRO Synthetic Aperture Radar (NISAR) Mission, in: 2020 IEEE Aerospace Conference, pp. 1–21, IEEE, https://doi.org/10.1109/AERO47225.2020.9172638, 2020.

Kennedy, R. E., Andréfouët, S., Cohen, W. B., Gómez, C., Griffiths, P., Hais, M., Healey, S. P., Helmer, E. H., Hostert, P., Lyons, M. B., Meigs, G. W., Pflugmacher, D., Phinn, S. R., Powell, S. L., Scarth, P., Sen, S., Schroeder, T. A., Schneider, A., Sonnenschein, R., Vogelmann, J. E., Wulder, M. A., and Zhu, Z.: Bringing an ecological view of change to Landsat-based remote sensing, Frontiers in Ecology and the Environment, 12, 339–346, https://doi.org/10.1890/130066, 2014.

Key, C. H. and Benson, N. C.: Landscape assessment: Ground measure of severity, the Composite Burn Index; and remote sensing of severity, the Normalized Burn Ratio, in: FIREMON: Fire Effects Monitoring and Inventory System, edited by Lutes, D. C., Keane, R. E., Caratti, J. F., Key, C. H., Benson, N. C., Sutherland, S., and Gangi, L. J., General Technical Report RMRS-GTR-164-CD, pp. LA1–LA51, USDA Forest Service, Rocky Mountain Research Station, Ogden, Utah, USA, 2006.

Kislov, D. E., Korznikov, K. A., Altman, J., Vozmishcheva, A. S., and Krestov, P. V.: Extending deep learning approaches for forest disturbance segmentation on very high-resolution satellite images, Remote Sensing in Ecology and Conservation, 7, 355–368, https://doi.org/10.1002/rse2.194, 2021.

Li, Y., Wulder, M. A., Zhu, Z., Verbesselt, J., Masiliūnas, D., Liu, Y., Bohrer, G., Cai, Y., Zhou, Y., Ding, Z., and Zhao, K.: DR-MAT: A multivariate algorithm for detecting breakpoints in multispectral time series, Remote Sensing of Environment, 315, 114 402, https://doi.org/10.1016/j.rse.2024.114402, 2024.

Martini, D., Sakowska, K., Wohlfahrt, G., Pacheco-Labrador, J., van der Tol, C., Porcar-Castell, A., Magney, T. S., Carrara, A., Colombo, R., El-Madany, T. S., and others: Heatwave breaks down the linearity between sun-induced fluorescence and gross primary production, New phytologist, 233, 2415–2428, 2022.

Marvasti-Zadeh, S. M., Goodsman, D., Ray, N., and Erbilgin, N.: Early Detection of Bark Beetle Attack Using Remote Sensing and Machine Learning: A Review, ACM Computing Surveys, 56, 1–40, https://doi.org/10.1145/3625387, 2024.

Marzo, T. D., Gasparri, N., Lambin, E., and Kuemmerle, T.: Agents of Forest Disturbance in the Argentine Dry Chaco, Remote. Sens., 14, 1758, https://doi.org/10.3390/rs14071758, 2022.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Matisone, I., Matisons, R., Jansone, D., and Liepiņa, A. A.: Large-Scale Post-Storm Salvage Logging Shows Transient Effects on Vegetation in Managed Hemiboreal Forest, Resembling Those of Conventional Wood Harvesting in the Long Term, Conservation, 6, 23, https://doi.org/10.3390/conservation6010023, 2026.

McDowell, N. G., Coops, N. C., Beck, P. S., Chambers, J. Q., Gangodagamage, C., Hicke, J. A., Huang, C.-y., Kennedy, R., Krofcheck, D. J., Litvak, M., Meddens, A. J., Muss, J., Negrón-Juarez, R., Peng, C., Schwantes, A. M., Swenson, J. J., Vernon, L. J., Williams, A. P., Xu, C., Zhao, M., Running, S. W., and Allen, C. D.: Global satellite monitoring of climate-induced vegetation disturbances, Trends in Plant Science, 20, 114–123, http://www.sciencedirect.com/science/article/pii/S1360138514002726, 2015.

Meddens, A. J., Hicke, J. A., Vierling, L. A., and Hudak, A. T.: Evaluating methods to detect bark beetle-caused tree mortality using single-date and multi-date Landsat imagery, Remote Sensing of Environment, 132, 49–58, https://doi.org/10.1016/j.rse.2013.01.002, 2013.

Mercado, J. E., Walker, R. T., Franklin, S., Kay, S., Gomez, S. K., and Hubbard, R.: Changes in Transpiration and Leaf Water Potential in Douglas-Fir Trees following Douglas-Fir Beetle Attack and Mechanical Girdling, Forests, https://doi.org/10.3390/f12121722, 2021.

Meyer, H. and Pebesma, E.: Predicting into unknown space? Estimating the area of applicability of spatial prediction models, Methods in Ecology and Evolution, 12, 1620–1633, https://doi.org/10.1111/2041-210X.13650, 2021.

Minařík, R., Langhammer, J., and Lendzioch, T.: Automatic Tree Crown Extraction from UAS Multispectral Imagery for the Detection of Bark Beetle Disturbance in Mixed Forests, Remote Sensing, 12, 4081, https://doi.org/10.3390/rs12244081, 2020.

Minařík, R., Langhammer, J., and Lendzioch, T.: Detection of Bark Beetle Disturbance at Tree Level Using UAS Multispectral Imagery and Deep Learning, Remote Sensing, 13, 4768, https://doi.org/10.3390/rs13234768, 2021.

Mitchell, S. J.: Wind as a natural disturbance agent in forests: a synthesis, Forestry, 86, 147–157, https://doi.org/10.1093/forestry/cps058, 2013.

Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovský, Z., Gastellu-Etchegorry, J.-P., Miller, J. R., Guanter, L., Moreno, J., Moya, I., Berry, J. A., Frankenberg, C., and Zarco-Tejada, P. J.: Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress, Remote Sens. Environ., 231, 111 177, 2019.

Montero, D., Aybar, C., Mahecha, M. D., Martinuzzi, F., S"ochting, M., and Wieneke, S.: A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research, Scientific Data, 10, 197, 2023.

Morley, P., Jump, A., West, M. D., and Donoghue, D.: Spectral response of chlorophyll content during leaf senescence in European beech trees, Environmental Research Communications, 2, https://doi.org/10.1088/2515-7620/aba7a0, 2020.

Mosig, C.: cmosig/sentle, https://github.com/cmosig/sentle, original-date: 2024-01-23T13:18:49Z, 2024.

Müller, F., Eifler, L., Cremer, F., Beck, P. S. A., Camps-Valls, G., and Bastos, A.: Sentinel-1 disturbance map, 2026.

Müller, F., Eifler, L., Cremer, F., Beck, P., Camps-Valls, G., and Bastos, A.: Hybrid forest disturbance classification using Sentinel-1 and inventory data: a case-study for Southeastern USA, Natural Hazards and Earth System Sciences, 26, 2785–2815, https://doi.org/10.5194/nhess-26-2785-2026, 2026.

Nowosad, J., Bonannella, C., Görgen, D., Jemeljanova, M., Kattenborn, T., Linnenbrink, J., Meyer, H., Nussbaum, M., Patelli, L., Simoes, R., and Uuemaa, E.: Navigating challenges in spatial machine learning: Validation, uncertainty, algorithms, and reproducibility, ERDKUNDE, https://doi.org/10.3112/erdkunde.2026.04.01, 2026.

Pan, Y., Birdsey, R. A., Phillips, O. L., Houghton, R. A., Fang, J., Kauppi, P. E., Keith, H., Kurz, W. A., Ito, A., Lewis, S. L., Nabuurs, G.-J., Shvidenko, A., Hashimoto, S., Lerink, B., Schepaschenko, D., Castanho, A., and Murdiyarso, D.: The enduring world forest carbon sink, Nature, 631, 563–569, https://doi.org/10.1038/s41586-024-07602-x, number: 8021, 2024.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Patacca, M., Lindner, M., Lucas-Borja, M. E., Cordonnier, T., Fidej, G., Gardiner, B., Hauf, Y., Jasinevičius, G., Labonne, S., Linkevičius, E., Mahnken, M., Milanovic, S., Nabuurs, G.-J., Nagel, T. A., Nikinmaa, L., Panyatov, M., Bercak, R., Seidl, R., Ostrogović Sever, M. Z., Socha, J., Thom, D., Vuletic, D., Zudin, S., and Schelhaas, M.-J.: Significant increase in natural disturbance impacts on European forests since 1950, Global Change Biology, 29, 1359–1376, https://doi.org/10.1111/gcb.16531, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.16531, 2023.

Petit, G., Netherer, S., Biruk, L., Candotti, A., Casolo, V., Faccoli, M., Hölttä, T., Losso, A., Mayr, S., Paljakka, T., Petruzzellis, F., Pitacco, A., Prendin, A., Riggins, J., Savi, T., Schuler, H., Squartini, A., Tomelleri, E., Battisti, A., and Zapponi, L.: From Colonization to Collapse: a Mechanistic Perspective on Tree Mortality by Bark Beetles, Current Forestry Reports, https://doi.org/10.1007/s40725-026-00279-7, 2026.

Pierdicca, N., Davidson, M., Chini, M., Dierking, W., Djavidnia, S., Haarpaintner, J., Hajduch, G., Laurin, G. V., Lavalle, M., López-Martínez, C., Nagler, T., and Su, B.: The Copernicus L-band SAR mission ROSE-L (Radar Observing System for Europe), in: Active and Passive Microwave Remote Sensing for Environmental Monitoring III, SPIE, 2019.

Quegan, S., Le Toan, T., Chave, J., Dall, J., Exbrayat, J.-F., Minh, D. H. T., Lomas, M., D'Alessandro, M. M., Paillou, P., Papathanassiou, K., et al.: The European Space Agency BIOMASS mission: Measuring forest above-ground biomass from space, Remote Sensing of Environment, 227, 44–60, https://doi.org/10.1016/j.rse.2019.03.032, 2019.

Rajwa-Kuligiewicz, A. and Bojarczuk, A.: Streamflow response to catastrophic windthrow and forest recovery in subalpine spruce forest, Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2024.131078, 2024.

Rodríguez Paulino, E., Schlerf, M., Röder, A., Stoffels, J., and Udelhoven, T.: Forest disturbance characterization in the era of earth observation big data: A mapping review, International Journal of Applied Earth Observation and Geoinformation, 128, 103 755, https://doi.org/10.1016/j.jag.2024.103755, 2024.

Rouse, J., Haas, R. H., Schell, J. A., and Deering, D.: Monitoring vegetation systems in the great plains with ERTS, https://ntrs.nasa.gov/citations/19740022614, 1973.

Roy, D. P., Wulder, M. A., Loveland, T. R., et al.: Landsat-8: Science and product vision for terrestrial global change research, Remote Sensing of Environment, 145, 154–172, https://doi.org/10.1016/j.rse.2014.02.001, 2014.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, Nature, 323, 533–536, https://doi.org/10.1038/323533a0, 1986.

Rüetschi, M., Small, D., and Waser, L. T.: Rapid Detection of Windthrows Using Sentinel-1 C-Band SAR Data, Remote Sensing, 11, 115, https://doi.org/10.3390/rs11020115, 2019.

Sakoe, H. and Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition, IEEE Transactions on Acoustics, Speech, and Signal Processing, 26, 43–49, https://doi.org/10.1109/TASSP.1978.1163055, 1978.

Savitzky, A. and Golay, M. J. E.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures., Analytical Chemistry, 36, 1627–1639, https://doi.org/10.1021/ac60214a047, 1964.

Schäfer, P.: The BOSS is concerned with time series classification in the presence of noise, Data Mining and Knowledge Discovery, 29, 1505–1530, https://doi.org/10.1007/s10618-014-0377-7, 2014.

Schleeweis, K. G., Moisen, G. G., Schroeder, T. A., Toney, C., Freeman, E. A., Goward, S. N., Huang, C., and Dungan, J. L.: US National Maps Attributing Forest Change: 1986–2010, Forests, 11, 653, https://doi.org/10.3390/f11060653, 2020.

Schroeder, T. A., Schleeweis, K. G., Moisen, G. G., Toney, C., Cohen, W. B., Freeman, E. A., Yang, Z., and Huang, C.: Testing a Landsat-based approach for mapping disturbance causality in U.S. forests, Remote Sensing of Environment, 195, 230–243, https://doi.org/10.1016/j.rse.2017.03.033, 2017.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

Schwartz, N., Uriarte, M., DeFries, R., Bedka, K., Fernandes, K., Gutiérrez-Vélez, V., and Pinedo-Vasquez, M.: Fragmentation increases wind disturbance impacts on forest structure and carbon stocks in a western Amazonian landscape., Ecological applications : a publication of the Ecological Society of America, 27 6, 1901–1915, https://doi.org/10.1002/eap.1576, 2017.

Sebald, J., Senf, C., and Seidl, R.: Human or natural? Landscape context improves the attribution of forest disturbances mapped from Landsat in Central Europe, Remote Sensing of Environment, 262, 112 502, https://doi.org/10.1016/j.rse.2021.112502, 2021.

Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., Wild, J., Ascoli, D., Petr, M., Honkaniemi, J., Lexer, M. J., Trotsiuk, V., Mairota, P., Svoboda, M., Fabrika, M., Nagel, T. A., and Reyer, C. P.: Forest disturbances under climate change, Nature Climate Change, 7, 395–402, https://doi.org/10.1038/NCLIMATE3303, 2017.

Senf, C. and Seidl, R.: Natural disturbances are spatially diverse but temporally synchronized across temperate forest landscapes in Europe, Global Change Biology, 24, 1201–1211, https://doi.org/10.1111/gcb.13897, 2018.

Senf, C. and Seidl, R.: Storm and fire disturbances in Europe: Distribution and trends, Global Change Biology, 27, 3605–3619, https://doi.org/10.1111/gcb.15679, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.15679, 2021.

Senf, C. and Seidl, R.: Post-disturbance canopy recovery and the resilience of Europe's forests, Global Ecology and Biogeography, 31, 25–36, https://doi.org/10.1111/geb.13406, 2022.

Senf, C., Pflugmacher, D., Wulder, M. A., and Hostert, P.: Characterizing spectral–temporal patterns of defoliator and bark beetle disturbances using Landsat time series, Remote Sensing of Environment, 170, 166–177, https://doi.org/10.1016/j.rse.2015.09.019, 2015.

Shao, Z., Cai, J., Fu, P., Hu, L., and Liu, T.: Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product, Remote Sensing of Environment, 235, 111 425, https://doi.org/10.1016/j.rse.2019.111425, 2019.

Sigurdsson, J., Armannsson, S. E., Ulfarsson, M. O., and Sveinsson, J. R.: Fusing Sentinel-2 and Landsat 8 Satellite Images Using a Model-Based Method, Remote Sensing, 14, 3224, https://doi.org/10.3390/rs14133224, 2022.

Stahl, A. T., Andrus, R., Hicke, J. A., Hudak, A. T., Bright, B. C., and Meddens, A. J. H.: Automated attribution of forest disturbance types from remote sensing data: A synthesis, Remote Sensing of Environment, 285, 113 416, https://doi.org/10.1016/j.rse.2022.113416, 2023.

Stephens, S. S., Romero, S. A., and Krist, F. J.: Major Forest Insect and Disease Conditions in the United States: 2022, Tech. Rep. FS-1219, U.S. Department of Agriculture, Forest Service, Washington, DC, 2023.

Storch, T., Honold, H.-P., Chabrillat, S., Habermeyer, M., Tucker, P., Brell, M., Ohndorf, A., Wirth, K., Betz, M., Kuchler, M., Mühle, H., Carmona, E., Baur, S., Mücke, M., Löw, S., Morgenroth, D., Zimmermann, S., Lenzen, C., Wiesner, S., Aida, S., Kahle, R., Willburger, P., Hartung, S., Dietrich, D., Plesia, N., Tegler, M., Schork, K., Alonso, K., Marshall, D., Gerasch, B., Schwind, P., Pato, M., Schneider, M., de los Reyes, R., Langheinrich, M., Wenzel, J., Bachmann, M., Holzwarth, S., Pinnel, N., Guanter, L., Segl, K., Scheffler, D., Foerster, S., Bohn, N., Bracher, A., Soppa, M. A., Gascon, F., Green, R., Kokaly, R., Moreno, J., Ong, C., Sornig, M., Wernitz, R., Bagschik, K., Reintsema, D., La Porta, L., Schickling, A., and Fischer, S.: The EnMAP imaging spectroscopy mission towards operations, Remote Sens. Environ., 294, Not Available, 2023.

Tagliabue, G., Boschetti, M., Bramati, G., Candiani, G., Colombo, R., Nutini, F., Pompilio, L., Rivera-Caicedo, J. P., Rossi, M., Rossini, M., Verrelst, J., and Panigada, C.: Hybrid retrieval of crop traits from multi-temporal PRISMA hyperspectral imagery, ISPRS J. Photogramm. Remote Sens., 187, 362–377, 2022.

Thom, D. and Seidl, R.: Natural disturbance impacts on ecosystem services and biodiversity in temperate and boreal forests, Biological Reviews, 91, 760–781, https://doi.org/10.1111/brv.12193, 2016.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., et al.: GMES Sentinel-1 mission, Remote Sensing of Environment, 120, 9–24, https://doi.org/10.1016/j.rse.2011.05.028, 2012.

https://doi.org/10.5194/egusphere-2026-7087
Preprint. Discussion started: 23 November 2026
© Author(s) 2026. CC BY 4.0 License.

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

Urquiza-Muñoz, J. D., Negrón-Juárez, R. I., et al.: Windthrows Inventory Data Base (WInD), Version 1.0, https://doi.org/10.5281/zenodo.11168104, 2024a.

Urquiza-Muñoz, J. D., Trumbore, S., Negrón-Juárez, R. I., et al.: Increased Occurrence of Large-Scale Windthrows Across the Amazon Basin, AGU Advances, 5, https://doi.org/10.1029/2023AV001030, 2024b.

U.S. Department of Agriculture, Forest Service: Forest Service Regional Boundaries (Feature Layer), https://data-usfs.hub.arcgis.com/datasets/usfs::forest-service-regional-boundaries-feature-layer/about, [Data set]. Accessed: 2026-08-04, n.d.

US Forest Service: US Forest Service - FIA Forest Types of the Southeastern United States | Data Basin, https://databasin.org/datasets/e874a247b7e44693ad0792fb1a252aba/, 2013.

van Eupen, F., Kamionka-Kanclerska, K., Żmihorski, M., Ottenburghs, J., and Walesiak, M.: Effect of bark beetle outbreak and salvage logging on tree-related microhabitats in Białowieża Forest, Forest Ecology and Management, 586, 122 711, https://doi.org/10.1016/j.foreco.2025.122711, 2025.

Velichkova, K. and Krezhova, D.: Extraction of the red edge position from hyperspectral reflectance data for plant stress monitoring, 2075, 130 018, https://doi.org/10.1063/1.5091303, 2019.

Viana-Soto, A. and Senf, C.: The European Forest Disturbance Atlas: a forest disturbance monitoring system using the Landsat archive, Earth System Science Data, 17, 2373–2404, https://doi.org/10.5194/essd-17-2373-2025, 2025.

Viehweger, J., Hirschmugl, M., Puhm, M., Burtscher, R., Gallaun, H., and Janik Deutscher: Dense labelled time-series for mapping European forest disturbance agents, 2026.

Wang, L., Liu, S., Song, W., Ding, S., and Zhang, J.: Global high-resolution forest disturbance type dataset, Earth System Science Data, 18, 1601–1617, https://doi.org/10.5194/essd-18-1601-2026, 2026.

White, P. S. and Pickett, S. T. A.: Chapter 1 - Natural Disturbance and Patch Dynamics: An Introduction, in: The Ecology of Natural Disturbance and Patch Dynamics, edited by Pickett, S. T. A. and White, P. S., pp. 3–13, Academic Press, San Diego, https://doi.org/10.1016/B978-0-12-554520-4.50006-X, 1985.

Wilson, E. H. and Sader, S. A.: Detection of forest harvest type using multiple dates of Landsat TM imagery, Remote Sensing of Environment, 80, https://doi.org/10.1016/s0034-4257(01)00318-2, 2002.

Wittich, D., Rottensteiner, F., Voelsen, M., Heipke, C., and Müller, S.: DEEP LEARNING FOR THE DETECTION OF EARLY SIGNS FOR FOREST DAMAGE BASED ON SATELLITE IMAGERY, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, V-2-2022, 307–315, https://doi.org/10.5194/isprs-annals-V-2-2022-307-2022, 2022.

Zhai, L., Coyle, D. R., Li, D., and Jonko, A.: Fire, insect and disease-caused tree mortalities increased in forests of greater structural diversity during drought, J. Ecol., 110, 673–685, 2022.

Zheng, a. c.-a.: A Novel Framework for Long-Term Forest Disturbance Monitoring: Synergizing the LandTrendr Algorithm with CNN in Northeast China, Remote Sensing, 17, 3521, https://doi.org/10.3390/rs17213521, 2025.