ForestPaths

Co-designing holistic forest-based policy pathways for climate change mitigation

D2.5 Map data products

03/09/2024

Author(s): Cornelius Senf, Alba Viana-Soto, Wanda De Keersmaecker, Luc Bertels & Ruben Van De Kerchove

Prepared under contract from the European Commission and the United Kingdom Research and Innovation Council.
Grant agreement No. 101056755
EU Horizon Europe Research and Innovation Action

Project acronym:	**ForestPaths**
Project full title:	**Co-designing Holistic Forest-based Policy Pathways for Climate Change Mitigation**
Project duration:	01.09.2022 – 28.02.2027 (54 months)
Project coordinator:	Dr. Hans Verkerk, European Forest Institute (EFI)
Call:	HORIZON-CL5-2021-D1-01
Deliverable title:	Map data products
Deliverable n°:	D2.5
WP responsible:	WP2
Nature of the deliverable:	Report
Dissemination level:	Public
Lead partner:	TUM
Recommended citation:	Cornelius Senf, Alba Viana-Soto, Wanda De Keersmaecker, Luc Bertels & Ruben Van De Kerchove (2024). ***Map data products***. ForestPaths project deliverable D2.5.
Due date of deliverable:	Month 24
Actual submission date:	03.09.2024

Deliverable status:

Version	Status	Date	Author(s)
1.0	Draft	21.08.24	Cornelius Senf, Alba Viana-Soto (TUM) & Wanda De Keersmaecker, Luc Bertels, Ruben Van De Kerchove (VITO)
	Review	28.08.24	Mikko Peltoniemi (LUKE) & Hans Verkerk (EFI)
2.0	Final	03.09.24	Cornelius Senf (TUM)

Table of contents

Key takeaway messages

- Forest disturbances for wind and bark beetle, fire, and harvest mapped from Landsat data at 30 m spatial resolution since 1985
- Quantification of map uncertainties shows improvement over existing forest disturbance map products
- Forest composition (genus) and structure (canopy cover and heights, foliage height diversity) mapped from Sentinel-1/2 data at 10 m spatial resolution for the year 2020
- Comparison to existing forest composition and structure datasets shows higher accuracy

Summary

Remote sensing offers novel opportunities for monitoring forest status and change, but methodological challenges still exist. In ForestPaths, we fill this gap by providing maps on forest disturbances (D2.1) and forest composition and structure (D2.2). We here provide the technical background and validation of the map data products. First, we deliver a detailed description of the next-generation forest disturbance map, produced from Landsat data at 30 m spatial resolution. We also performed an in-depth validation through space and time. The next-generation forest disturbance map includes annual layers of forest disturbance, disturbance agent and disturbance severity, and summary layers across the full period. The validation revealed high accuracies compared to existing products, but inconsistencies in accuracies over time, which calls for caution in deriving trends from the maps. We discuss potential solutions to this problem. We also discuss how the new map contains multiple disturbances per pixel, which was absent in past products, and likely has led to underestimation of disturbances in past reports. Second, we provide a detailed description of the forest structure maps created from Sentinel-1/2 data at 10 m spatial resolution. Validation includes a comparison to an airborne laser scanning-based reference database covering large parts of Europe. The comparison revealed better performance of the new map compared to existing maps. Finally, we provide a detailed description of the composition maps generated from Sentinel-1/2, also provided at 10 m spatial resolution. The maps were trained and validated on a pan-Europe sample of plots. Initial tests showed lower skill in separating species, but moderate to high skill in separating genera. We thus proceeded with mapping genera of the most important tree species across Europe. All in all, the three map products described here set a new standard for remote sensing-based forest monitoring in Europe and provide the most accurate maps of forest disturbances, forest structure and forest composition to date.

1 General introduction

Europe's forests cover more than one third of the continent and they provide essential ecosystem services to society (FOREST EUROPE, 2020), spanning from timber production and carbon storage (Lindner et al., 2010) over water purification and regulation (Orsi et al., 2020) to recreation and spiritual value (Saarikoski et al., 2015). Europe's forests are shaped by a long history of forest management, with past management decisions driving their current structure and composition (Ciais et al., 2008; Sabatini et al., 2018; Seidl & Senf, 2024). Past forest management has also focussed on increasing forest resistance to natural disturbances (Seidl, 2014), sustaining a continuous provision of timber and other ecosystem services (Thom & Seidl, 2016). In the recent years, however, evidence for changing natural disturbance regimes has been reported globally (McDowell et al., 2020) and in Europe, climate change and climate extremes have led to increased natural disturbances (Patacca et al., 2023; Senf et al., 2018; Senf & Seidl, 2021b). The trend of increasing natural disturbances coincides with an increasing demand for wood globally, with reported increases in timber harvest rates observed across most of Europe (Breidenbach et al., 2022; Ceccherini et al., 2020; Palahí et al., 2021; Seidl & Senf, 2024). Forests in Europe are thus undergoing complex changes that require a fundamental monitoring of where and when disturbances have occurred in the past, and up-to-date information on their current structure and compositions. Remote sensing can provide such information, but there is yet a lack of operational methods for monitoring forest disturbance and mapping forest composition and structure. With this deliverable we will fill this gap by outlining the technical details and validation of the next generation disturbance map published in D2.1 and the forest structure and composition maps published in D2.2.

2 Next generation disturbance maps

2.1 Introduction

Remote sensing data can provide detailed and consistent information on forests disturbances (Hirschmugl et al., 2017; Senf, 2022), and the Landsat archive (Wulder et al., 2022) – covering more than four decades – plays a key role in this regard (Banskota et al., 2014), because it covers longer time frames than other optical satellites archives (e.g. Sentinel-2). In particular, Landsat contributed to a better understanding of the causes and consequences of changing disturbance regimes in Europe (Grünig et al., 2022; Lecina-Diaz et al., 2024; Stritih et al., 2021) and on disturbance interactions (Buma, 2015; Hermosilla et al., 2019; Turner & Seidl, 2023). Since the opening of the Landsat archive, different methodologies and algorithms have emerged for monitoring forest cover change across large areas (Zhu, 2017), with a regional focus on forests of the United States and Canada (Kennedy et al., 2010; Verbesselt et al., 2010; Zhao et al., 2019; Zhu & Woodcock, 2014). Such forest change detection algorithms can broadly be grouped into four types: (1) trajectory segmentation approaches (Hughes et al., 2017; Kennedy et al., 2010; Moisen et al., 2016), (2) time series decomposition approaches (Verbesselt et al., 2010; Zhao et al., 2019), (3) threshold-based methods (Huang et al., 2010), and (4) classification approaches (Hansen et al., 2013; Hermosilla et al., 2015). Time series segmentation or decomposition methods rely on the detection of statistical breakpoints in a spectral time series to detect forest change. While easy to implement and requiring little to no reference data, such approaches make use of only a limited set of spectral characteristics (usually one spectral index) and they require re-calibration when new observations are incorporated (Hermosilla et al., 2017). Threshold-based methods, in turn, can be easily updated with new data but they rely on simple thresholds that might be difficult to apply across different ecosystems, because thresholds will vary depending

on forest composition and site conditions (Cardille et al., 2022). Classification-based methods, particularly machine learning-based approaches (Belgiu & Drăguţ, 2016), can capture complex patterns and adapt to different types of disturbances and varying environmental conditions (Cardille et al., 2022). This makes them more flexible to map changes across large areas (Hansen et al., 2013). Classification-based approaches also apply a strict definition of disturbances (i.e. by a well-defined classification label) and they can reliably reduce commission errors by filtering out falsely detected disturbances resulting from noise in the time series (W. Cohen et al., 2017; Hermosilla et al., 2015). Finally, annual classification approaches can also facilitate the detection of more than one disturbance per pixel time series, which can be challenging with trajectory-based approaches (Hermosilla et al., 2015).

Due to the rapid development of new remote sensing approaches, there is growing interest to also develop operational forest monitoring systems at the level of the European Union (Fassnacht et al., 2024; Ferretti, 2024; Ferretti et al., 2024; Nabuurs et al., 2022) and in particular for the monitoring of tree cover change and disturbance (Dutrieux et al., 2023; European Commission et al., 2023). As global tree cover change products contain high uncertainty when analyzed regionally (Ceccherini et al., 2020, Breidenbach et al., 2022; Palahí et al., 2021), efforts have been made to develop forest disturbance monitoring approaches specifically for Europe (Francini et al., 2021; Senf & Seidl, 2021a; Turubanova et al., 2023). Francini et al. (2021) proposed an automated algorithm (3I3D) for mapping forest disturbances based on spectral change in three indices within three consecutive years. The approach has shown good results for harvest disturbance detection in Italy but was yet not applied at continental scale. Shortly after, Senf and Seidl (2021a) created the first pan-European characterization of forest disturbance by combining a trajectory-segmentation algorithm (LandTrendr; Kennedy et al., 2010) with a random forest classification approach. While delivering spatially consistent data across Europe, this approach cannot include multiple disturbances per pixel and recurring or interacting disturbance events thus cannot be detected. Turubanova et al. (2023) modelled changes in tree crown height across Europe from 2001 to 2021 by integrating Landsat imagery and Lidar data (ALS and GEDI). It constitutes the first attempt to analyse the evolution of tree canopy height at the European scale, but it only includes data from 2001 onwards and thus lacks a baseline for understanding more recently observed changes and for quantifying trends over time. As such, while there is a multitude of different approaches and datasets mapping forest disturbance across Europe, none of the existing approaches fulfils all requirements for an operational monitoring of forest change in Europe.

We here present the next-generation European forest disturbance map, which is a Landsat-based approach for mapping annual forest disturbances across continental Europe since 1985. The maps contain annual layers on disturbance occurrence, severity and agent, as well as aggregated layers on the latest and greatest disturbance year and on the number of disturbances. The aim of this report is to (i) explain and document the data and processing routines used for Landsat-based disturbance mapping, (ii) quantify uncertainties and errors in the disturbance maps, (iii) describe the individual map products derived from the disturbance maps, and (iv) discuss the use (and misuse) of the map products for forest monitoring.

2.2 Data and Methods

The overall workflow behind the next-generation forest disturbance map contains six processing steps summarized in Figure 1 and described in full detail in the following sub-sections: First, building a consistent Landsat data cube for Europe. Second, compiling a reference dataset on forest land use and forest disturbances. Third, creating a consistent forest land use mask. Fourth, developing a classification-based approach to detect disturbances annually. Fifth, validating the disturbances maps based on an independent reference sample. Sixth, creating a set of summary layers.

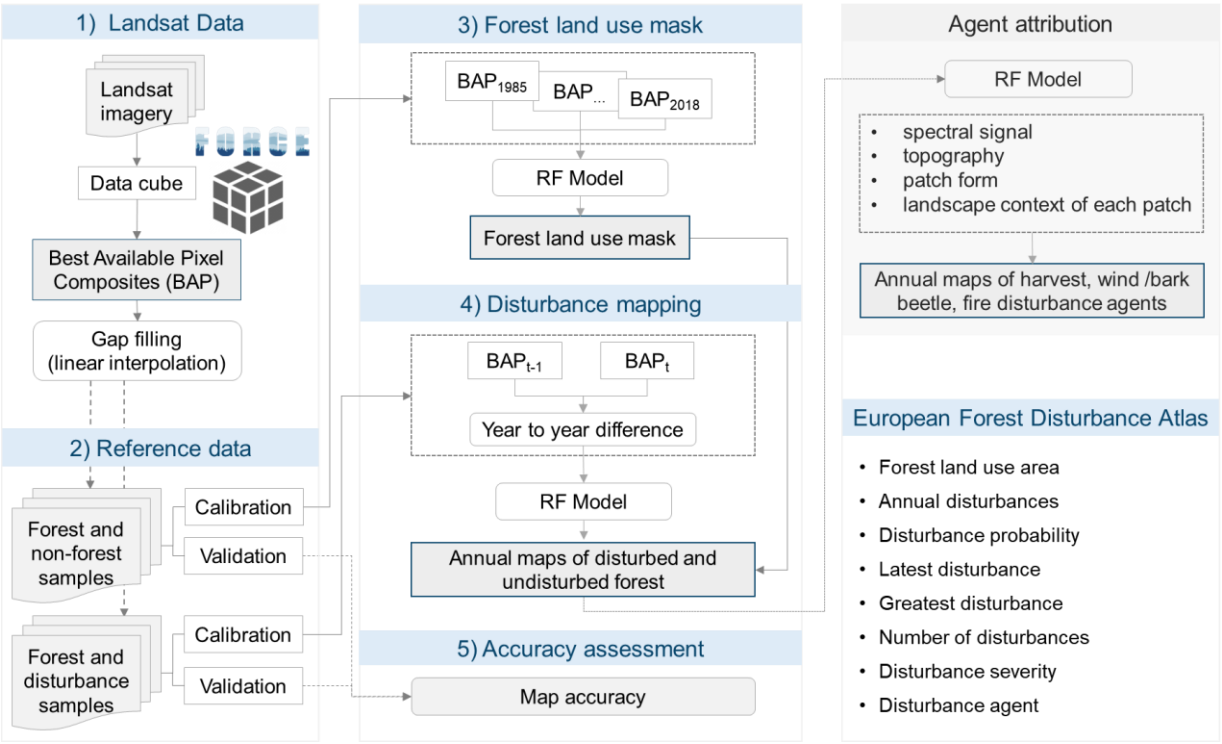

Figure 1. Overview of the workflow for creating the next-generation forest disturbance map.

2.2.1 Landsat data cube

Our analysis area covers continental Europe, which was defined as all European countries except for Russia, Malta, Cyprus (excluded due to their small size and low forest cover) and overseas territories and which covers a total of 5,749,424 km^2 of land area. For the analysis area, we identified all Level-1 images collected from Landsat 4 to Landsat 9 since 1984 and with a cloud cover below 60 %. We only searched for images within the growing season (defined as the period between 1st June and 30th September) to prevent differences in reflectance caused by phenological variations and sun angle changes. This resulted in a total of 115,663 images until 2023, which we downloaded from the United States Geological Survey (Figure 2a). The Level-1 Landsat images were further processed to surface reflectance (Level-2) and organized in a data cube structure of non-overlapping tiles of 150 x 150 km (Figure 2b) using the Framework for Operational Radiometric Correction for Environmental Monitoring (FORCE version 3.7.9, Frantz, 2019). Processing steps included atmospheric corrections using a precompiled water vapor database (Frantz et al., 2016), topographic correction using the ASTER Global Digital Elevation

Model Version 3 (Abrams et al., 2020), bidirectional reflectance distribution function corrections, as well as cloud and cloud shadow masking (Frantz et al., 2016; Roy et al., 2016; Zhu & Woodcock, 2012).

From the Level-2 data cube, we created annual, cloud-free best available pixel composites across the whole analysis area. Best available pixel compositing has been demonstrated to yield temporally and radiometrically consistent data for large area mapping (Hermosilla et al., 2022a), and recent research has shown their superiority in detecting disturbances compared to other temporal aggregation methods (Francini et al., 2023). For creating the best available pixel composites, we selected for each pixel the best observations based on a parametric weighting scheme established in previous research (Griffiths et al. 2013). Observations were ranked per pixel according to distance to clouds and cloud shadows, haze opacity and proximity to a predefined target date (1st of August; Figure 3). As there were still remaining gaps in the composites (i.e., areas where no high-quality observation was found during the summer season, 10.1% of the total area on average (see Figure 4 for annual details), we applied a linear gap-filling algorithm extrapolating the previous year's spectral values to fill remaining gaps (Figure 2c), reducing data gaps to 4.9% per year on average.

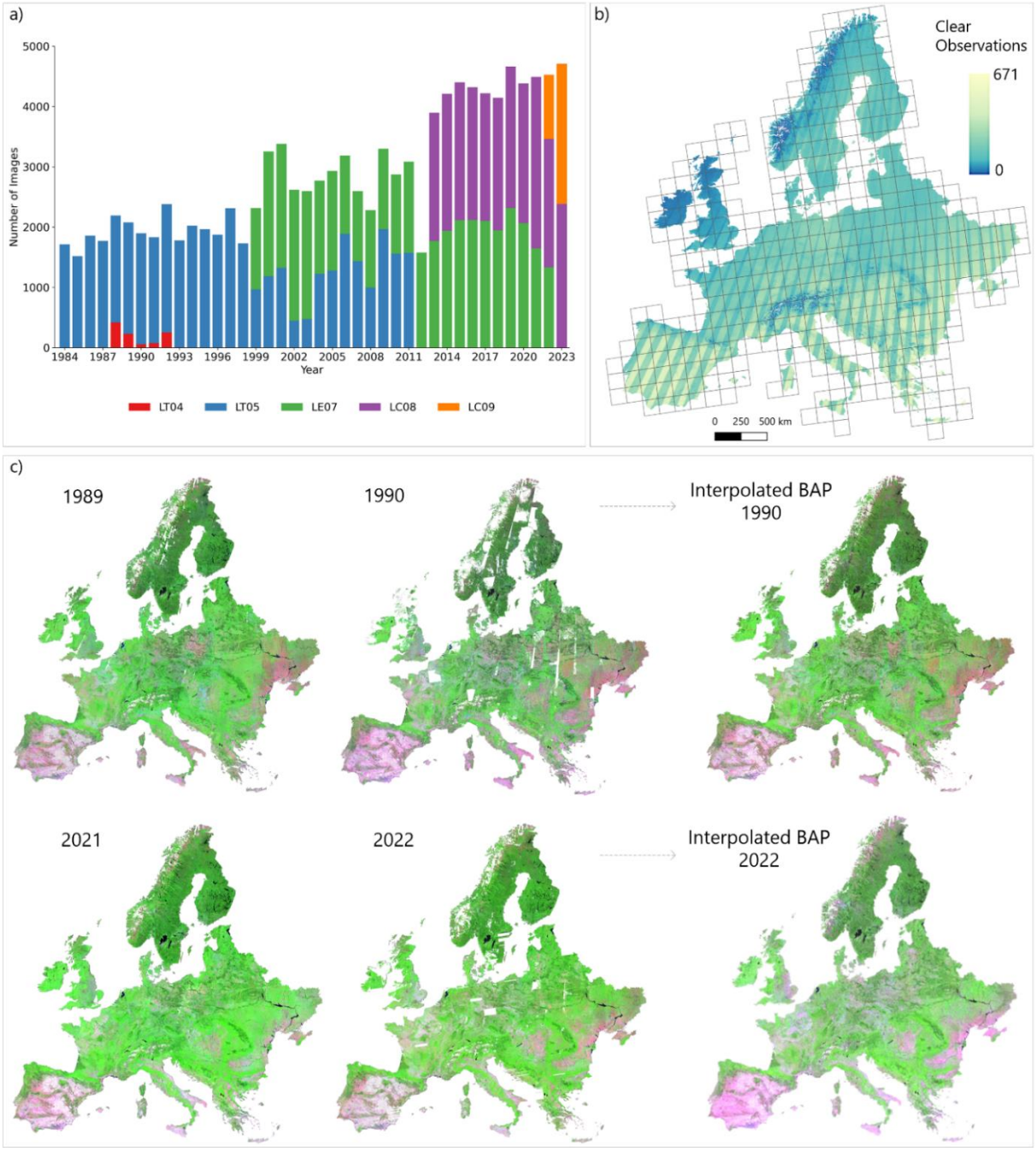

Figure 2. a) Number of images processed per year and satellite/sensor (LT04 = Landsat 4 TM, LT05 = Landsat 5 TM, LE07 = Landsat 7 ETM+, LC08 = Landsat 8 OLI, LC09 = Landsat 9 OLI2). b) Clear Sky Observations across Europe for the entire time series and tile system of the data cube. c) Examples of best available pixel (BAP) composites for Europe and resulting composites after gap-filling. The images are represented as false-colour composites in RGB-space with band 7 on red, band 5 on green and band 4 on blue.

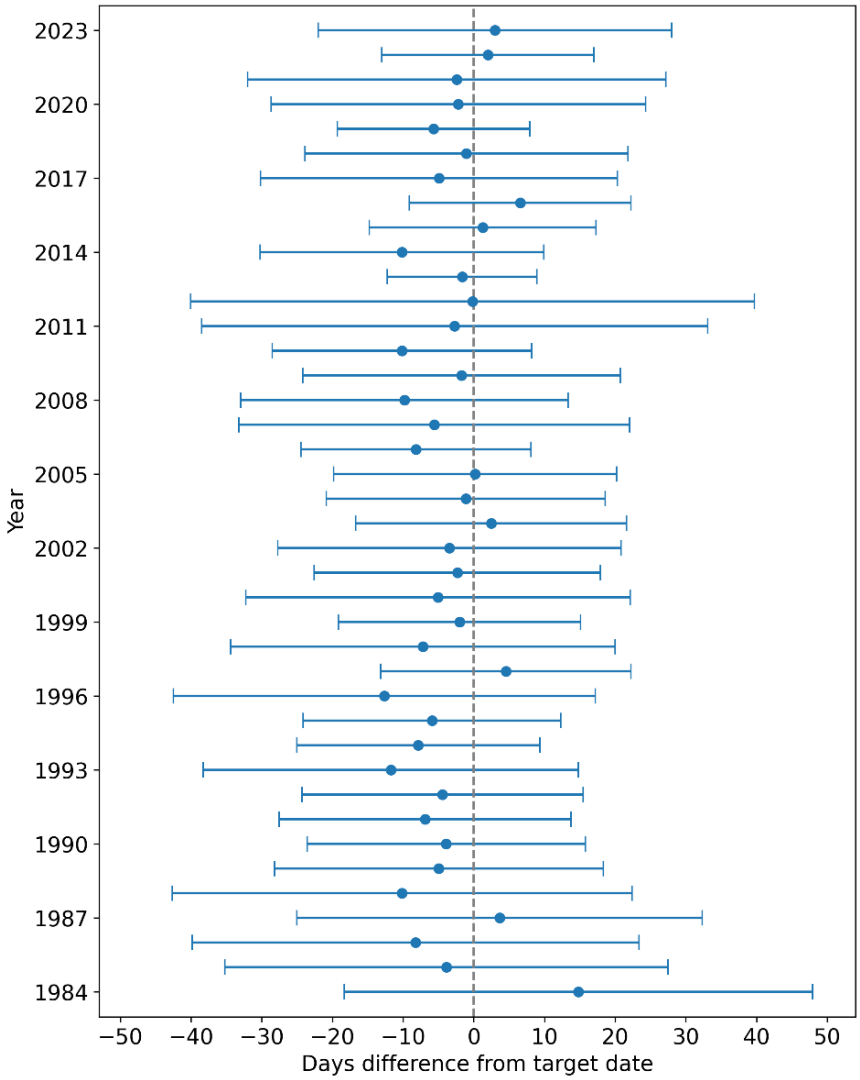

Figure 3. Days difference from target date (1st of August) per year (Mean +- Standard Deviation).

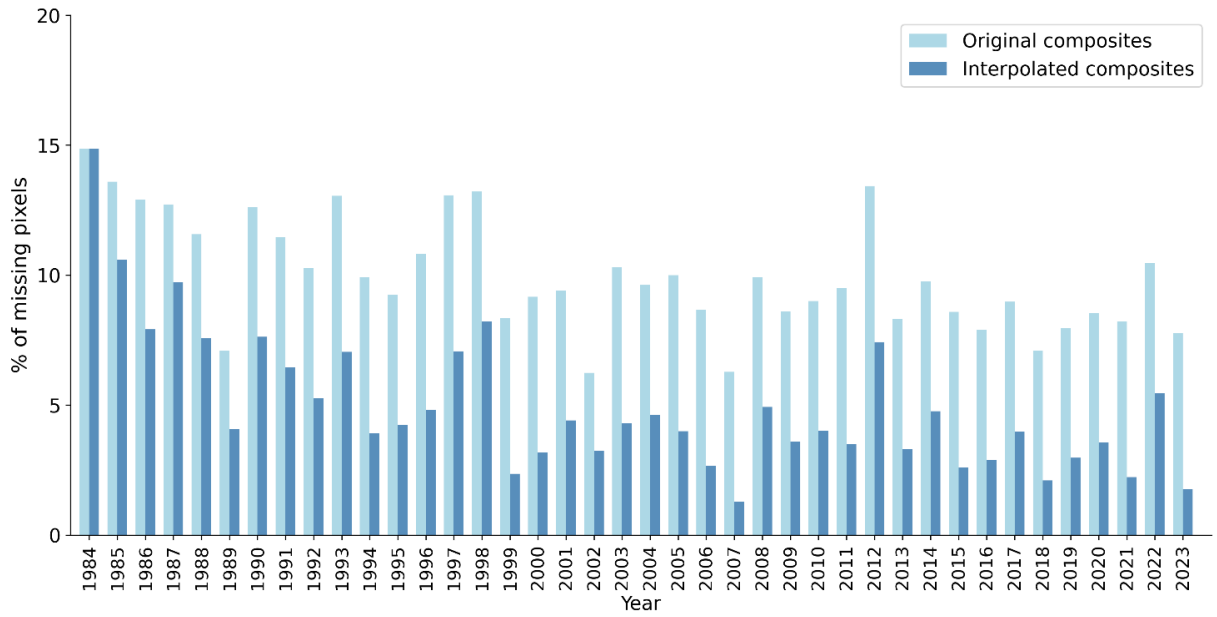

Figure 4. Annual percentage of no-data pixels for best available pixel composites and after linear gap-filling interpolation.

2.2.2 Reference data

We compiled reference data for both forest and non-forest land use pixels, used to train a classifier that discriminates forest from other land use, as well as for disturbance and undisturbed pixels, used to train a classifier that detects tree cover change within forest land use annually. For the latter, we used a previously established dataset of manually interpreted Landsat pixels for 35 countries available in https://zenodo.org/records/3561925. The dataset was built from two samples: The first one containing 24,000 randomly selected Landsat pixels for six countries across all land uses for the period 1985 to 2017 (Central Europe; Senf et al., 2018). The second one containing 500 randomly selected Landsat pixels for 29 countries (excluding Central Europe) within forest land use for the period 1985 to 2018 (Senf et al., 2021). After removing non-forest land use pixels from both samples, we were left with a total of 20,084 reference pixels within forest land use across Europe. For each pixel, trained interpreters segmented the spectral time series into linear segments of stable, disturbance and recovery (Figure 5) using an established interpretation tool (W. B. Cohen et al., 2010). The interpreters additionally recorded the land cover for each node of each segment (tree [>=50% tree cover] or non-treed [<50% tree cover]), which allowed us to disentangle stand-replacing disturbances (land cover changes from treed to non-treed) from non-stand-replacing disturbances (land cover stays treed despite disturbance). For full details on the interpretation, we direct the reader to the original publications (Senf et al., 2018, 2021). We converted the linear segments into annual binary information on disturbance occurrence (Figure 5), which yielded 662,772 data points (pixel-year combinations) to train an annual classification model described below.

For masking out all non-forest land use, we complemented the above sample within forest land use with non-forest land uses reference points from the LUCAS database (Land Use and Coverage Area frame Survey, Eurostat: https://ec.europa.eu/eurostat/web/lucas). The LUCAS database is a spatially explicit database using a stratified random sampling design and trained

field surveyors to assess land use and land cover (among other parameters) at 651,676 locations across Europe over the years 2006 to 2018 (d'Andrimont et al., 2020). LUCAS has been successfully used for land cover mapping (Pflugmacher et al., 2019) and thus is an ideal database for generating reference information for large-scale remote sensing analyses. From LUCAS, we extracted samples on the following categories: artificial land, croplands, grassland, bare land, water, and wetlands. We used only direct observations made at the plot in the field or via photointerpretation for which the land cover area was >1 ha, the land cover proportion was >80%, and plots for which the field-observed GPS location was <15m away from the LUCAS point. We collected a total of 46,651 non-forest reference pixels that way, matching the forest to non-forest land use ratio per country following statistics available in the FAOSTATS database (FOREST EUROPE, 2020). Doing so yields a random sample per country, i.e. the share of samples within and outside forest land use corresponds to the share of forest to non-forest land use of the country.

From both samples, a validation subsample of 2500 pixels each was retained for independent validation of the forest land use and forest disturbance classification. The land use validation sample was drawn proportionally to each country's land area and stratified by forest share within each country (see Table 1), which emulates a random sample across Europe's land area. That said, due to the absence of LUCAS information for 10 countries, the non-forest sample was slightly smaller (2066 samples) than the planned 2500 samples (see Table 1). The forest disturbance validation sample was drawn proportionally to each countries forest area, which emulates a random sample across all forests in Europe. Both samples are thus independent, random draws from the full population and allow for estimating unbiased map accuracies of the final map products.

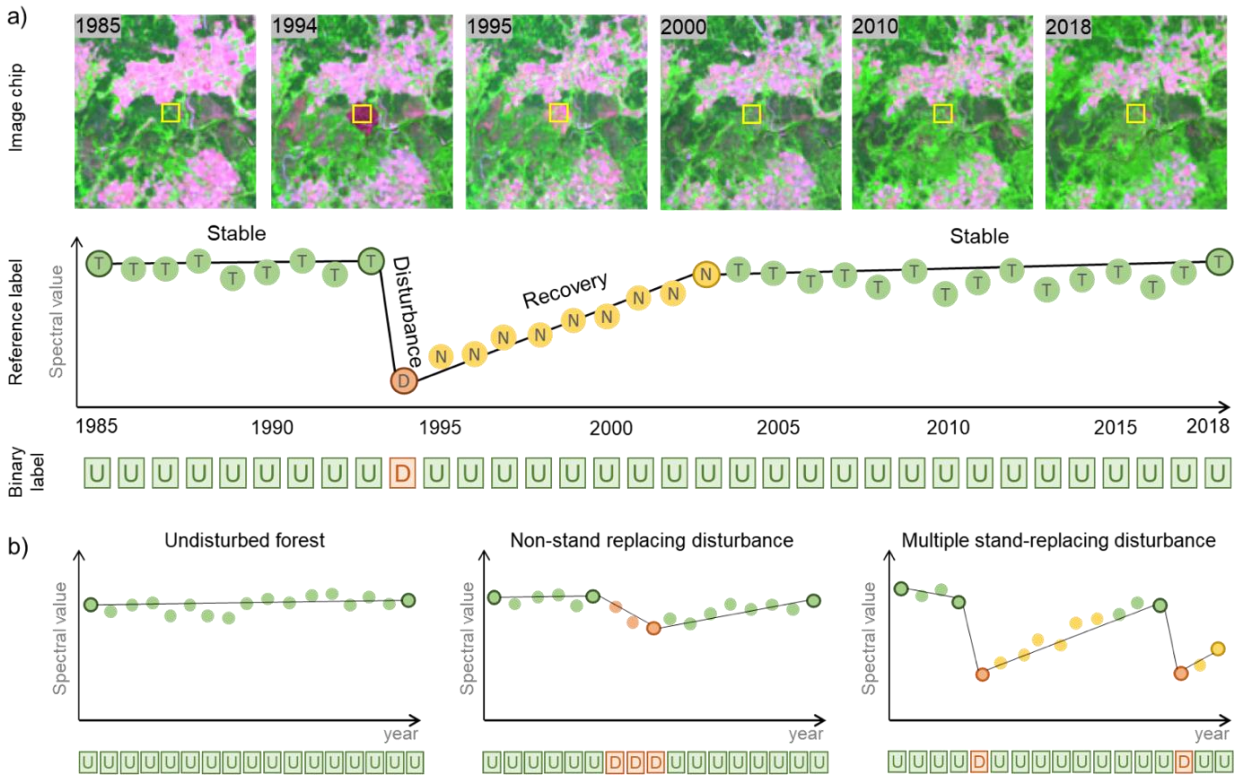

Figure 5. Schematic representation of the original reference data from Senf et al. 2018 and Senf et al. 2021: a) Image chips and corresponding segmentation labels of stable forest, disturbance and recovery segments and corresponding land cover of treed (T) and non-treed (N), ultimately labelled into annual binary classes of Undisturbed (U) and Disturbed (D); b) examples of undisturbed forest, a gradual non-stand replacing disturbance and multiple stand-replacing disturbances, and their corresponding binary labels.

Table 1. Sizes of the validation samples for the land use and disturbance classifications and the corresponding weights used for calculating the sample size (missing samples in the LUCAS database indicated by "-")

Country name	Country area (km^2)	Forest area (km^2)	Land area proportion	Forest area proportion	Land use validation sample		Disturbance validation sample
					forest	non-forest	forest
Albania	28786.07	7716	0.00501	0.268	3	-	9
Austria	83988.21	39600	0.01461	0.471	17	20	47
Belarus	207575.27	80334	0.03610	0.387	35	-	96
Belgium	30587.77	6834	0.00532	0.223	3	10	8
Bosnia and Herzegovina	51030.41	25599	0.00888	0.502	11	-	31
Bulgaria	110953.91	36250	0.01930	0.327	16	32	43
Croatia	57017.2	24901	0.00992	0.437	11	14	30
Czechia	78842.74	26000	0.01371	0.330	11	23	31
Denmark	43501.59	6120	0.00757	0.141	3	16	7
Estonia	45405.32	23066	0.00790	0.508	10	10	27
Finland	338434	233320	0.05886	0.689	101	46	278
France	549006.24	246640	0.09549	0.449	107	132	294
Germany	357454.99	114190	0.06217	0.319	50	105	136
Greece	124885.96	37600	0.02172	0.301	16	38	45
Hungary	93001.36	20990	0.01618	0.226	9	31	25
Ireland	70243.37	7540	0.01222	0.107	3	28	9
Italy	300887.48	106736	0.05233	0.355	46	85	127
Latvia	64549.9	28807	0.01123	0.446	12	16	34
Lithuania	64941.46	21223	0.01130	0.327	9	19	25
Moldova	33847.27	3290	0.00589	0.097	1	-	4
Montenegro	13764.39	6252	0.00239	0.454	3	-	7
Netherlands	35162.75	3650	0.00612	0.104	2	13	4
North Macedonia	25438.31	10285	0.00442	0.404	4	-	12
Norway	311654.47	121120	0.05421	0.389	53	-	145
Poland	311759.97	90000	0.05422	0.289	39	97	107
Portugal	88700.69	31820	0.01543	0.359	14	25	38
Romania	238289.83	69610	0.04145	0.292	30	74	83
Serbia	88372.28	27200	0.01537	0.308	12	-	32
Slovakia	49036.66	20006	0.00853	0.408	9	12	24
Slovenia	20221.19	12574	0.00352	0.622	6	3	15
Spain	498518.52	184180	0.08671	0.369	80	136	219
Sweden	450040.79	280730	0.07828	0.624	122	74	335
Switzerland	41239.21	12540	0.00717	0.304	5	-	15
Ukraine	597120.26	105000	0.10386	0.176	46	-	124
United Kingdom	245164.44	28650	0.04264	0.117	13	94	34

	912 1154 2066	2500

2.2.3 Forest land use mask

A forest land use mask was required to mask out any non-forest land use areas, which often exhibit strong spectral changes that could be confused with disturbances. Since there is no consistent forest land use mask available across all continental Europe, we created our own forest land use mask using a multitemporal classification approach. We broadly followed the FAO definition of forest land use, which is any area that has or will reach tree cover greater 10% in the near future, is larger than 0.5 ha and 20 m in width, and which is primarily not used under urban or agricultural land use (FAO, 2020). This definition thus includes young forests of lower tree cover, areas of reforestation and temporally unstocked areas (e.g. forestry roads). To match this definition, we assigned each reference pixel (see Section 2.2.) that has had tree cover (even temporally) since 1985 as forest land use. All reference pixels of other land uses (i.e. artificial, croplands, water bodies) were assigned to non-forest land use. We used such a broad land use definition, instead of a recent tree cover map, to avoid masking recently disturbed areas that are temporally unstocked but still considered forest land use. Using the reference categories assigned to each pixel, we trained a random forest model with a stack of all annual best available pixel composites as input (i.e. 34 years with 6 spectral layers and 6 spectral indices, resulting in 408 features, see Figure 6). We then applied the model to the raster stack to yield one classification of forest/non-forest land use for our entire study period. According to the FAO definition, we also defined a minimum mapping unit (MMU) consisting of 6 Landsat pixels for the forest mask (0.54 ha). All non-forest land use pixels were excluded from following analyses.

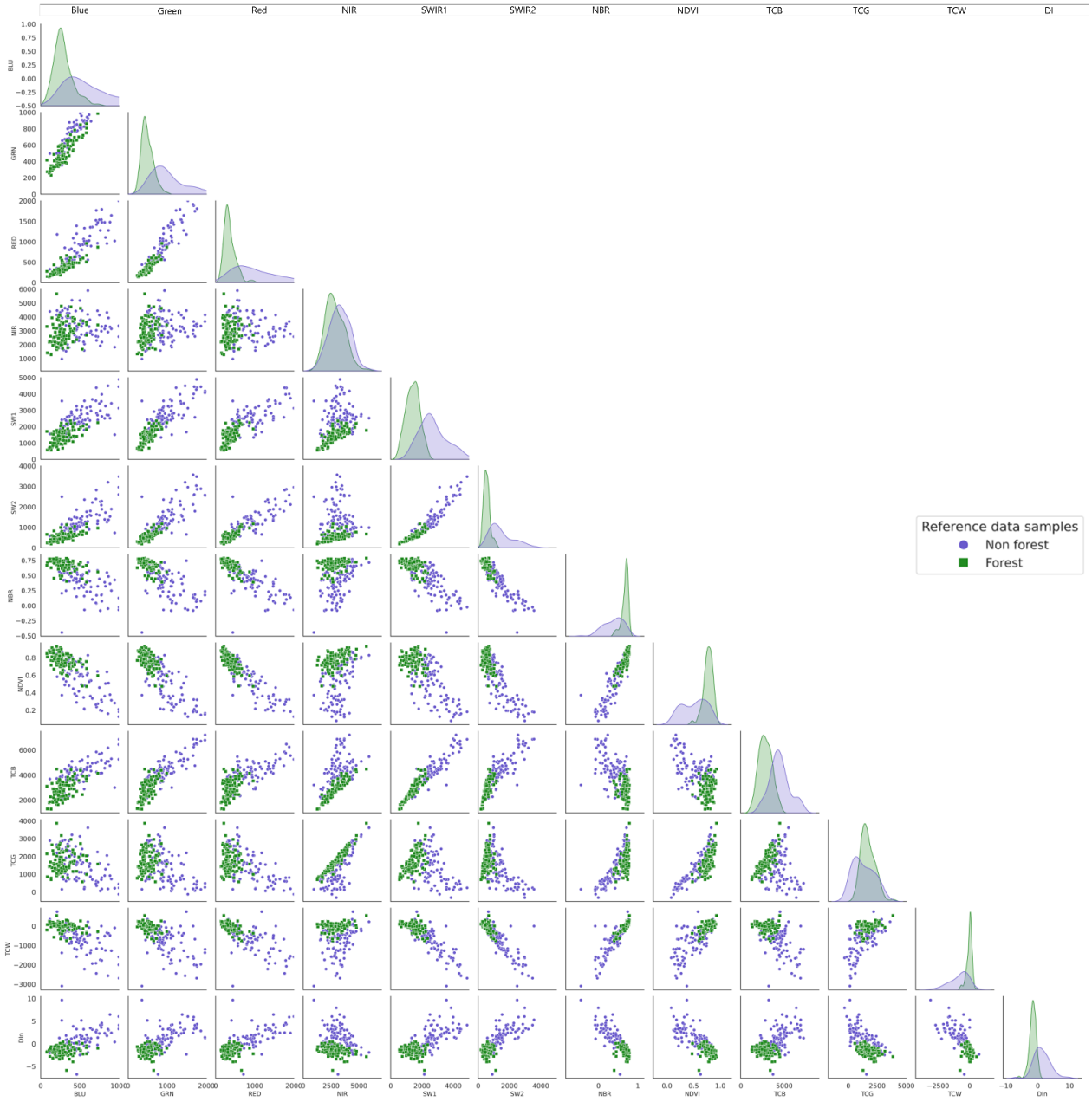

Figure 6. Feature space of spectral bands and indices for forest and non-forest reference samples.

2.2.4 Annual disturbance mapping

For each forest land use pixel, forest disturbances were classified on an annual basis since 1985. Disturbed and undisturbed forests have a distinctive spectral signal across a two-year period, as spectral characteristics change significantly during disturbance while they remain constant for undisturbed forests (Kennedy et al., 2007). We hence calibrated a random forest model to identify disturbed and undisturbed pixels using the spectral information from a given year (target year t_0) and the previous year (reference year, t_{-1}) to disentangle inter-annual stability and change. As classification input, we used a set of spectral indices (NDVI, Tucker, 1979 and NBR, García and

Caselles, 1991), the Tasseled Cap components (Crist, 1985), and the Disturbance Index (Healey et al., 2005) (Figure 7). As reference we used undisturbed and stand-replacing disturbances (i.e. disturbances that led to a change in land cover; see Section 2.2), which indicate a clear opening of the top canopy and thus allow the model to learn the distinct spectral differences between undisturbed and disturbed pixels. We use the full sample for validation though (see section below). The trained model was then applied annually, which facilitates updating the map product as soon as images are available for the next year.

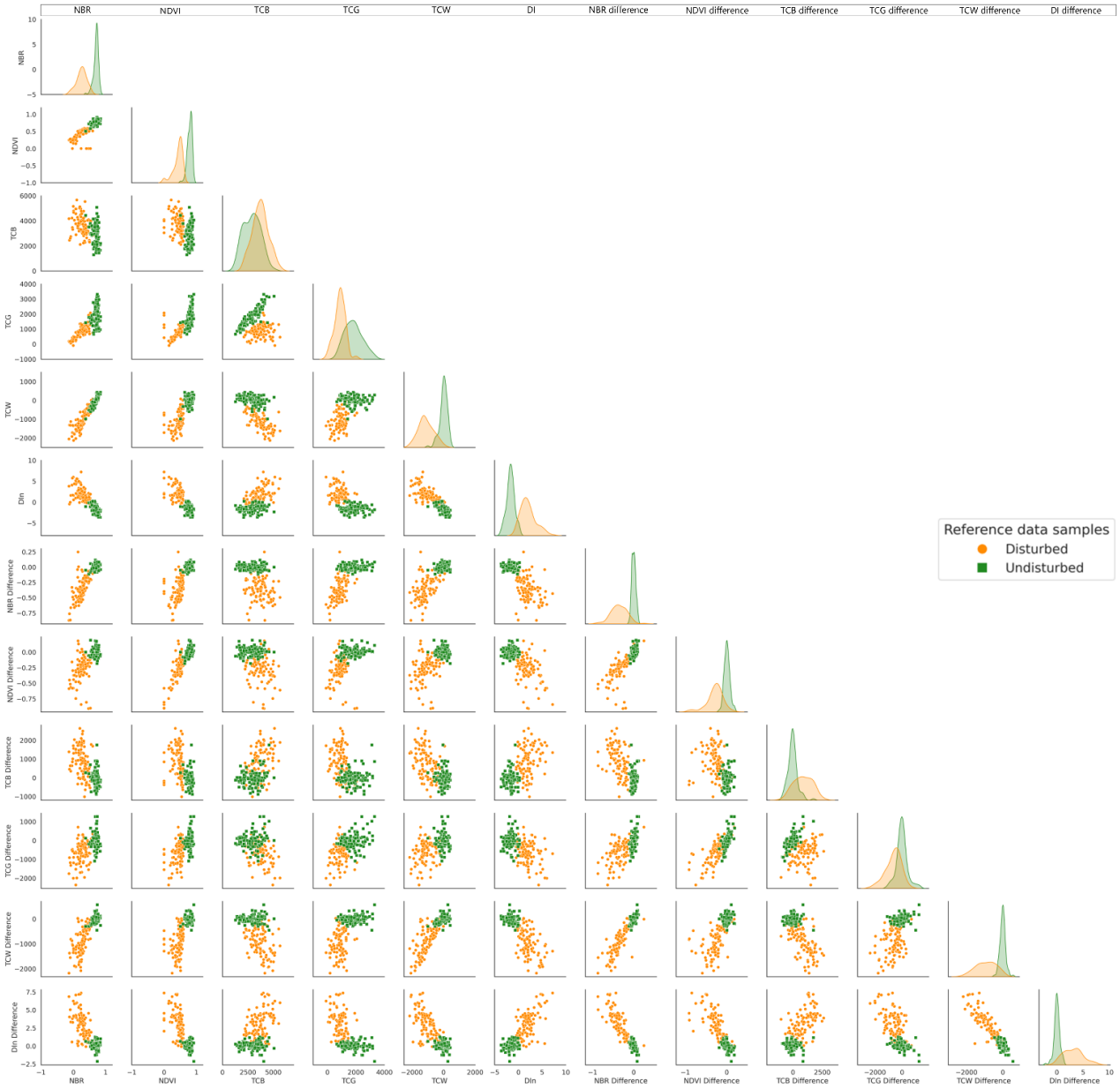

Figure 7. Feature space of spectral indices for disturbed and undisturbed reference samples. NBR = Normalized Burn Ratio, NDVI = Normalized Vegetation Index, TCB = Tasseled Cap Brightness, TCG = Tasseled Cap Greeness, TCW = Tasseled Cap Wetness, DIn = Disturbance Index.

Due to the imbalanced nature of the sample (disturbances representing only 3%), we applied the SMOTE method (Chawla et al., 2002), which helps by lowering the overall learning cost, assigning higher costs to misclassifying the minority class, and balancing the data through under-sampling the majority and over-sampling the minority class. We further used the module RandomizedSearchCV (Pedregosa et al., 2011) to efficiently find the best set of hyperparameters using cross-validation, selecting the combination that resulted in the highest area under the receiver operating characteristic curve score. To classify the probability of disturbance into the binary categories disturbed and undisturbed, we optimized the probability cutoff using the F1-score derived from cross-validation. The optimal threshold was 0.5 with an overall F1-score of 91%.

Finally, we applied several post-processing steps: First, annual maps were masked according to the forest land use classification and a minimum mapping unit of 3 Landsat pixels was assigned to disturbance pixels (0.27 ha), reducing noise from single pixels falsely flagged as disturbance (so-called *salt-and-pepper* effect typical for pixel-based classification). The smallest disturbance in the disturbance map thus is 0.27 ha. Second, we applied a collapsing step in which multiple disturbances detected in a time window of <4 years since the last disturbance were reduced to the first year the disturbance sequence was detected (Figure 8). This was done to avoid illogical changes (i.e. disturbance, no disturbance, disturbance) and double counting of disturbances in consecutive years that arise from the same disturbance event. That is, if a disturbance event extends over multiple, consecutive years, we will map only the first year of the disturbance segment as disturbed.

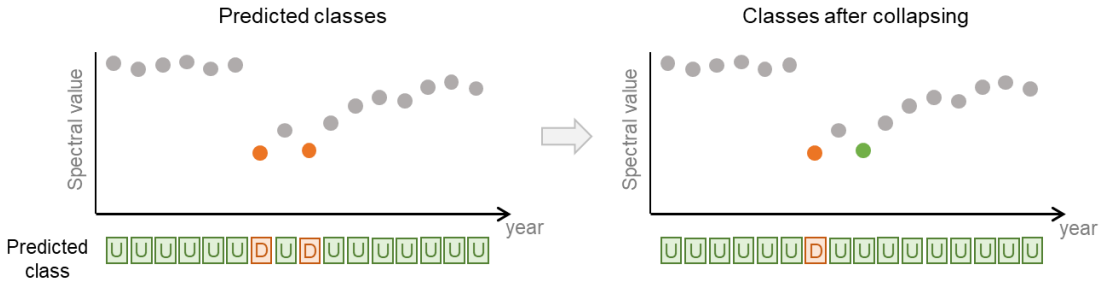

Figure 8. Schematic view of collapsing step to relabel illogical and consecutive disturbance events.

2.2.5 Accuracy and uncertainty assessment

Robust estimation of map accuracies is important for appropriate use and interpretation of maps and for subsequent analyses (Olofsson et al., 2014). We quantified map accuracies of both the forest land use and forest disturbance maps using the independent samples described in 2.2 and standard confusion matrices, from which we estimated commission errors/producer's accuracies (or recall), omission errors/user's accuracies (or precision) and over errors/accuracies and F1-scores (Forman & Scholz, 2010). Commission errors estimate how many pixels were falsely classified as disturbance/no disturbance, whereas omission errors estimate how many disturbance/no-disturbance pixels were missed by the classified. Uncertainties in each accuracy estimate were quantified using bootstrapping with 1000 repetitions, which simulates sampling variability due to hypothetical repeated sampling for each estimate. We further assessed how accuracies of the disturbance map varied spatially and temporally. For assessing spatial variability, we computed confusion matrices per region (North - 7 countries, Central - 20 countries, and South - 11 countries), according to FOREST EUROPE 2020 definitions). For assessing

temporal variability, we computed confusion matrices and corresponding accuracy estimates per year and for different periods. We also calculated confusion matrices using only stand-replacing disturbances (i.e. pixels where the interpreter assigned a change in land cover) and both stand-replacing and non-stand-replacing disturbances (i.e. pixels where the interpreter assigned no change in land cover). We did so to test the sensitivity of our maps and map accuracies to low-severity disturbances. We finally compared the forest land use area estimated from our map to national level statistics on forest land use area obtained from the FAOSTATS database (FOREST EUROPE, 2020).

2.2.6 Agent attribution

To identify the agent of disturbances, we adapted an existing attribution algorithm (Sebald et al., 2021; Senf & Seidl, 2021b) for the next-generation disturbance map. The algorithm first detects disturbance patches by grouping pixels disturbed in the same year that are connected by an edge or corner. That is, the analysis is performed at the patch-level and not at the pixel level. To correct for timing errors in the disturbance map (e.g., a fire spreading over two years might appear as two separate fires), patches from consecutive years that share an edge are merged, with the disturbance year assigned based on a majority vote. For each patch, we generated a set of predictors, including its size, shape, spectral characteristics, and surrounding landscape (see Table 2 for details).

Table 2: Predictors used in the agent attribution model.

Predictor group	Predictor	Description
Spectral characteristics	Spectral value (NDVI, NBR, TCW, TCB and TCG)	The absolute value in the respective spectral index in the year of the disturbance event.
	Spectral magnitude (NDVI, NBR, TCW, TCB and TCG)	The absolute change in the respective spectral index during the disturbance event.
Patch characteristics	Patch size	The total size of the disturbance patch in ha.
	Patch fractional dimension index	The fractional dimension index, given an indication of patch complexity, with larger values indicating more complex patches.
Landscape	Pulse-dynamics	The proportion of disturbances in a 5 km radial buffer occurring in the same year as the focal patch. The values range between zero and one, with zero indicating no other disturbances in the same year in the 5 km neighbourhood (high spatiotemporal segregation), whereas a value of one indicates that all other disturbances in the surrounding neighbourhood occurred in the same year (high spatiotemporal clustering).
	Number of patches	Overall number of patches in the 5 km radial buffer.

As reference data, we used an extended reference database of 12,571 points for fires, windthrows, and bark beetle disturbances. This database builds upon an existing reference database of 11,364 points of fires and windthrows developed in (Senf & Seidl, 2021b), and which was extended to 12,571 points in the RESONATE project to also including bark beetle disturbances (Seidl & Senf, 2024). Each point in the reference database was linked to a disturbance patch by a unique patch id. In many cases, two or more occurrence points fell within

the same disturbance patch (especially for large fires or windthrows), which reduced the number of patches with an agent label to 9,256 (6,986 for windthrow, 727 for bark beetles and 1,543 for wildfire). The database thus covers the three most important natural disturbance agents in Europe (Patacca et al. 2022). We note, however, that the focus of the attribution is on the root cause of disturbance. That is, an area affected by natural disturbances (e.g. wind) that is salvage logged afterwards is considered a wind disturbance. We did not have dedicated information on harvest in the reference database and interpreting harvest is difficult, because harvest can happen in reaction to natural disturbances (i.e. salvage logging). We thus used an approach developed in Senf & Seidl, 2021b and selected a random background sample from all disturbance patches. As harvest is assumed to be the major disturbance agent in Europe (Patacca et al. 2022), this background sample will represent harvest conditions in contrast to the agent information available in the existing databases. We drew a total of 9,256 harvest patches, equalling to the sum of wind, fire and bark beetle disturbances, and resulting in a final reference database of 18,512 patches.

We trained a random forest model using the agent labels in the reference database and predictors described above. We initially trained a model classifying each patch into one of the four agent categories (wind, fire, bark beetle and harvest) but found high confusion between wind and bark beetle disturbances. This was mostly due to bark beetle reference data being available only for very recent, large-scale bark beetle outbreaks in Central Europe that resemble past windthrows in their form and shape. As little to no historic information was available on bark beetle disturbances that could have improved model skill, we decided to group wind and bark beetle disturbances into one category. Ecologically, both disturbance agents form a disturbance complex with wind disturbances often triggering bark beetle disturbances and vice versa (Seidl and Rammer 2017). That said, as bark beetle has been a minor disturbance agent prior to the recent Central European drought event (Patacca et al. 2022), disturbances in the wind/bark beetle category prior to 2017 can be considered as mostly wind disturbances. As there was no independent sample of reference data available at the agent level, we were unable to provide map accuracies on the resulting agent maps. We refrained from providing model accuracies as those can be misleading as model accuracies are typically much higher than true map accuracies.

2.2.7 Summary layers

Using the annual maps of disturbances and the underlying spectral information, we computed a series of layers that compose the next-generation forest disturbance map (see Table 3 for a detailed summary). The resolution of all layers is 30-m, containing information from 1985 to 2023 (in the current version 2.11). The data is distributed at the country level using the ETRS89 coordinate reference system and the Lambert Azimuthal Equal-Area projection (ETRS89-extended / LAEA Europe; EPSG:3035) and can be downloaded from https://doi.org/10.5281/zenodo.13333034 and viewed under following link https://albaviana.users.earthengine.app/view/european-forest-disturbance-map.

Table 3. Summary layers of the next-generation forest disturbance map

Dataset	Name	Valid range	Description
Forest land use mask	forest_mask_{country}	1	Forest land use mask

Annual disturbance	annual_disturbances_1985_2023_{country}	(0, 1)	Stack of annual disturbances indicating undisturbed (0) and disturbed (1)
Disturbance probability	disturbance_probability_1985_2023_{country}	(0, 100)	Stack of annual disturbance probabilities within forest areas
Latest disturbance	latest_disturbance_{country}	(1985, 2023)	Indicates the year of the most recent disturbance
Greatest disturbance	greatest_disturbance_{country}	(1985, 2023)	Indicates the year of the highest disturbance severity (greatest spectral change)
Number of disturbances	number_disturbances_{country}	Categorical (1, 2, 3, 4)	Indicates the number of disturbance events within the time-series (1, 2, 3, or more than 3 events, i.e. 4)
Disturbance severity	disturbance_severity_1985_2023_{country}	(-10,000, 10,000)	Stack of spectral change in NBR for disturbance patches ($t - t_{-1}$).
Disturbance agent	disturbance_agent_1985_2023_{country}	Categorical (1, 2, 3)	Annual classification of disturbance agents (1- wind/bark beetle, 2- fire, 3- harvest) implemented in Seidl and Senf, (2024) following Sebald et al. (2021) and Senf and Seidl (2021b)
Aggregated disturbance agent	disturbance_agent_aggregated_{country}	Categorical (1, 2, 3, 4)	Aggregated layer of agents that summarises the dominant agent within the time-series. In those cases where a pixel has been disturbed more than once and by multiple agents, the class mixed (4) is assigned.

2.3 Results

2.3.1 Map products in the next-generation European forest disturbance atlas

Our newly created Forest Disturbance Atlas for Europe provides a set of variables at 30-m describing forest disturbances annually since 1985 until 2023 (Figure 9). By mapping disturbances annually, we capture multiple disturbance events to obtain direct information on the number of disturbances, the latest disturbance (most recent disturbance event), and the greatest disturbance (event with the greatest spectral change). The annual disturbance maps are coupled with the estimated annual disturbance probabilities from Random Forest as a proxy of uncertainty on disturbance detection. Maps are constrained to our forest land use mask, which includes areas that have been forest at some point within the time series and thus avoids masking recently disturbed areas.

Figure 9 shows a Europe-wide overview of the area affected by disturbances over the last four decades. The individual layers show the ability of our atlas to precisely retrieve disturbance characteristics across different disturbance regimes, such as a recent bark beetle outbreak interspersed with harvest activities in the Thuringian Forest in Germany (Figure 9.1), interactions between different agents in an intensively managed forest plantation in the Gascony region in France (Figure 9.2), where a windstorm was captured in 1990 and a recent fire occurred in 2023, as well as recurrent fires in planted forests in central Portugal (Figure 9.3).

Based on the annual disturbance estimates (Figure 10), we quantified a total forest disturbed area of 439,000 km^2 (22% of the total forest area) that increases to 610,000 km^2 when accounting for

multiple disturbance events (i.e. considering area disturbed in multiple years, such as reburns or multiple harvests). From those disturbed areas, 72% correspond to areas with one disturbance event while 18.9% (115,000 km^2) have two disturbance events, and 9.1% (55,000 km^2) have three or more disturbance events. That is, a total of 28% of all pixels disturbed in Europe in the past four decades had multiple disturbance events. These multiple disturbances mostly occur in southern Europe (e.g. Portugal, Spain), where multiple fires occurred interspersed with harvesting activities, as well as in regions characterized by short-rotation plantation systems, e.g. the south of France or Hungary, with harvests return periods of 15-20 years.

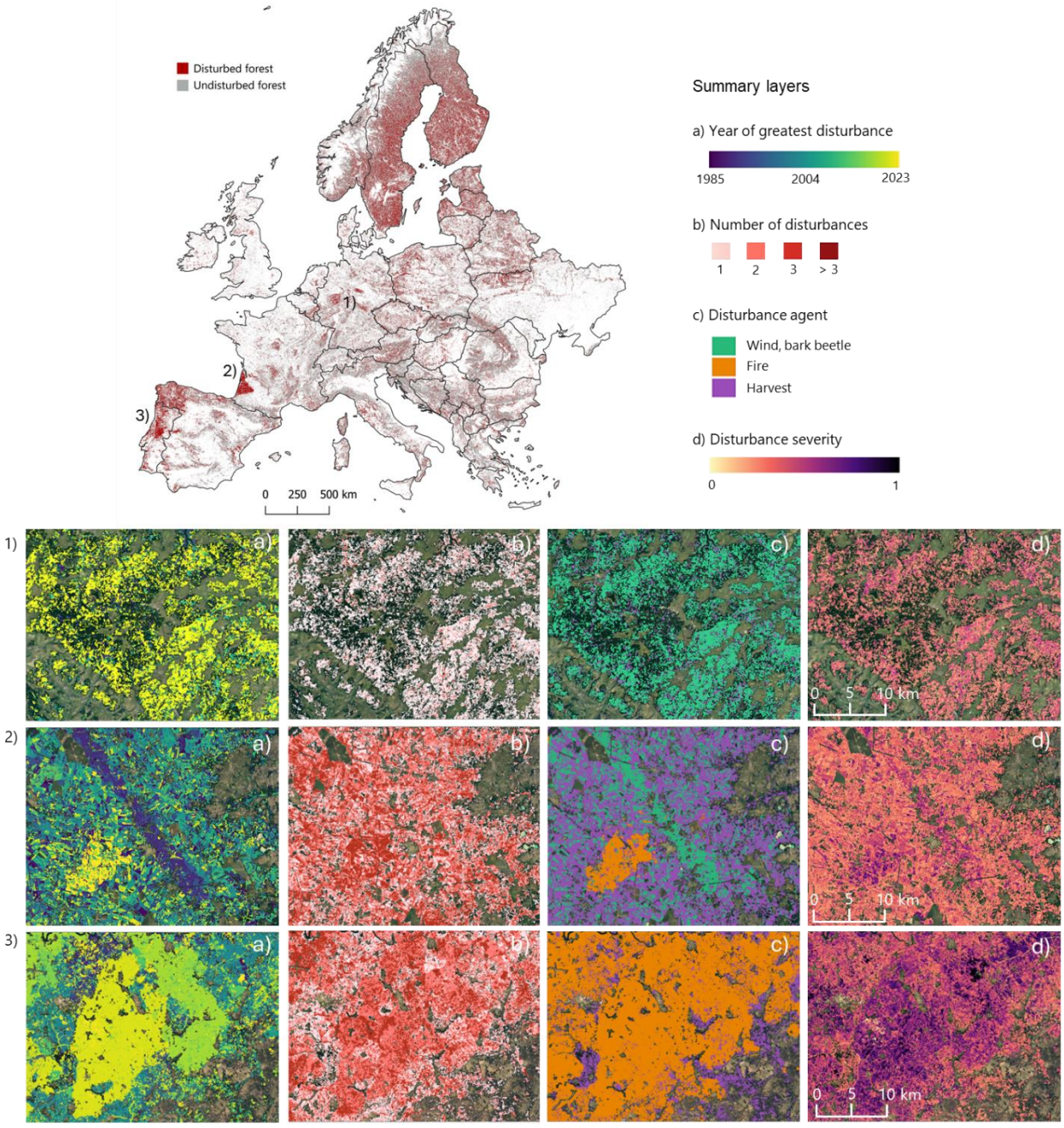

Figure 9. Forest disturbances in Europe (1985-2023). Details show (a) bark beetle outbreaks in central Germany; (b) a combination of a windstorm event and a recent fire in a managed forest plantation in Gascony (France); and (c) recurrent fire disturbances in central Portugal.

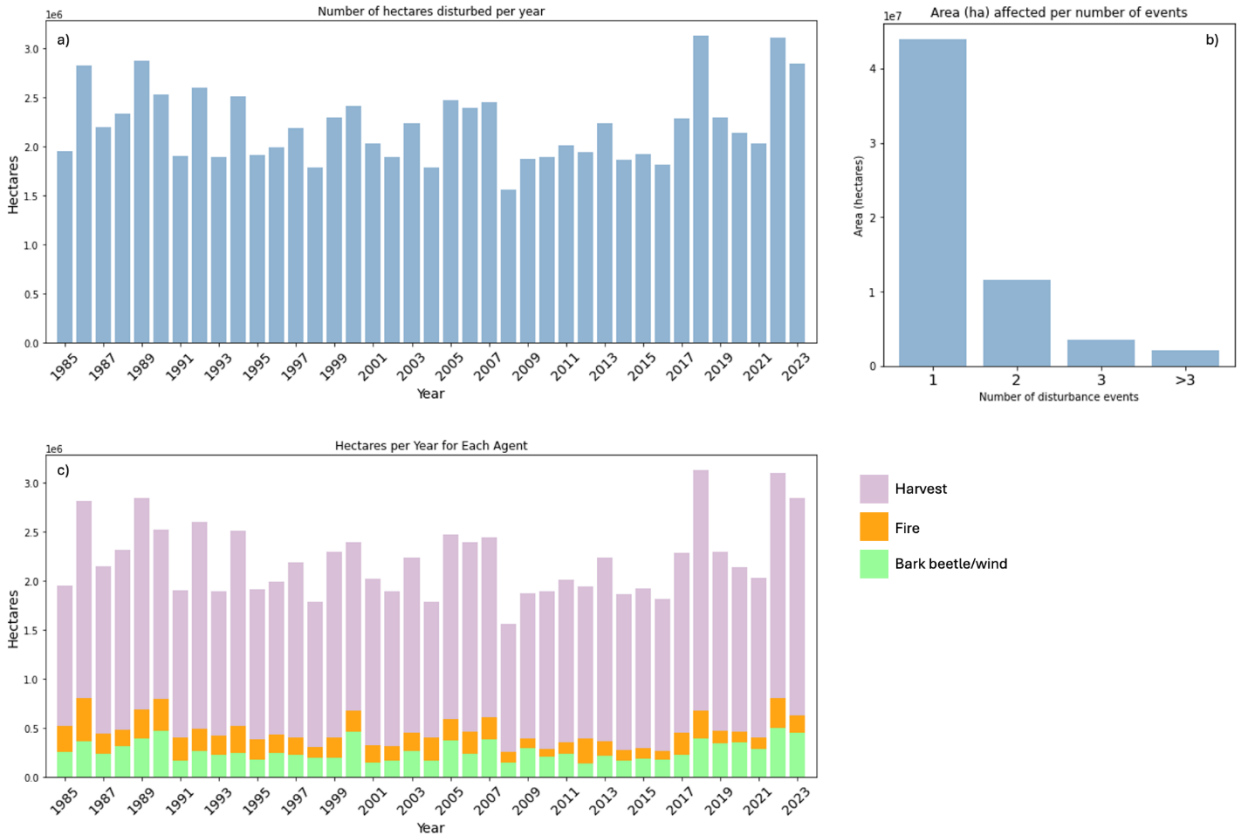

Figure 10. Area disturbed per year (a,) area disturbed per number of events (b), and area disturbed by agent (c).

Breaking down total disturbance area by agent we found harvest to be the most important disturbance agent overall, causing approximately 79 % of all disturbances since 1985. The second most important agent was wind and bark beetle, causing approximately 12 % of all disturbances, followed by fire with 9% of all disturbances. From a temporal perspective (Figure 10c), one can identify years of large-scale windthrow (e.g. 1990 [storm Vivian/Wiebke], 2000 [storm Klaus happening late in 1999] or 2007 [storm Kyrill]), but also the more recent years (i.e. 2018, 2022, 2023) characterized by amplified bark beetle activity across Central Europe. That said, those waves of natural disturbances remain small in comparison to overall harvest volume.

2.3.2 Accuracies and uncertainties

The assessment of the land use mask, discriminating between non-forest and forest land use, yielded high accuracies with an overall F1-score of 0.92, with F1-scores of 0.93 for non-forest and 0.91 for forest land use, respectively (Table 4). Commission and omission errors were thus overall low (<10%) and well balanced, indicating that on average 7.5% of the pixels in the land use mask will be wrongly classified. Further, comparison of the estimated forest area at national level with forest land use area from FAOSTATS showed a high agreement (R^2 = 0.98; MAE = 6,019km^2, 16.9%; Figure 11), with a slight tendency of overestimating forest area.

Table 4. Confusion matrix of the forest land use map

	Reference data

	Non-forest	Forest	N	Commission errors (%)	Confidence intervals
Non-forest	1025	61	1086	**5.6**	[90.1 - 99.9]
Forest	91	851	942	**9.7**	[87.1 - 93.9]
N	1116	912			
Omission errors (%)	**8.2**	**6.7**			
Confidence intervals	[87.9 - 94.8]	[89.9 - 98.2]			
F1 score	**0.93**	**0.91**	Overall accuracy = 92.5 % | Overall error = 7.5 %		

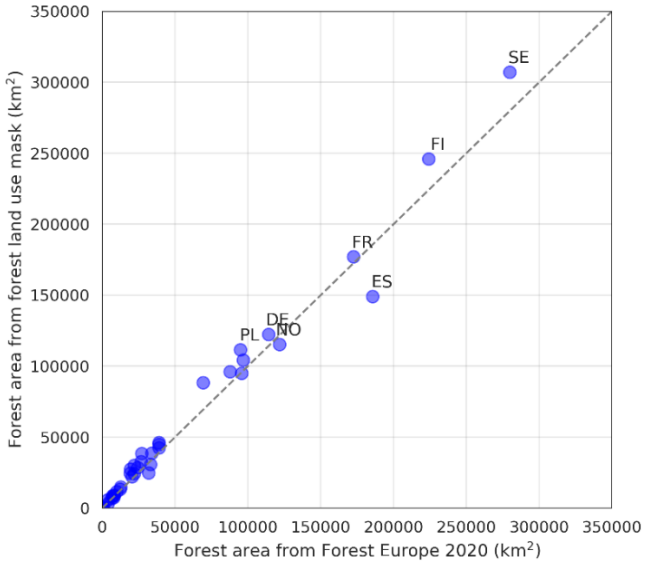

Figure 11. Forest area according to FAOSTATS 2020 versus forest land use estimated for the series 1985-2023.

Map accuracies of the disturbance classification showed an overall F1-score of 0.89, with F1-scores of 0.99 for undisturbed and 0.80 for disturbed pixels (Table 5). Accuracies were thus less balanced between both classes, with very low commission/omission errors (<1%) for the undisturbed class (i.e. the map rarely misses or falsely classifies something as undisturbed), but higher commission and omission errors for the disturbed class (17.3% and 22.5%, respectively). That is, for 17% of all pixels classified as disturbance the classification will be wrong, whereas 23% of true disturbances will be missed in our map. Those numbers changed significantly when only considering reference points interpreted as stand-replacing disturbances in the reference data (Table 6), with a substantially reduced omission error (14.2% vs. 22.5%) and a slightly increased commission error (22.7% vs. 17.3%). The omission error of our map is thus driven by low-intensity disturbances that remove the tree canopy only partially. We found significant variation in accuracies among regions, with higher commission error in Northern Europe compared to Central and southern Europe, whereas omission was lower in Northern Europe compared to Central and Southern Europe (Table 7). In Northern Europe, it is thus more likely to randomly select a pixel falsely detected as disturbance but less likely that a true disturbance is missed, while in Central and Southern Europe it is more likely to miss a true disturbance than falsely identifying a pixel as disturbed.

In the temporal domain we found overall high fluctuation between years, yet an overall decreasing trend in commission error but no clear trend in omission error (Figure 12). That is, disturbances mapped before the year 2000 will have an 19.5% chance of being falsely classified as disturbance, whereas commission substantially decreases to 10.6% after the year 2000. The high commission error prior to 2000 was driven significantly by the very early years of our time series (<1990), where commission error was on average 22.5% (compared to 16.4% between 1990 and 1999). Omission errors showed a less clear pattern, with also high omission rates in the 80s (23.3%) but less pronounced differences before/after the year 2000 (22.5% and 20.5%, respectively). Finally, we found high agreement between estimated and manually interpreted disturbance years (R^2 =0.73; Figure 13), with a mean absolute error of 1.91 years.

Table 5. Confusion matrix of disturbance detection assessment

	Reference data				
	Undisturbed	Disturbed	N	Commission errors (%)	Confidence intervals
Undisturbed	83206	220	83426	**0.26**	[99.70 - 99.89]
Disturbed	158	756	914	**17.29**	[77.30 - 87.78]
N	83364	976			
Omission errors (%)	**0.18**	**22.54**			
Confidence intervals	[99.65 - 99.83]	[71.98 - 82.27]			
F1 score	**0.99**	**0.80**	Overall accuracy = 89.9 % | Overall error = 10.1 %		

Table 6. Confusion matrix of disturbance detection assessment leaving out non-stand-replacing disturbances

	Reference data				
	Undisturbed	Disturbed	N	Commission errors (%)	Confidence intervals
Undisturbed	83233	74	83307	**0.1 %**	[99.83 - 99.97]
Disturbed	131	447	578	**22.66 %**	[74.30 - 82.87]
N	83364	521			
Omission errors (%)	**0.2**	**14.20**			
Confidence intervals	[99.72 - 99.85]	[79.98 - 88.27]			
F1 score	**0.99**	**0.81**	Overall accuracy = 90.7 % | Overall error = 9.28 %		

Table 7. Errors in disturbance detection per region

	North		**Central**		**South**	
	Commission errors (%)	Omission errors (%)	Commission errors (%)	Omission errors (%)	Commission errors (%)	Omission errors (%)
Undisturbed	0.14	0.2	0.2	0.1	0.13	0.1
Disturbed	23.32	18.1	13.62	26.96	16.99	22.97

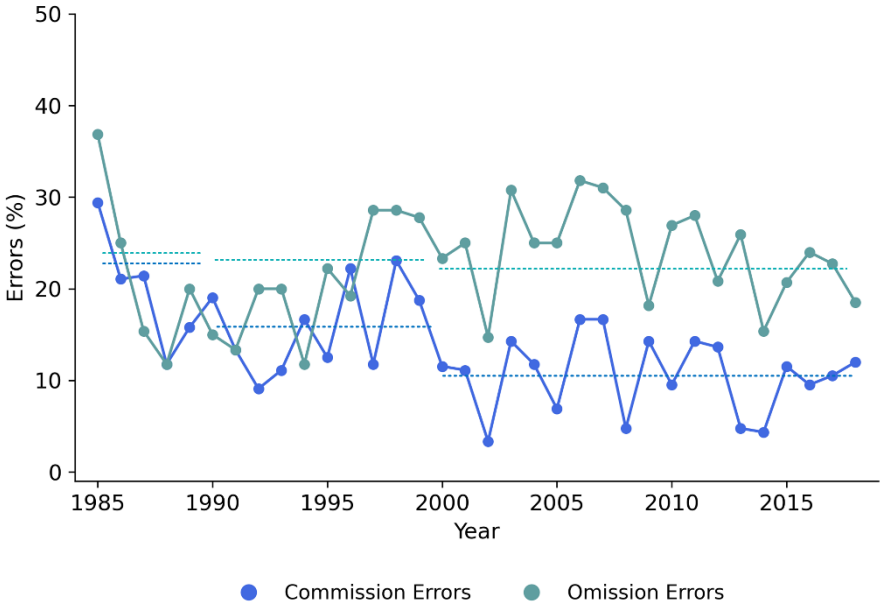

Figure 12. Omission and commission errors per year. Dashed horizontal lines indicate the averages per period 1985-1989, 1990-1999, 2000-2018.

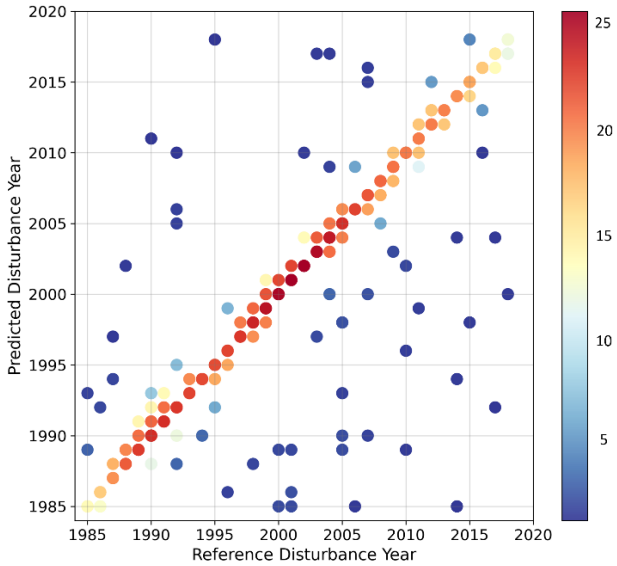

Figure 13. Validation of the disturbance year. Estimated disturbance year versus manually interpreted year of disturbance for independent reference pixels (colour bar indicates the number of points)

2.4 Discussion

We presented the next-generation European forest disturbance map, providing spatially explicit information on disturbances since 1985 and demonstrating the importance of long time series for understanding disturbance change, disturbance regimes and disturbance interactions. While

there are already several disturbance products for Europe (Senf & Seidl, 2021a; Turubanova et al., 2023), our new dataset moves beyond existing products in at least four important ways: First, annual disturbance mapping enables to capture multiple disturbance events, providing direct information on disturbance frequencies and interactions. Multiple disturbance information is often not assessed or simplified to the greatest or latest disturbance of the analysis period. Without capturing disturbances on an annual basis, however, we lose valuable information on forest dynamics (Buma, 2015; Hermosilla et al., 2019), such as disturbance return times (Pugh et al., 2019). Including multiple disturbances was thus an important improvement over past map products available for Europe. Second, the long-time frame of our map products offers a baseline for understanding more recently observed changes and for quantifying trends over time. Furthermore, our workflow is designed to be operationally updated when new data arrives and therefore enables to create up-to-date maps annually after the summer season. Third, our approach is consistently applicable over all European forests with variable forest types and forest disturbances. It thus is likely to be adoptable and replicable in other regions. Fourth, we provide a characterization of disturbance agents and thus the root cause of disturbance. To facilitate user uptake and reproducibility, we provide a full open access framework that can be implemented on any computer system, independent of commercial cloud environments, data providers or software tools (see here for all codes: https://github.com/albaviana/European-Forest-Disturbance-Atlas). The next-generation forest disturbance map thus contributes to a future operational forest monitoring envisioned for Europe.

One of the aims of this paper was to provide a full characterization of map uncertainties. Compared to the first pan-European disturbance product developed by Senf and Seidl (2021a), which reported a commission error of 17.1% and an omission error of 36.9% for detecting disturbances, we reduced omission errors considerably (22.5%) while maintaining low commission errors (17%). That is, the next-generation European forest disturbance map detects more true disturbances compared to the past state-of-the-art product, while making similar errors in falsely detecting disturbances. Disturbance areas derived from Senf and Seidl (2021a) are thus likely conservative estimates that underestimate true disturbance area. A similar improvement was found in the timing of disturbances, with a mean absolute error of 3 years for Senf and Seidl (2021a) compared to 1.9 years for our map, indicating that fewer disturbances are attributed to a wrong year compared to manually interpreted data. The disturbance detection accuracies found here are further consistent with more local approaches implemented in Europe, such as Francini et al. (2021), who reported omission/commission errors of 27% and 30% for clear-cut mapping in Italy with the 3I3D algorithm; or recently developed disturbance maps for the European Alps (Morresi et al., 2024), with omission/commission errors of 16.9% and 16.5%, and for continental Spain (Miguel et al., 2024), with omission/commission errors of 12.4% and 18%, respectively. Our map was of slightly lower accuracy than a very recent pan-European forest disturbance map developed by Turubanova et al. (2023), with omission and commission errors of ~19% and ~7%, but their map only covers disturbances since 2001. Comparing our map accuracies to Turubanova et al. (2023) for the same period yields approximately similar accuracies between both map products (see Figure 12). Finally, our forest land use mask yielded high accuracies, and the forest area estimated per country aligned well with estimates provided by FOREST EUROPE (2020). Remaining differences can be explained by the different temporal reporting periods, as FOREST EUROPE reports the forest area in 2020 but we provide a multitemporal land use definition, comprising all areas that have been forest land use at some point since 1985. Our forest land use map is also distinctly different from recent tree cover products (Liu et al., 2023; Turubanova et al., 2023), as it includes recently disturbed areas that might be temporally unstocked but return to

closed canopy forests in the near future (Mandl et al., 2024) and thus meet our forest land use definition. Using static tree cover maps would mask out those areas based on the (arbitrary) year they were created for. We thus move away from a static forest cover map to a forest land use map that reflects the spatial and temporal dynamics of forests (Chazdon et al., 2016).

Despite the good accuracies compared to past and other map products, there are several limitations and possible sources of uncertainties that should be considered when using the maps. The higher omission error for the disturbed class and wider confidence intervals indicates a possible limitation of our approach when mapping low severity disturbances that do not produce a clear or complete opening of the canopy (e.g. thinning). The effect of such non-stand replacing disturbances is clearly seen when redoing the accuracy assessment without them, which nearly halved the omission error (Table 5). Non-stand replacing disturbances are quite widespread in central and southern Europe, leading to higher omission errors in those regions compared to northern Europe characterized by a high prevalence of clearcut harvesting compared to Central and Southern Europe. In northern Europe, however, a higher frequency of missing data and noise (i.e. remaining clouds, difficult illumination conditions) led to higher commission errors in comparison to central and southern Europe, where data availability was overall higher (see Figure 2b). We further found that omission errors were caused by a delayed detection of disturbances (i.e. disturbances detected one year after the reference). The detection of disturbances with a delay of one year is a common problem when working with annual summer composites, where disturbances can occur in the same year but after the compositing date (i.e. a pixel is selected from August, but the disturbance occurs in November and will thus only be detected the next summer). Finally, the use of a minimum mapping unit may contribute to the omission of small-scale disturbances < 0.27 ha but reduces false positives from individual isolated pixels (Reinosch et al., 2024). We thus argue that the loss of information imposed by the minimum mapping unit outweighs the potential error by included many small-scale disturbances with high errors. The application of a minimum mapping unit is also a common practice to facilitate the integration with different data sources such as national forest inventories (Wulder et al., 2024) or national land use products (Gómez, et al., 2019).

The temporal validation revealed variability in accuracy over the time series, with the 1980's and 1990s having considerably lower accuracies (especially in terms of commission error) than disturbances detected after 2000s (Figure 12). This result highlights challenges with reliability estimating disturbance trends from remote sensing data (see Olofsson et al., 2014 for an indepths discussion), as trends derived from disturbance maps with variable accuracy will be biased (Palahí et al., 2021). Quantification of disturbances and trends should thus be based on a manually interpreted sample or based on models taking in to account the variable accuracies over time (Francini et al., 2022; Olofsson et al., 2014). Finally, despite careful processing and checking, there will be remaining errors in the underlying remote sensing data and specific regions and years will have large errors resulting from noise. For example, some peaks of higher omission errors displayed in Figure 12 coincide with years of higher percentage of no-data pixels as for example in the 80s, 1998, 2008 or 2012 (Figure 4). This limitation is intrinsic to the data used for producing the next generation forest disturbance map, but so far there is no other consistent alternative to provide spatially explicit and consistent information for monitoring Europe's forest over time.

We provide an agent attribution layer based on a previous established model (Senf and Seidl 2021b; Seidl & Senf, 2024), complementing the disturbance layers of the next-generation European forest disturbance map described in this study. Although comparison of estimates with

existing databases for fire and storms in Senf and Seidl (2021b) and salvage logging in Seidl and Senf (2024) showed high agreement of the attribution model with external datasets, an independent validation sample on the actual occurrence of disturbance agents is still missing for robust reporting of uncertainties. Validation of agent-level disturbance maps remains still an ongoing challenge, as no independent reference dataset spanning the whole period of our map is unavailable. Creating such a data set through manual interpretation, as was done for the overall disturbance reference dataset, is also challenging, as differentiating between different disturbance agents is impossible without additional data such as high-resolution imagery that is often only available for more recent years. Further, the fact that management superimposes most natural disturbances though salvage logging makes it difficult to disentangle the root cause of disturbance. Hence, change attribution needs to be further refined, especially when considering the addition of further agents (Stahl et al., 2023), such as drought and more gradual non-stand replacing disturbances that also have a markedly impact on forest condition globally (Coops et al., 2020; Hammond et al., 2022). The agent attribution provided should thus be taken with caution, as reliable estimates of map accuracies are absent. Novel data collected across Europe might help with overcoming this data challenge (Forzieri et al., 2020, 2023; Patacca et al., 2023), but most data does not cover the 80s and 90s, which are especially important for improving map accuracy and quality. Further research should thus increasingly focus on generating reference data that can be used for remote sensing applications, and we urge authors to make their reference data openly available, ultimately leading to a better understanding of forest change across Europe.

There is high need for an operational forests monitoring system at European scale and our newly developed next-generation European forest disturbance map constitutes a first step in this direction by producing disturbance information in a standardised way, with consistent quality across all forests in Europe. Our approach further relies on open-source data and tools and is implemented in an open-access framework, which allows full reproducibility and easy updating into the future and thus paves the road for operationalisation. Our data covers already 40 years of forest disturbance dynamics, constituting the longest remote sensing-based disturbance products available at continental scale today. Covering such long-time series allows for a better characterisation of disturbance regimes, requiring several decades of data to reliably quantify underlying distributions (Maroschek et al., 2024), as well as they can help to improve disturbance modelling by providing empirical data on disturbance occurrence (e.g. Grünig et al., 2022). Hence, the next generation European forest disturbance map will improve our understanding of European forest disturbance dynamics beyond simple monitoring forest disturbances.

2.5 The computational costs of disturbance mapping

The next generation disturbance map is based on approximately 50 TB of raw data. Downloading the raw data takes approximately two weeks using a standard internet connection. Updating the data by year is much faster and can be done within a day. The processing to Level-2 and Level-3 is done on a Linux-cluster with 48 multi-threading cores (resulting in a maximum of 96 parallel processes). Processing time of individual images are low (1-3 minutes), but the processing time is largely driven by the connection of the storage system to the cluster. We use a high-performance storage system that has a very fast connection to the cluster, which effectively reduces the processing time of an individual image to <1 minute. Storing the raw as well as Level-2 and Level-2 data requires approximately 150 TB of data. The operationalization of the process

would thus require, at minimum, a Linux cluster with 48 CPUs or larger and a high-performance storage space of 150 TB.

3 Forest structure maps

3.1 Introduction

Spatially explicit information on forest structure is needed for forest monitoring, parameterization of forest simulation and growth and yield models, and to support carbon accounting. Traditionally, forest structure has been monitored using field inventories. Although these data produce valuable estimates on forest resources at regional and national scales, the data are typically based on a sample and limited plot area and have a low update frequency due to practical and economic reasons. More recently, structural information has been generated by remote sensing information, e.g., from satellites that enable wall-to-wall mapping of forest structure over continental to global extents. Canopy structure, for instance, has been estimated from satellite imagery using machine learning approaches, which are often calibrated with ALS or spaceborne LiDAR data (Lang et al., 2023; Liu et al., 2023; Potapov et al., 2021).

The launch of spaceborne LiDAR data, such as acquired by Global Ecosystem Dynamics Investigation (GEDI) instrument or Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2) mission, facilitated the continuous, large-scale mapping and monitoring of forest structure. GEDI is a full waveform lidar instrument that is mounted on the International Space Station (ISS). By measuring the reflected energy from emitted pulses of laser light, it quantifies the vertical distribution of the vegetation, from which forest structure metrics can be derived. The instrument covers the Earth surface between 51.6°N and 51.6°S using footprints of about 25m in diameter (Dubayah et al., 2020). In contrast to GEDI, the Advanced Topographic Laser Altimeter System (ATLAS) instrument onboard ICEsat-2 is a photon-counting laser altimeter that was mainly optimized to monitor glaciers, sea ice, and ice sheets and covers the Earth surface between 88°N and 88°S (Abdalati et al., 2010).

Forest structure metrics derived from LiDAR data are usually divided into three main groups: metrics related to (i) canopy height, (ii) canopy cover, and (iii) vertical variability. The GEDI instrument provides metrics within each of the three categories. Canopy height is often represented by the relative height (RH) of the returns. Here, we focus on the relative height at the 98^{th} percent of the returns (RH98), as the maximum can have unrealistic height values due to noise in the waveform. Canopy cover is the percentage of the ground covered by the vertical projection of canopy material (i.e. leaves, branches and stems only) and thus represents forest land cover (in contrast to forest land used in Section 1). GEDI provides cover information at multiple heights, representing the horizontally intercepted canopy elements at a given height. Here, we focus on the canopy cover at 5 m height using training data from the GEDI instrument. Note that this metric differs from other widely used cover types, i.e. "the proportion of the vegetation over a segment of the sky hemisphere at one point on the ground", or crown cover as "the percentage of the ground covered by a vertical projection of the outermost perimeter of the natural spread of the foliage of plants". Foliage height diversity (FHD) is a measure of vertical variability and was selected as a metric describing the vertical complexity of forest stands. FHD is calculated using the Shannon's diversity index of the vertical LAI profile and provides an

indicator of the canopy structure complexity. A high FHD value is often associated with more complex forest structure (e.g. caused by multiple canopy layers).

Spaceborne LiDAR datasets have been intensively used to train models that allow the mapping of forest structure from satellite imagery. For example, Potapov et al. (2021) made a global canopy height map at 30m resolution by modelling the 95th percentile (RH95) canopy height from GEDI using annual Landsat metrics and random forest models. Over Europe, Turubanova et al. (2023) mapped changes in canopy height and extent between 2001 and 2021 from the Landsat archive. This model was calibrated using a combination of GEDI and ALS data. Several studies used deep learning approaches to map canopy height. Lang et al. (2023) trained a U-Net model using the 98th percentile height from GEDI and Sentinel-2 data, resulting in a global canopy height map at 10m resolution. Pauls et al. (2024) also trained a U-Net architecture to model and map canopy height globally at 10m resolution from Sentinel-1 and Sentinel-2 data using the RH100 metric from GEDI. The study focused on optimizing preprocessing techniques and employed a novel loss function to account for geolocation inaccuracies. Finally, several studies focused on deep learning approaches applied on high resolution imagery. Liu et al. (2023) for instance trained a deep learning model on ALS data and PlanetScope imagery to produce 3m resolution canopy height maps across Europe. Meta and World Resources Institute launched a global 1-meter resolution canopy height map using natural color imagery from Maxar Technologies. The map is the result of the combination of low-resolution predictions from a convolutional network trained on GEDI RH95 data and high-resolution predictions from an DPT decoder trained with aerial LiDAR data (Tolan et al., 2024).

While the mapping of canopy height using spaceborne and airborne LiDAR data has already been intensively studied, large scale mapping of the other forest structure metrics has received less attention. Kacic et al. (2023) for instance modelled and mapped both canopy height, canopy cover, and forest biomass dynamics over Germany for the years 2017 to 2022. Matasci et al. (2018) mapped several forest structure attributes, including canopy cover, at 30m resolution over the Canadian boreal zone using Landsat data and LiDAR plot data. In addition, - despite the relatively large number of canopy height maps that have been developed - most large-scale studies only focus on GEDI and/or ALS data to train the canopy height models. The use of GEDI data is however not available in the northern part of Europe, in contrast to ICESat-2 data.

Therefore, we aim to:

- Model canopy height, canopy cover, and FHD at 10m resolution for the year 2020 using Copernicus Sentinel-1 and Sentinel-2 data across Europe,
- Compare canopy height models trained using GEDI, ICESat-2, and the combination of GEDI and ICESat-2 data, and
- Evaluate the canopy height models across Europe using ALS data

3.2 Materials and methods

3.2.1 Spaceborne LiDAR data

3.2.1.1 GEDI

To train the forest structure models, we collected spaceborne LiDAR data acquired by Global Ecosystem Dynamics Investigation (GEDI) (Dubayah et al., 2020). GEDI L2A relative height metrics (RH98) and GEDI L2B canopy cover and FHD data were collected over the growing season (months April until October) and of the years 2019-2021. The data were filtered to keep high quality observations (based on their quality flag) with a sensitivity larger than 0.95 and without degradation, acquired during nighttime, during the leaf-on season (i.e., based on the leaf-on flag), and using a power beam.

Since LiDAR observations close to buildings may introduce noise in the dataset, these observations were removed using a 50m-by-50m window based on OpenStreetMap building data from the year 2020. Moreover, since LiDAR observations are known to be affected by topography (Adam et al., 2020; Fayad et al., 2021), leading to height overestimations, we further excluded observations with a slope larger than 10% using the 30m resolution Copernicus Global DEM. Finally, noisy observations were further reduced by excluding observations with a height larger than 60m.

3.2.1.2 ICESat-2

Since the GEDI data do not cover the northern part of Europe, we additionally collected relative height metrics derived from Ice, Cloud and land Elevation Satellite-2 (ICESat-2) observatory data (ATL08 product v005) to train a canopy height model. The ICESat-2 observatory and Advanced Topographic Laser Altimeter System (ATLAS) instrument utilize a photon-counting lidar that covers the Earth between a latitude of approximately 88° N to 88° S, and thus complements GEDI data that is only available until 53° North. The instrument covers three pairs of spots on the ground, forming six ground tracks that are typically about 14m wide. The beams have different transmit energies (i.e. weak and strong beams), with an energy ratio between the weak and strong beams of about one against four (Neuenschwander et al., 2019).

We used the 98% height of all the individual relative canopy heights (height above terrain) for the 20m geosegments of the ATL08 product as a measure of canopy height. Heights are only provided for segments that have sufficient photons, i.e., at least 10 signal photons and 3 canopy photons are required. Note that – in contrast to GEDI - the ICESat-2 canopy height metric does not include ground photons. Hence, it represents the height above the ground surface.

The ICESat-2 observations were collected over Europe for the months April – September over the years 2019 – 2021. Like the GEDI observations, we only selected observations from strong beams, that were not located close to buildings, were not located on slopes higher than 10%, and had a height lower than 60m.

3.2.2 Predictor variables

3.2.2.1 Sentinel-1 and -2 features

Yearly temporal statistics derived from Copernicus Sentinel-1 and Sentinel-2 data were used as predictor variables in the canopy height models. The Sentinel-2 features were derived from Sentinel 2 Level 2A products of the year 2020, which were first selected and filtered based on cloud cover. We excluded clouds, cloud shadows, cirrus, and saturated pixels from the Sentinel-2 bands based on the scene classification layer of the L2A product. To further reduce residual

noise in the time series, ten days median composites with twenty days moving window were subsequently computed. The remaining missing values were filled using a linear interpolation of neighboring available pixels in the time series. After pre-processing the Sentinel-2 data, a set of vegetation indices were calculated, including the NDVI, EVI, NIRV, NDWI, NDGI, NDMI, NBR, NBR2, REP, ANIR, NDRE2, and NDRE3. Next, the time series of pre-processed Sentinel-2 data and vegetation indices were used to derive a set of descriptive statistics, i.e. features. These include the 10th, 50th, 90th, and interquartile range (difference between 75th and 25th percentile) for each band (B02, B03, B04, B05, B06, B07, B08, B11, and B12) and vegetation index. In addition, NDVI time series that were down-sampled to 6 timestamps using Fourier methods were added as feature.

The Sentinel-1 features were derived in a similar fashion as the Sentinel-2 features. The Sentinel-1 Ground Range Detected (GRD) products were first corrected to Gamma0 backscatter and geocoded to the Sentinel-2 grid. Maximum one observation per day and per tile was processed to reduce data redundancy. A multitemporal speckle filter was then applied to reduce noise, after which the time series were composited to ten daily composites. Finally, the percentiles and interquartile range were computed on the Sentinel-1 VV and VH backscatter, the difference between the VH and VV backscatter, and the radar vegetation index (RVI) time series.

3.2.2.2 Localization features

Next to the Sentinel-1 and Sentinel-2 data, several features were used to enable the model to localize in space, including the terrain height from the Copernicus DEM, and the latitude and longitude of the observation. The inclusion of such localizing features facilitates the usage of a single model instead of multiple regional models and has previously been shown to improve the prediction accuracy (Lang et al., 2023).

3.2.3 Model training

3.2.3.1 Canopy height models

Separate canopy height models were trained for each of the spaceborne LiDAR datasets, i.e. using GEDI, ICESat-2, and the combination of GEDI and ICESat-2 data as reference data.

First, training, validation, and test data were sampled using a fixed 100x100 km grid. Since the spaceborne LiDAR data may show high values for the relative height metric outside areas covered by forests (e.g. due to the presence of buildings, tall crops, or water bodies) and hence are not always representing canopy height in these areas, we used a separate sampling approach inside, and outside areas covered by woody vegetation. Per grid cell, we sampled maximally 2500 observations over treed areas and targeted a ratio of 0.4 between the locations outside and within woody areas. 80% of the grid cells were then assigned to training, 10% to validation, and 10% to testing.

Over areas with woody vegetation, relative height observations were sampled from locations where GEDI or ICESat-2 data were available. For each of the samples, the independent variables (i.e., the features) were extracted by taking the mean over a 3x3 10m pixel window to represent the 25m GEDI observations and a single 10m pixel to represent the ICESat-2 observations. Areas covered with trees were delineated based on the intersection of the Copernicus High Resolution

forest type layer of 2018 and the forest and shrubland areas within the ESA WorldCover land cover map of 2020 and 2021 (EEA, 2020; Zanaga et al., 2022; Zanaga et al., 2021).

Over areas without woody vegetation, the relative height metric was set to zero. Areas outside woody vegetation were delineated according to intersection of the no-forest areas within the Copernicus High Resolution layer of 2018 and the areas without woody vegetation within the ESA WorldCover land cover map of 2020 and 2021 (EEA, 2020; Zanaga et al., 2022; Zanaga et al., 2021). To ensure that the samples outside woody vegetation represent different land cover types without oversampling dominant classes, the sampling was performed using the ESA WorldCover land cover map of 2020. We sampled 10% of the pixels per land cover type, with a maximum of 50 samples per land cover type within a 10x10 km area.

Next, a multi-quantile CatBoost regression model (Dorogush et al., 2018) was trained using the set of Sentinel-1 and Sentinel-2 temporal features of the year 2020, elevation, and location (i.e. latitude and longitude) as independent variable. To avoid the model overfit on the location, we added noise to the latitude and longitude that was randomly sampled between -1 and 1 degrees. Moreover, since the number of data samples was not equally distributed over the range of the canopy height metric, we computed a histogram and assigned a weight inversely proportional to the squared root of the number of observations within each bin. To avoid height bins with very few observations to result in very high weights and lead to overfitting, we saturated the weight for heights higher than 35 m.

A multi-quantile loss function was applied to enable the model to predict multiple target quantiles, allowing to gain information about the conditional target distribution and provide an uncertainty estimate. Like Poggio et al. (2021), three quantiles were used: 0.05 (q0.05), 0.5 (median, q0.5), and 0.95 (q0.95). The 0.5 quantile (q0.5) predictions were used to estimate canopy height, where negative estimation values were set to zero. Uncertainty was mapped using the 90th prediction interval (PI90) and prediction interval ratio (PIR) as indictor (Eq. 1 $PI_{90} = q_{0.95} - \ q0.05$

and 2).

$$PI_{90} = q_{0.95} - \ q_{0.05} \quad (1)$$

$$PIR = \frac{q_{0.95} - q_{0.05}}{q_{0.5}} \quad (2)$$

The canopy height models were evaluated over the test set using the root mean squared error (RMSE), coefficient of determination (R^2), the Mean Absolute Error (MAE), and the Mean Bias Estimate (MBE).

3.2.3.2 Foliage height diversity and canopy cover models

A similar modelling approach as the canopy height models was used to train the FHD and canopy cover models. Yet, there are three differences between the approaches. First, since there is no logical FHD value outside woody areas, we only trained the FHD model on observations located

in woody areas. During inference, we used the canopy height estimations to identify the pixels that are likely not located in woody areas (i.e. they have a height lower than 3 m) and set the values of these pixels to zero. Second, to ensure consistency between the canopy height and the canopy cover estimations, we also set the canopy cover of pixels with a canopy height lower than 3 m to zero. Third, no saturation of the weights was applied to the canopy cover and FHD models.

3.2.4 Validation of canopy height estimations with ALS data

To independently validate the canopy height datasets, a validation dataset was prepared using freely available Airborne Laser Scanning (ALS) data over a set of 640 by 640 m patches across Europe. The patches were selected to (i) be equally distributed across Europe, and (ii) cover areas that are partially covered by trees or have been covered by trees in the past.

To ensure an equal distribution of patches across Europe, we defined a grid of 100 by 100 km cells over Europe and tried to sample ten patches within each grid cell. Patch locations were selected that were covered at least for 50% by trees as defined by the union of the Copernicus High Resolution Forest Type layers of 2012, 2015, and 2018. The idea of this approach is to collect data over areas that were potentially covered by trees. Yet, since the ALS data over the patches are collected over different dates, the actual tree cover over the patch at the flight date might be lower than 50%. In addition, we noticed that for a minority of the patches also urban areas with high tree cover were included.

Since each of the selected patches can be covered by one or more LAZ-files, we first collected the LAZ files per patch, clipped the data to the dimensions of the patch, and merged them into a single LAZ file. Only the first returns were utilized during this stage and, consequently, throughout all subsequent processing steps.

The point cloud data of the patch were subsequently normalized, resulting in the above-ground height of each LiDAR point. The normalization uses the a-priori classification and identification of ground points within the downloaded point clouds. These ground points were triangulated into a ground triangular irregular network (TIN) and the elevation of each point with respect to this TIN was then computed.

Since the LAZ-files are commonly in a local coordinate system, the LiDAR point cloud data were reprojected to the Universal Transverse Mercator (UTM) projection, aligning with the Copernicus Sentinel-2 grid. The resulting LiDAR point cloud data may still contain noise, such as isolated points due to power lines or birds. To reduce the effect of these noisy, isolated points on the vegetation structure metrics, a denoising operation was applied. Here, points are eliminated if they have less than f_d other points in their surrounding 3 by 3 by 3 grid of cells, each having a size of 2 by 2 by 1 meter. The most suitable denoising parameter f_d depends on the characteristics of the data (such as point density) as was thus defined per country.

Finally, the data was filtered by flight date to maximize dataset consistency. The LiDAR point cloud data for a given patch may originate from various flight lines collected on different dates. Significant time differences between these dates, such as one part of the patch being flown in summer and another in winter, can result in heterogeneous data. To avoid this issue, a maximum

allowable time difference of 30 days was established. If the time difference between flight dates exceeded 30 days, only data from the most recent flight date was used.

The height at the 98th percentile, i.e. the height at which 98 percent of the first returns above a height cutoff of zero meter has been recorded, was then derived from the LiDAR point cloud as a measure of canopy height. The canopy height metric was calculated on a grid of 10 by 10 meter, aligning with the grid and spatial resolution of Copernicus Sentinel-2 data and resulting in a raster size of 64 by 64 pixels. Additionally, per-pixel metadata such as point density and flight date were recorded alongside the vegetation structure metrics.

Despite denoising efforts, invalid canopy information may remain in the dataset due to e.g. remaining noisy points over power lines or windmills. Hence, a thorough screening and visual inspection of the vegetation structure metric dataset was performed. Pixels that showed artificial anomalies were identified and tagged using a point vector file. These invalid pixels were then assigned a NODATA value. Moreover, patches with clear large-scale artefacts were excluded from the dataset. This resulted in about 3400 patches across 18 countries in Europe. The patches cover a gradient from the north of Europe (Scandinavia) to the south of Europe (Spain) but show a gap in the Balkan countries and Italy (Figure 14).

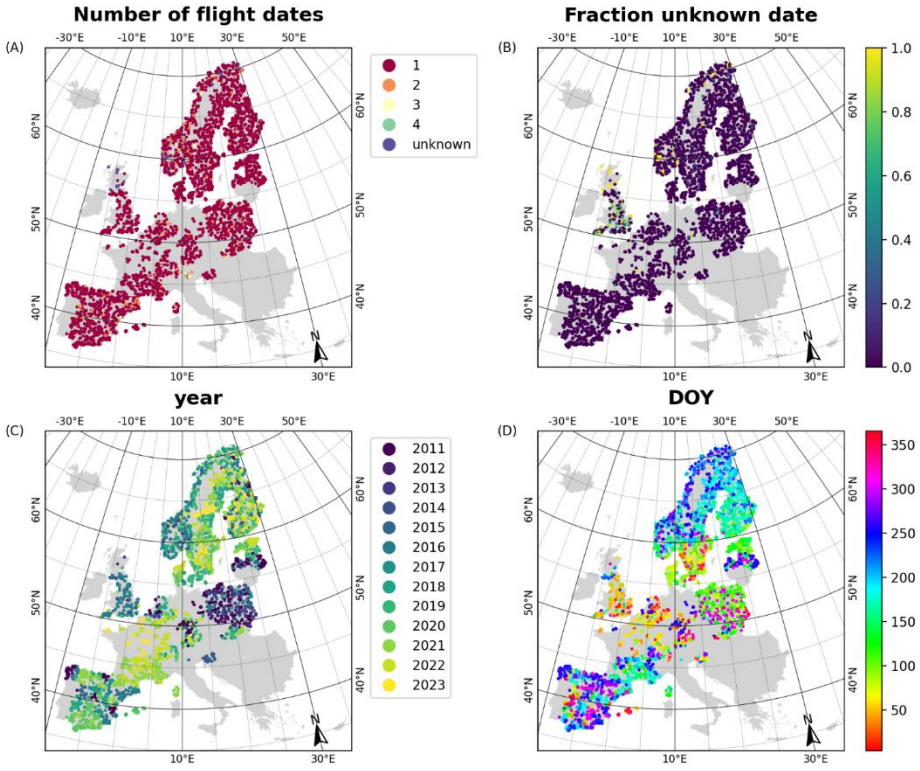

Figure 14: Flight days for the ALS patches. The number of flight days within each patch (A), the fraction of pixels within each patch with an unknown date (B), the flight year for patches with a single flight date (C), and the DOY for patches with a single flight date (D).

The ALS-based validation dataset contains about 3500 10 m resolution patches and covers a large share of Europe, except for Italy and the Balkan countries. This extensive dataset spans various and diverse climate and forest zones, ranging from boreal coniferous forest to temperate oceanic, and subtropical dry forest (De Rigo et al., 2016). Gathering detailed information on vegetation structure across such a vast area and at this resolution would be too expensive and labour intensive without LiDAR observations.

An important challenge in preparing the ALS-based dataset was the data acquisition. Since the ALS data were stored and maintained at regional or national levels, information about the data was sometimes only available in local languages, hampering finding the correct access points to the data. Moreover, information about the tiling grid was not always available or was provided in a non-georeferenced format. The means for accessing the data also varied significantly. In some countries, the data could easily be downloaded using a script, while in others, it required manually clicking on download links.

The validation dataset was constructed by sampling ten pixels within each of the ALS patches. Since the relative height metric might be impacted by the presence of buildings, water, or steep slopes, we excluded ALS data that was flagged as build-up areas or water in the WorldCover V200 dataset (Zanaga et al., 2022) and slopes larger than 20%.

As the GEDI data only covers an area up to 51.6° N, we evaluated the canopy height models using (i) all ALS patches, (ii) the patches north of 51.6°N (i.e. outside the area covered by GEDI training data), and (iii) the patches south of 51.6°N.

3.3 Results and discussion

3.3.1 Canopy height

3.3.1.1 Canopy height maps

The prediction interval ratio (PIR) of the height estimations trained on the combination of ICESat-2 and GEDI data (Figure 15) shows that the highest relative model uncertainty can be found in Spain, England, Ireland, Italy, and Norway. The model trained on GEDI data shows higher height estimations over most of the northern part of Europe (above 55 degrees latitude) than the model trained on both spaceborne LiDAR datasets. The difference between the height estimates below 55 degrees latitude are less pronounced and mixed. Also, for the model trained on ICESat-2 data, higher estimations can be found in the north of Europe, yet less pronounced as compared to the GEDI data. Below 55 degrees latitude, the model trained on ICESat-2 data tends to produce slightly lower height estimations than the model using both LiDAR datasets.

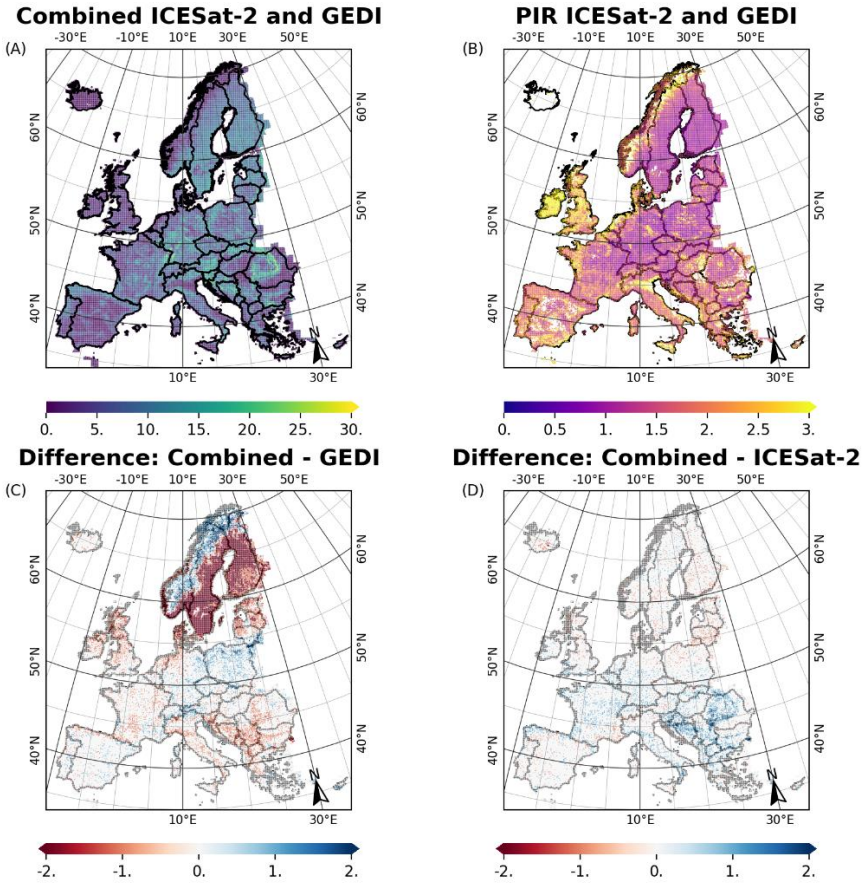

Figure 15: Canopy height map based on the combination of ICESat-2 and GEDI data (A), the prediction interval ratio (PIR) of the canopy height estimation using both ICESat-2 and GEDI data (B), the difference between the canopy height estimations using the combination of ICESat-2 and GEDI data and GEDI data (C) the difference between the canopy height estimations produced by the model using the combination of ICESat-2 and GEDI data and ICESat-2 data (D)

Comparing the canopy height map based on GEDI and ICESat-2 data with the average height map produced by Pucher et al. (2022) at 8 km resolution shows general agreement between both datasets (Figure 16). The average height of Pucher et al. (2022) however shows higher values over Hungary and lower values over the Carpathian Mountains.

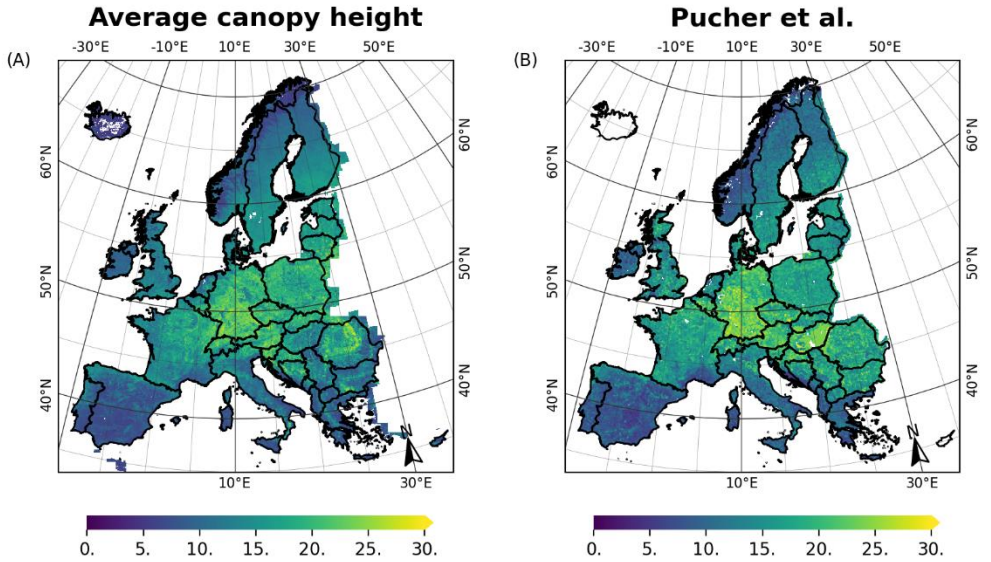

Figure 16: Canopy height map based on GEDI and ICESat-2 data, aggregated to 8 km by taking the mean over pixels that show a height larger than 5m (A) and the average canopy height map produced by Pucher et al. (2022) (B).

3.3.1.2 Evaluation of the canopy height maps on the test set

Evaluating the height maps on the test set results in a Mean Absolute Error (MAE) around 3m, a Mean Bias Estimate (MBE) around 0.5m, an R^2 around 0.75, and a Root Mean Squared Error (RMSE) around 4.5 m. The accuracy metrics are a bit worse when using only observations that were labelled as woody vegetation, with a Mean Absolute Error around 4 m, an R^2 around 0.5, and a Root Mean Squared Error around 5.5 m (Table 8 and Figure 17).

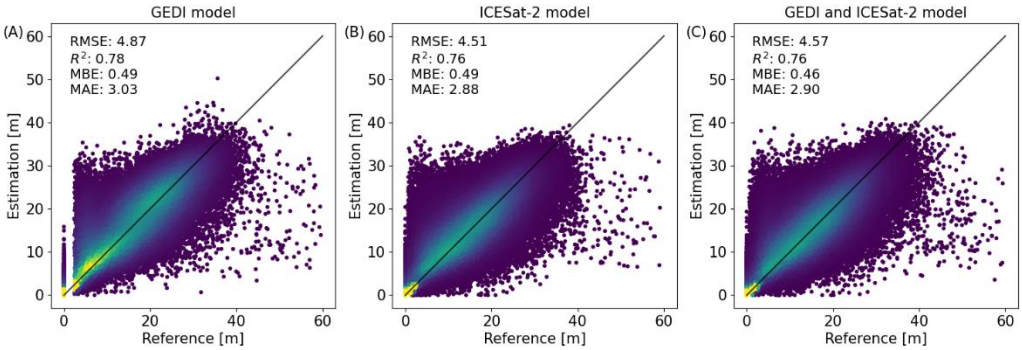

Figure 17: Evaluation of the models on the test set for the model: trained on GEDI data only (A), ICESat-2 data only (B), and GEDI and ICESat-2 data (C). The colors denote the density of the points.

Table 8: Overview of the accuracy metrics (RMSE, R2, MBE, MAE) over the test set for the three models. Models were evaluated using all observations (inside and outside woody areas) and using observations that are located within woody areas.

System	Subset	RMSE	R^2	MBE	MAE
GEDI	All observations	4.87	0.78	0.49	3.03
	Woody observations	5.75	0.56	0.60	4.15
ICESat-2	All observations	4.51	0.76	0.49	2.88
	Woody observations	5.30	0.49	0.58	3.92
GEDI and ICESat-2	All observations	4.57	0.76	0.46	2.90
	Woody observations	5.38	0.51	0.55	3.95

3.3.1.3 Evaluation of the canopy height maps with ALS data

Comparing the height estimations of the three models with ALS data shows that the canopy height models using GEDI data only, ICESat-2 data only, and both datasets perform similarly in the southern part (Figure 20). In the northern part, no GEDI data was available for training, resulting in a drop in accuracy of the GEDI-based model (Figure 18 and Figure 19). Overall, the ICESat-2 and combined models reach a RMSE of 5.8 m and a MAE of 5.1 m over the ALS-based validation set, which is an improvement in accuracy compared to existing maps.

Evaluating the estimation errors over height bins shows that all canopy height maps have the tendency to underestimate height for taller trees (Figure 21). Over the lower canopy height range, the canopy height estimations of the GEDI and ICESat-2 based model performs similarly as the maps provided by Potapov et al. (2021) and Tolan et al. (2024), whereas Lang et al. (2023) tends to overestimate height over the 0-15m range and Liu et al. (2023) over the 0-5m range. Over the higher canopy height range, the height underestimation of our canopy height map is a bit better than those of Potapov et al. (2021). The height underestimation of maps provided by (Liu et al., 2023) and Tolan et al. (2024) are the worst, whereas the error is the lowest for the map provided by Lang et al. (2023).

Besides errors in the height estimations, the maps have different spatial resolutions. The high spatial resolution (one meter) of the canopy height map of Tolan et al. (2024) has the advantage of better discriminating fine details (e.g. small gaps in a forest or sparse trees) (Figure 22). Yet, the map also shows some block artifacts, which are, however, more pronounced outside Europe (e.g. in the tropics). The map derived by Liu et al. (2023) was also produced at high resolution (3 m) but was made available on Zenodo at 30 m resolution and was therefore only evaluated at the latter resolution. The GEDI and ICESat-2 canopy height map tends to better capture spatial details than the 10m resolution map of Lang et al. (2023) (e.g. Figure 22 C and H), or the 30m resolution map of Potapov et al. (2021).

Even though the ALS-based validation dataset provides unique and valuable information on vegetation structure, it also shows several limitations. The ALS data across Europe is heterogeneous, having been collected using different instruments and on various dates. For instance, some data from Poland was collected in 2011, whereas parts of France were surveyed in 2023. Secondly, also the point density of the ALS measurements differs across the dataset. Lower point densities reduce the capacity to capture the fine-scale spatial positioning of vegetation components, potentially leading to less reliable vegetation structure metrics. Due to these limitations, the accuracies of the maps may be underestimated. In addition, it should be noted that each of the maps uses different canopy height metrics. Whereas our study and the map of Lang et al. (2023) used the 98th percentile height, Tolan et al. (2024) and Potapov et al.

(2021) focused on the 95th percentile height. Since all maps were evaluated with the 98th percentile height, a part of the error could potentially be attributed to differences in the height metrics.

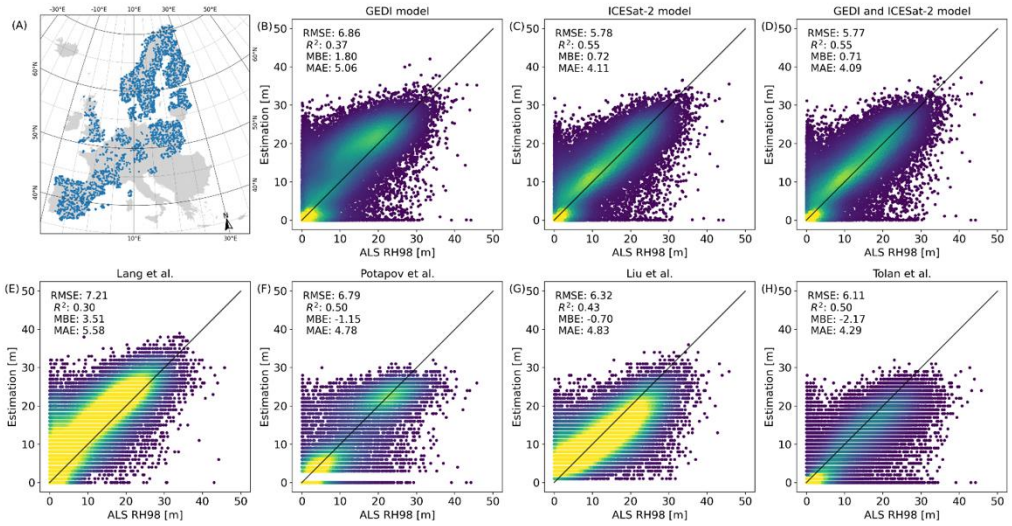

Figure 18: Validation of the canopy height estimations using ALS data (A) for the canopy height model using GEDI data only (B), ICESat-2 data only (C), the combination of GEDI and ICESat-2 data (D), and the height estimations from Lang et al. (2023) (E), Potapov et al. (2021) (F), Liu et al. (2023) (G), and Tolan et al. (2024) (H).

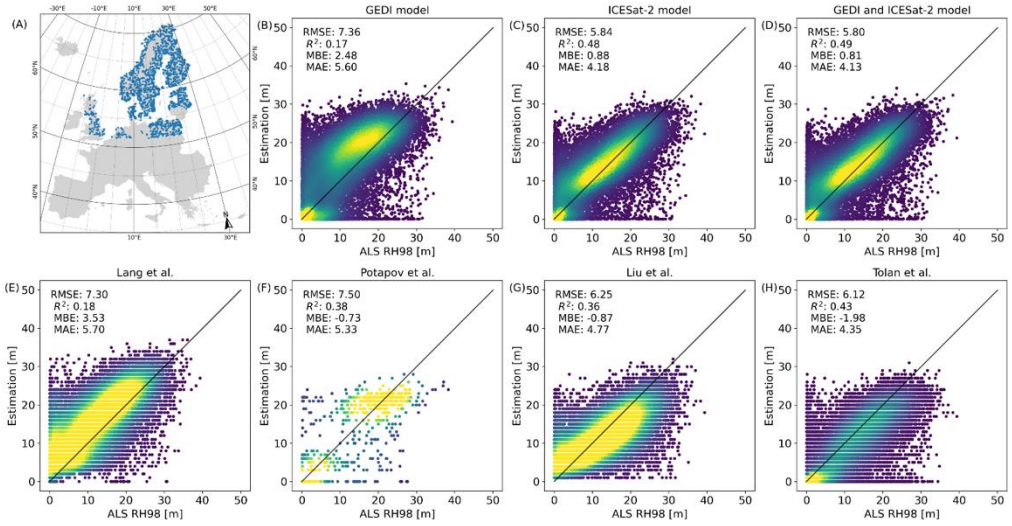

Figure 19: Validation of the canopy height estimations using ALS data in the northern part of Europe (A) for the canopy height model using GEDI data only (B), ICESat-2 data only (C), the combination of GEDI and ICESat-2 data (D), and the height estimations from Lang et al. (2023) (E), Potapov et al. (2021) (F), Liu et al. (2023) (G), and Tolan et al. (2024) (H).

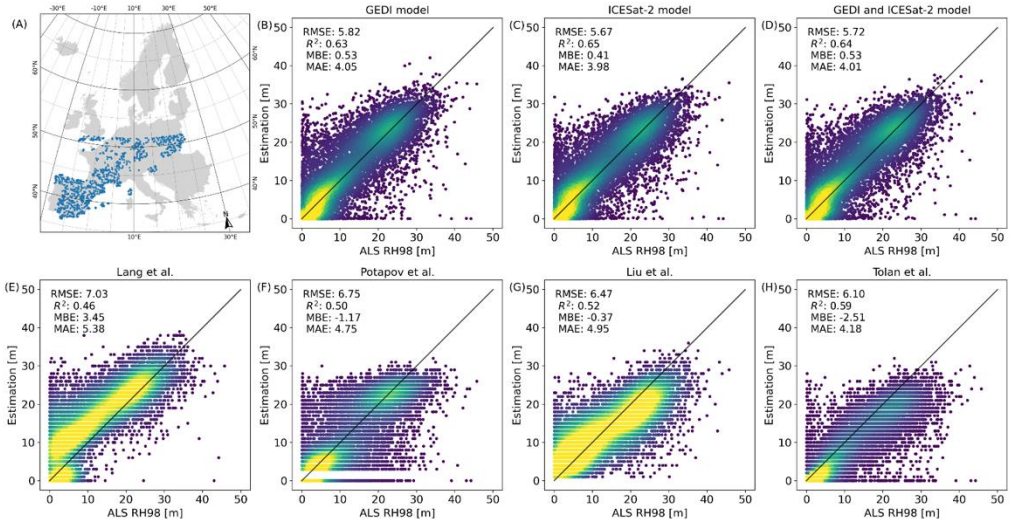

Figure 20: Validation of the canopy height estimations using ALS data in the southern part of Europe (A) for the canopy height model using GEDI data only (B), ICESat-2 data only (C), the combination of GEDI and ICESat-2 data (D), and the height estimations from Lang et al. (2023) (E), Potapov et al. (2021) (F), Liu et al. (2023) (G), and Tolan et al. (2024) (H).

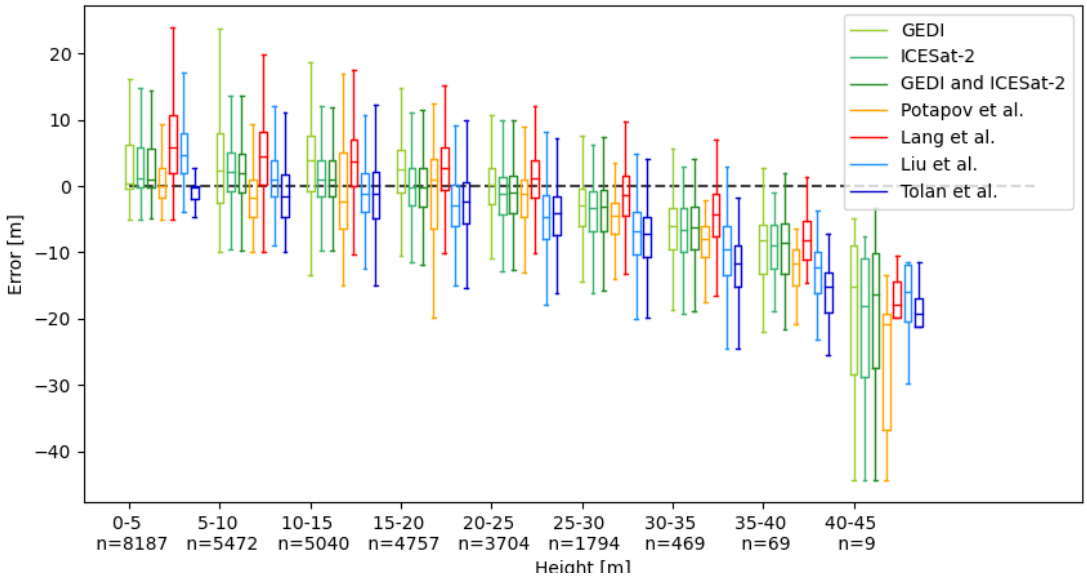

Figure 21: Error (estimated height – ALS height) per height interval. The number of observations is given along the x axis. For some datasets (e.g. those from Potapov et al.) this number might be smaller.

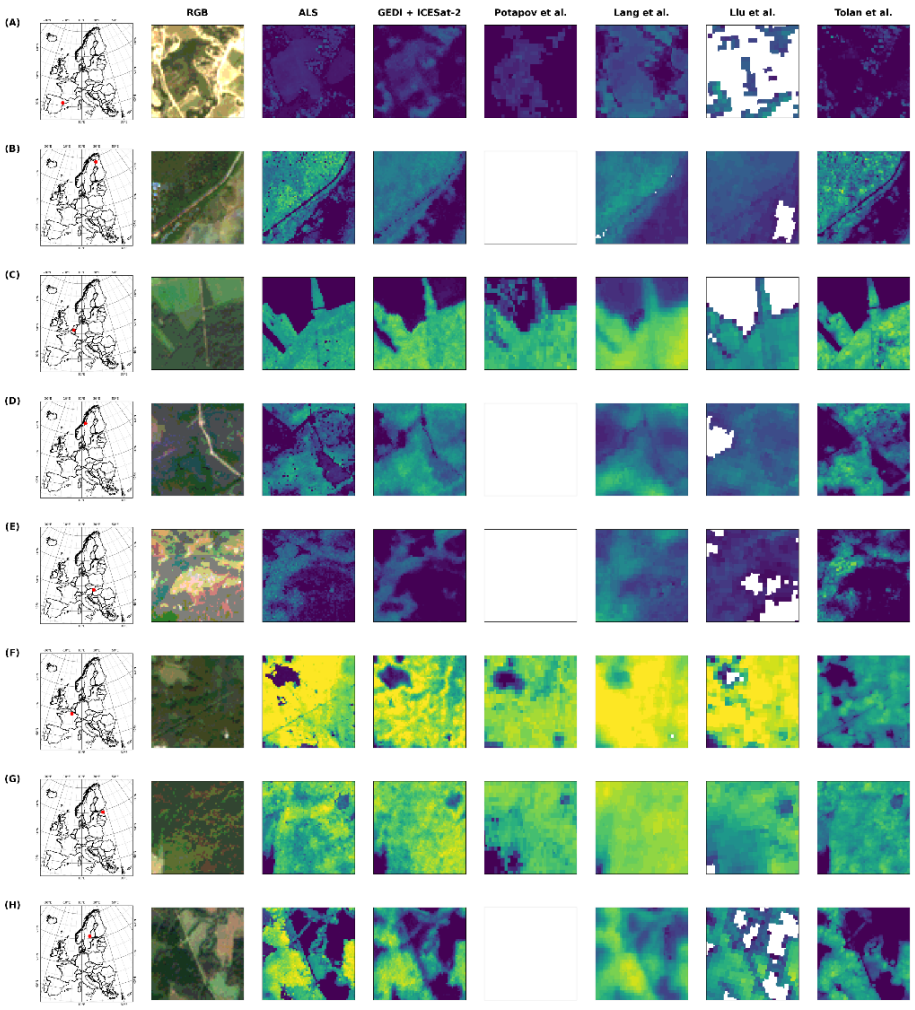

Figure 22: Illustration of canopy height estimations over a set of patches across Europe. From left to right, the location of the patch, an RGB image of the patch using Copernicus Sentinel-2 data, canopy height estimations from ALS data, the model using GEDI and ICESat-2 data, and the estimations from Potapov et al. (2021) Lang et al. (2023), Liu et al. (2023), and Tolan et al. (2024) are shown.

3.3.2 Canopy cover and Foliage Height Diversity

Whereas most studies focus on estimating canopy height, also canopy cover and FHD were estimated over Europe (Figure 23). The canopy cover model reaches a RMSE of 0.22, an R^2 of 0.59, MBE of 0.04, and MAE of 0.13 and the FHD reached a RMSE of 0.5, R^2 of 0.37, MBE of -0.01, and MAE of 0.34 over the test set (Figure 24). While the model can capture the main cover and FHD trend, there is some overestimation of low values (FHD lower than one and cover lower than 0.1 notable).

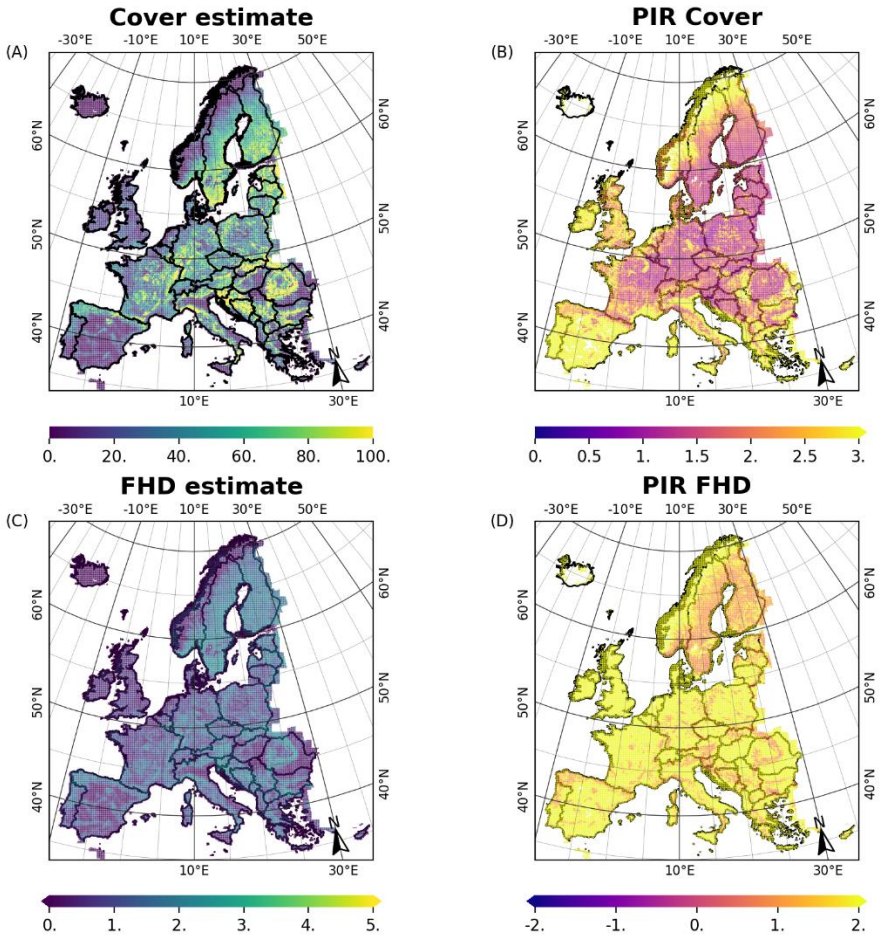

Figure 23: Canopy cover (at 5m height) (A) and FHD (B) maps over Europe and their associated PIR maps (B and D, respectively)

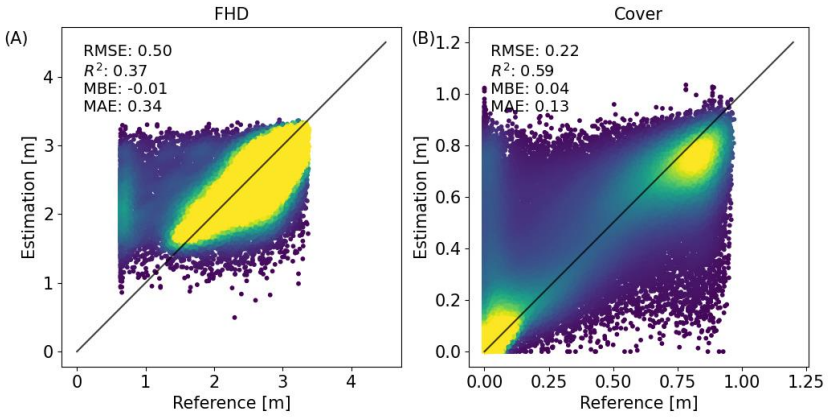

Figure 24: Evaluation of the FHD and canopy cover estimations over the test set

3.4 Computational costs of structure mapping

The forest structure maps were derived from a set of Sentinel-1 and -2 features that were pre-computed and stored in the framework of the ESA WorldCover project. The storage of these features over Europe takes around 2 TB. Further optimization of the processing and storage cost of the spectral features is, however, possible. As the spectral bands and vegetation indices are correlated, the number of features could likely be reduced with limited effects on the map accuracy. The inference of the canopy height map was run in Python on Linux and took around seven hours on a cluster where maximally 500 executors were utilized simultaneously, each having access to 4 GB memory.

3.5 Conclusions

Three types of forest structure metrics, i.e. canopy height, canopy cover, and foliage height diversity, were modeled and mapped at 10 m spatial resolution over Europe using spaceborne LiDAR data, Sentinel-1 and -2 imagery, and localizing features. Validation of the canopy height products with an independent airborne LiDAR dataset resulted in a mean absolute error around 4 m and a root mean squared error around 5.8 m and indicated that the maps performed similarly or showed an improvement against canopy height maps found in literature. It should be noted that the canopy height map tends to underestimate the height of taller trees, which is a phenomenon that was also found for other maps in literature.

The canopy height models were trained using two types of spaceborne LiDAR data: GEDI and ICESat-2 data. Since no GEDI data is available in the north of Europe, the model trained with only GEDI data showed larger errors in that area. The other models (based on ICESat-2 or the combination of GEDI and ICESat-2) are thus recommended above the GEDI-based model.

The modeling of the canopy cover and foliage height diversity metrics seemed to be more challenging. The overall trends in foliage height diversity and cover could be mapped, yet the model shows some overestimation for low foliage height diversity and canopy cover values.

4 Forest composition maps

4.1 Introduction

Wall-to-wall information on tree species distributions at high spatial resolution is important to assess biodiversity or woody biomass, initializing forest simulation models, to support sustainable forest management, and to support climate change adaptation and mitigation (Ruiz-Benito et al., 2020; Vihervaara et al., 2017). In many countries, national forest inventory (NFI) data provides a highly valuable source of information on tree species composition, enabling assessments at regional or country scale. Yet, the coverage of the in-situ data is often limited to a set of samples and does not provide spatial explicit, wall-to-wall information on tree species composition.

To overcome this limitation, several studies used in-situ data to develop wall-to-wall maps of tree species at continental scales. Over Europe, Brus et al. (2012) used a combination of compositional kriging and a multiple logistic regression model to map the distribution of 23 tree species groups at 1 km resolution. In addition, San-Miguel-Ayanz et al. (2016) derived the frequency and chorology of tree species and modeled their relative probability of presence and

habitat suitability at 1 km spatial resolution. The modelling was based on a harmonized dataset of field observations extracted from a set of surveys that are available within the Forest Information System for Europe (FISE) and a set of bioclimatic parameters, solar irradiation and elevation range. Bonannella et al. (2022) further described a framework to produce potential and realized distribution maps for 16 tree species at 30 m resolution over the years 2000-2020 using an ensemble machine learning approach. Although these maps are highly valued, they were produced at low to medium resolution.

The combination of in-situ observations with high resolution remote sensing data and the rise of computing power may however support the use of machine and deep learning methods for wall-to-wall mapping of tree species composition at high spatial resolution. While several sensor types have been explored over relatively small areas to support tree species mapping, including airborne LiDAR, hyperspectral, multispectral data, or a combination of the former (Fassnacht et al., 2016), multispectral data has mostly been used to map species over larger spatial scales given the availability of consistent data at relatively low cost. In particular, the launch of Sentinel-1 and -2 data facilitated the mapping of tree species through their sufficiently high spatial resolution (10-20 m) and revisit time. The high revisit time of the sensors allows us to overcome issues with missing data due to cloud cover and produce cloud-free composites. In addition, it allows to include spatio-temporal information into the model and capture vegetation phenology, improving the accuracy of tree species models (Francini et al., 2024).

Several studies have already explored the high resolution national-scale mapping of tree species. Francini et al. (2024) for example trained a random forest model to map seven groups of species over the Netherlands using Sentinel-2 features and NFI data. Tree species maps have been generated over Germany over seven classes using regional XGBoost models based over Sentinel-2 (Welle et al., 2022) and eleven classes using a random forest classifier based on both Sentinel-1 and -2 data and a set of environmental features, such as climate, topography, soil, and meteorology (Blickensdörfer et al., 2024). In Norway, Breidenbach et al. (2021) modelled the distribution of spruce, pine, and broadleaf trees using a random forest model. In addition, 37 tree species were mapped over Canada using Landsat imagery, climate, terrain, and phenology data (Hermosilla et al., 2022). Besides tree-based classifiers, the use of deep learning models has been explored. Bolyn et al. (2022) for instance mapped tree species proportions for nine species groups over the Wallonia region in Belgium using a U-Net++ architecture.

Here, we aim to model the distribution of seven classes related to tree genus over Europe at 10m resolution using the combination of Sentinel-1 and Sentinel-2 data, topography, and climate data. Genus classes were selected that are either frequently occurring in Europe or are expected to show a distinctive spectral signature, i.e., Larix, Pinus, Picea, Fagus, Quercus, other needleleaf, other broadleaf, and no woody vegetation.

4.2 Materials and methods

4.2.1 Genus information

To train the genus model, we collected information on tree genus over Europe. Inventory data, such as national forest inventories, are an important source of information on forest composition, yet the availability of these data with exact coordinates is limited across Europe. To reduce the spatial gaps in the training data, we combined data from different sources.

4.2.1.1 Harmonized Tree Species Occurrence Points for Europe

First, we collected species occurrence data from the harmonized tree species occurrence points dataset (Heisig & Hengl, 2020). This dataset harmonizes and combines occurrence data from GBIF, the EU-Forest project (Mauri et al., 2017) and the LUCAS survey for tree species that occur in the European Atlas of Tree Species (San-Miguel-Ayanz et al., 2016). Since some of the observations within this dataset may have a large spatial uncertainty, we used supplied variables to filter out observations with a location accuracy higher than 10m, that had no known issues regarding the date or location and were not reported to be outside the natural geographical species range. More information about the dataset can be found in Bonannella et al. (2022).

4.2.1.2 National Forest Inventories

Next, we collected observations from NFI over Flanders (Belgium), the Netherlands, Germany, Sweden, and Spain. All observations had either exact coordinates or the predictor variables were extracted by the NFI over the plot locations using the exact coordinates (e.g., for Germany).

The NFI data of Flanders contained measurements collected during the second (VBI2) and third inventory (VBI3), which started in 2009 and 2019, respectively (ANF, 2023). The Flemish NFI use a fixed sampling grid of 0.5 by 1 km, where plots located in forest are measured with a revisit cycle of about 10 years. The measurements are conducted in three concentric circles. The plot with the smallest radius (radius 4.5m) contains trees with a circumference smaller than 22cm and a height smaller than two meters, whereas the intermediate plot (radius of 9m) contains trees or coppice with a circumference between 22 and 122 cm, and the largest plot (radius 18m) contains trees or coppice with a circumference larger than 122 cm (Govaere, 2019; Govaere et al., 2009). More information on the Flemish NFI and how to access it can be found here: https://www.natuurenbos.be/dossiers/bosinventaris.

To derive the dominant genus from the individual tree measurements per plot, we first filtered the measurements based on tree status, where we only kept observations of trees that were alive. Next, the basal area of each tree was scaled based on the size of its concentric measurement plot. The fraction of the scaled basal area per genus class was then computed per plot and the genus class with the highest basal area fraction was assigned to each plot.

Next, we used individual tree measurements from the seventh forest inventory (NBI-7) of the Netherlands (Schelhaas et al., 2022), which were collected during the years 2017-2021. The NFI data in the Netherlands use a semi-regular grid as sampling design with one sample per 100 ha. The circular measurement plots have a variable radius, ranging between 5 to 20 m and including at least 20 trees. Like the Flemish data, we first filtered the tree measurements based on their status. Measurements of trees that were removed, dead, broken, or fallen were excluded. For each plot, the basal area fraction per genus class was subsequently computed and the genus class with the highest basal area fraction was assigned to the plot.

The NFI data over Germany were extracted from the third inventory, collected over the years 2011 and 2012 (Riedel, 2017). The NFI in Germany are measured over about 60 000 permanent plots, which are revisited every ten years. The permanent cluster sample plots are in a grid, with varying grid sizes among federal states. The clusters consist of four plots, each separated by 150m.

Variable-radius sampling is used to survey the trees, where the probability of sampling a tree depends on its basal area. Based on the individual tree measurements per plot, we computed per plot the fraction of the basal area for each genus class. In addition, we computed the basal area fraction using the species information available over a larger sample of trees at each plot (i.e. baanteile). Plots were only included if the dominant genus from the two assessments corresponded.

In Spain, NFI data from the fourth inventory, whose campaign started in the year 2008, were used (Alberdi et al., 2017; Álvarez-González et al., 2014). The permanent inventory plots are located on the nodes of a one-by-one km grid and follow an inventory cycle of ten years. The plots consist of four concentric circular areas, having a radius of 5, 10, 15, and 25 m. The dominant genus class per plot was derived from the basal area estimates that were provided by plot and grouped by species and diameter class. Like the other NFI data, the basal area fractions per genus class were computed and the dominant genus class was assigned per plot as the genus class with the highest basal area fraction.

In Sweden, inventory data collected on the NFI temporary plots over the years 2007 – 2022 were used (Fridman et al., 2014)[1]. The plots are sampled along five strata, with a decreasing sampling intensity towards the north. 40% of the annually measured plots are temporary plots, which are only measured once. These temporary plots have a radius of 7 m and consist of different concentric circles. The survey of trees in each of the concentric areas depends on their diameter at breast height. The NFI dataset over the temporary plots contains amongst others the basal area for a set of tree species per plot, i.e. Pine, Spruce, Birch, Contorta, and other deciduous trees. Based on these basal area estimates, we computed the basal area fraction for each genus class and defined the dominant genus as the genus class with the highest basal area fraction.

4.2.1.3 Private forest Finland

In addition to the NFI data, we used inventory data over private forests in Finland provided by the Finnish Forest center. The inventory data contains the diameter and species for sample trees within each of the plots. The Finnish Forest center additionally estimated the diameter of the remaining trees based on field data and remote sensing data. Like the NFI data, we computed the basal area per genus from the estimated and sampled data and assigned the genus with the highest basal area as the dominant genus per plot. Only plots with an area smaller than 400 m^2 were included.

4.2.1.4 BD Foret

Finally, we collection tree genus information in France through the BD Foret product (IGN, 2016). BD Foret has been produced between the years 2007 and 2018 by interpretation of orthophotos, resulting in a vector file for each of the departments over France with information on forest composition. The BD Foret classes that were associated with pure stands (e.g. pure pine or oak stands), were translated into the genus classes and point locations were subsequently sampled from the polygon layers. The data and further information about the dataset can be found here: https://geoservices.ign.fr/bdforet.

[1] https://www.slu.se/en/Collaborative-Centres-and-Projects/the-swedish-national-forest-inventory/foreststatistics/microdata-for-download/

4.2.1.5 Filtering

The collected genus classes may still contain noisy labels due to remaining localization and annotation errors (e.g. citizen science data collected from GBIF). To reduce these, additional filtering steps based on auxiliary information were applied to the Harmonized Tree Species Occurrence Points for Europe and BD Forest datasets. First, we excluded observations that were not entirely covered or surrounded by woody vegetation (e.g. observations at the edge of a forest, over sparse trees, or that were not covered by woody vegetation at all) using land cover data from the WorldCover v200 product of the year 2021 (Zanaga et al., 2022) and the Copernicus High Resolution forest type product of the year 2018 (EEA, 2020). Observations that were not fully covered by woody vegetation (shrubs or trees) in a 30m-by-30m window were discarded. Finally, we checked if the genus of the observation matched the forest type (coniferous or broadleaved forest) in a surrounding window of 30 by 30m. Observations were only kept if the genus of the observation matched with the forest type of all pixels in the window.

4.2.2 Predictor variables

4.2.2.1 Sentinel-1 and -2 data

Yearly temporal statistics derived from Sentinel-1 and Sentinel-2 data were used as predictor variables in the genus models. The Sentinel-2 features were derived from Sentinel 2 Level 2A products of the year 2020, which were first selected and filtered based on cloud cover. We excluded clouds, cloud shadows, cirrus, and saturated pixels from the Sentinel-2 bands based on the scene classification layer of the L2A product. To further reduce residual noise in the time series, ten days median composites with twenty days moving window were subsequently computed. The remaining missing values were filled using a linear interpolation of neighboring available pixels in the time series. After pre-processing the Sentinel-2 data, a set of vegetation indices were calculated, including the NDVI, EVI, NIRV, NDWI, NDGI, NDMI, NBR, NBR2, REP, ANIR, NDRE2, and NDRE3. Next, the time series of pre-processed Sentinel-2 data and vegetation indices were used to derive a set of descriptive statistics, i.e. features. These include the 10th, 50th, 90th, and interquartile range (difference between 75th and 25th percentile) for each band (B02, B03, B04, B05, B06, B07, B08, B11, and B12) and vegetation index. In addition, NDVI time series that were down sampled to 6 timestamps using Fourier methods were added as feature.

The Sentinel-1 features were derived in a similar fashion to the Sentinel-2 features. The Sentinel-1 Ground Range Detected (GRD) products were first corrected to Gamma0 backscatter and geocoded to the Sentinel-2 grid. Maximum one observation per day and per tile was processed to reduce data redundancy. A multitemporal speckle filter was then applied to reduce noise, after which the time series were composited to ten daily composites. Finally, the percentiles and interquartile range were computed on the Sentinel-1 VV and VH backscatter, the difference between the VH and VV backscatter, and the radar vegetation index (RVI) time series.

4.2.2.2 Localizing features

Given that the dominant genus classes are modeled over Europe, a large variability in growth conditions can be expected. Since tree species may adapt their phenology and morphology to

environmental conditions, the spectral features for the genus classes may vary geographically. To account for this spectral variation, we used information that allowed the model to localize, including information on altitude and climate. Climate information consisted of the minimum (10th percentile), mean (50th percentile), and maximum (90th percentile) temperature, mean rainfall, and mean dew temperature of the year 2020.

4.2.3 Model training

We trained a CatBoost (Dorogush et al., 2018) classification model that predicts the seven genus classes using the Sentinel-1 and Sentinel-2 features, altitude, and climate data as input. First, a training, validation, and test dataset was sampled from the genus information that was extracted from the NFI, harmonized tree species occurrence points, inventory of the private forests in Finland, and BD Forest datasets (Figure 25 and Figure 26). The data were sampled in blocks of 100 by 100 km, where 80% of the blocks were used for training, 10% for validation, and 10% for testing. Since the genus data showed a geographical bias (e.g. a sampling campaign can lead to many samples within one region), potentially leading to an unbalanced model, we sampled a maximum of 250 observations per genus per 100 x 100 km tile. As the inventory data provide information about the dominance of a genus and were considered as the most reliable source of genus information, samples originating from the NFI datasets or the inventory of private forests in Finland were prioritized over the other datasets. We sampled up to 250 observations per genus and per 100x100 km tile from the inventory data. If the total number of observations per genus was still lower than 250, also data from the other sources were sampled. To reduce mixed spectral signals in the training set, only samples were included where the basal area of the dominant genus was larger than 70%.

Besides the genus classes, also samples over areas without woody vegetation are needed to represent the ‘no tree’ class. These locations were sampled based on the intersection of the non-forest areas within the Copernicus High Resolution layer of 2018 and the areas without woody vegetation within the ESA WorldCover land cover map of 2020 and 2021. To ensure that the samples outside woody vegetation represent different land cover types without oversampling dominant classes, the sampling was performed using the land cover classes from the ESA WorldCover land cover map of 2020. We extracted 10% of the pixels per land cover type, with a maximum of 50 samples per land cover type within a 10x10 km area. Finally, 500 samples were maximally selected at random from the sampled locations per 100 x 100 km grid cell.

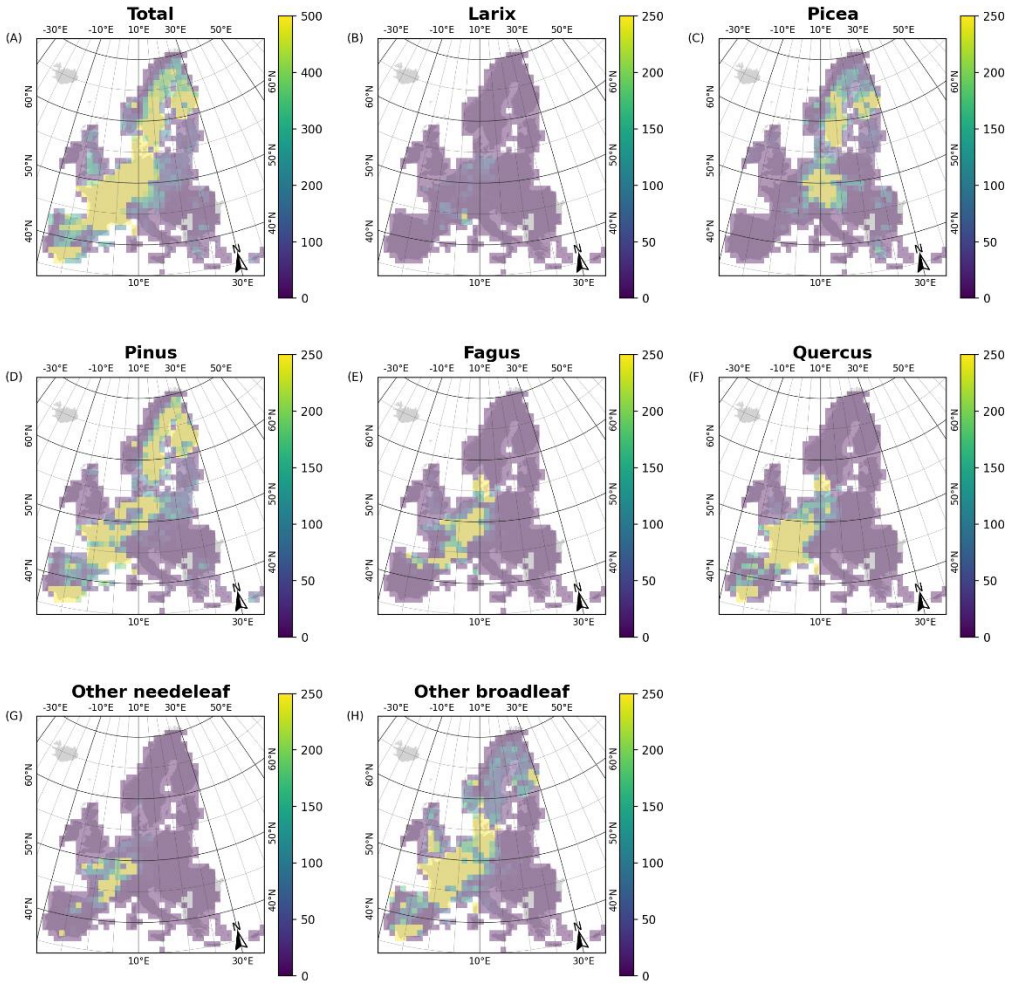

Figure 25: Distribution of samples in the training dataset per genus class: total number of samples (A), number of samples for Larix (B), Picea (C), Pinus (D), Fagus (E), Quercus (F), other needleleaf (G), and other broadleaves (H).

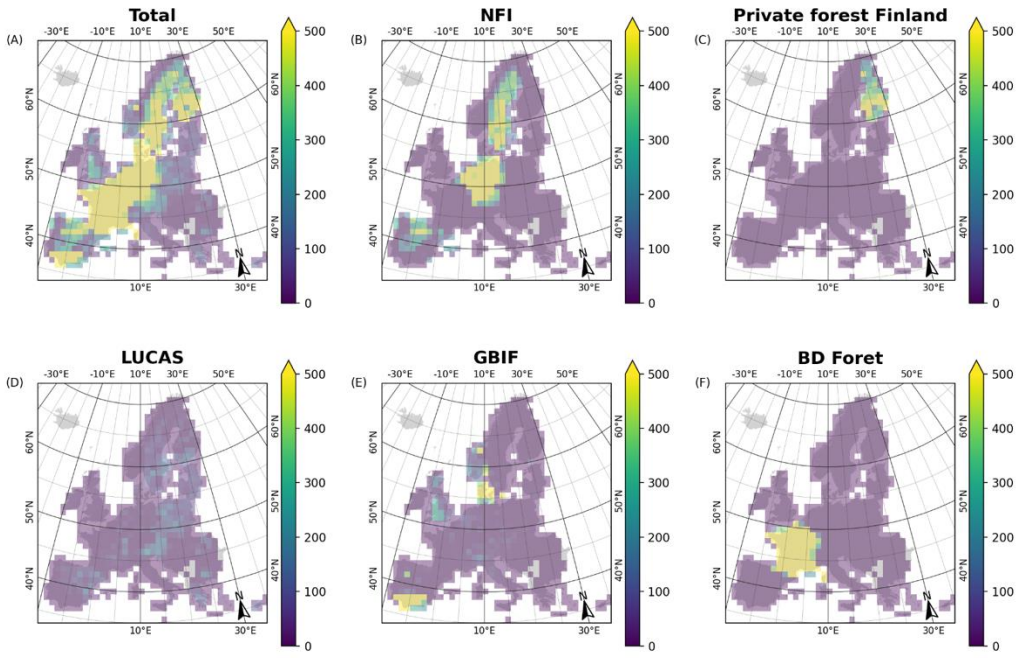

Figure 26: Distribution of the number of samples per data source: total number of samples (A), number of samples originating from NFI data (B), Private forest inventories in Finland (C), LUCAS (D), GBIF (E), and BD Forest (F).

Next, a CatBoost Classification model was trained using the set of Sentinel-1 and Sentinel-2 temporal features of the year 2020, altitude, and climate data as independent variable. Since the number of data samples was not equally distributed over the classes, we assigned a weight to each class that was inversely proportional to the squared root of the number of observations within the class. The model was trained using both the training and validation set, where the validation set was used as overfitting detector, to select the best iteration, and to monitor changes in the model loss.

We used a hold-out of the data (i.e. the test set) to assess the performance of the model. Based on this test set, we computed the overall accuracy, precision, recall, and F1 score per class. The F1 score is a combination of precision and recall, with an equal relative contribution of both metrics. The values range between 0 and 1. High F1 scores indicate a better classification performance than low values.

In addition to the validation, the geographical distribution of the probabilities of the genus classes was compared with the probability maps of Brus et al. (2012) and the relative probability of presence of De Rigo et al. (2016). To compare the maps over Europe, we derived the probabilities at lower resolution. Yet, since the probability of a genus class at low resolution may be affected by the fraction of the pixel that is covered by trees, we computed the relative probability: P_{genus} / (100-$P_{no\ tree}$). In addition, the maps of Brus et al. (2012) and De Rigo et al. (2016) provide probabilities for some of the genus classes at species level. These maps were aggregated per genus class by taking the maximum probability per pixel.

4.3 Results and discussion

4.3.1 Genus map

The prediction of the genus classes over Europe shows that Pinus is mostly dominating in the North (above 50 degrees North) and the Iberian Peninsula (Figure 27). Whereas Pinus can dominate throughout Europe, Quercus and Beech trees are mostly dominating below 55° North. Picea tends to dominate mostly in central Europe and Scandinavia.

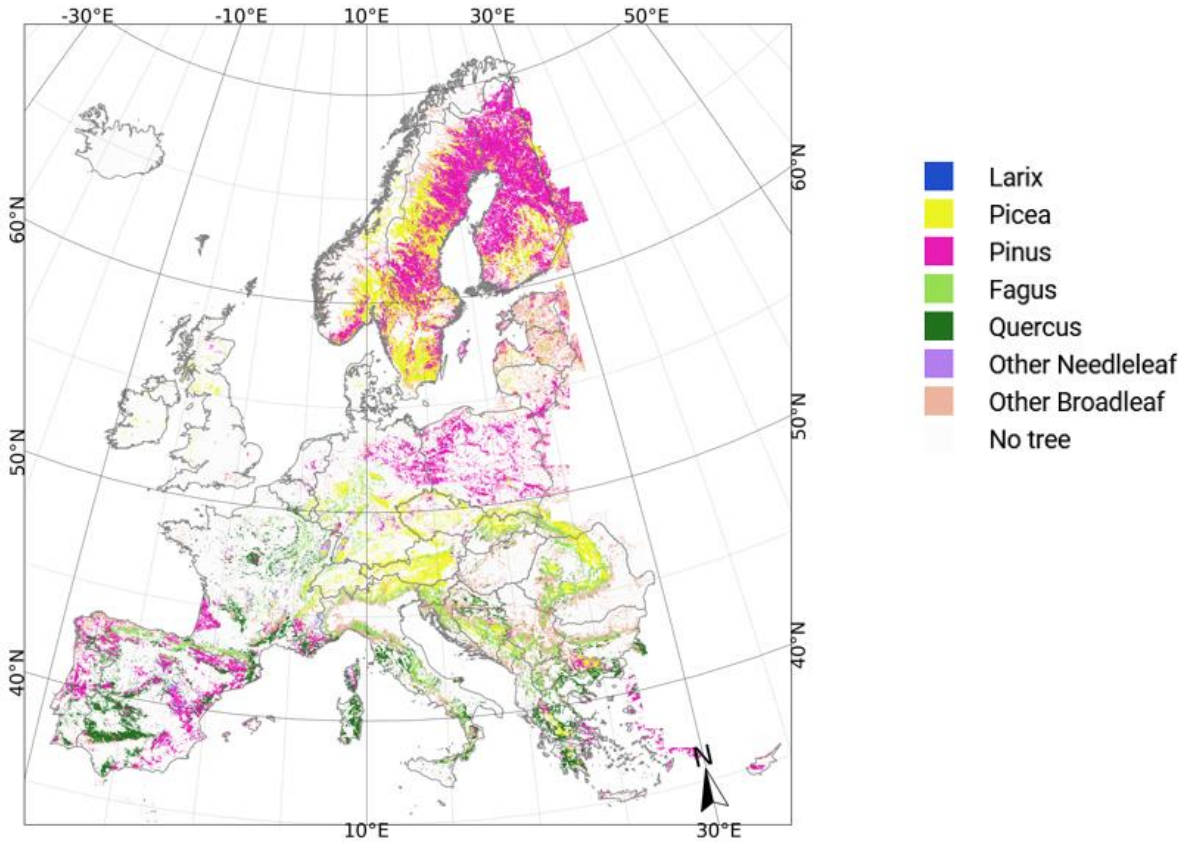

Figure 27: Prediction of genus classes over Europe

Zooming in on the prediction and associated probabilities over a set of patches across Europe (Figure 28) however shows that the probabilities may add important information to the predictions. In some cases (e.g. the Pinus class in Figure 28A) the model predicts the genus classes with high probability. Yet, in other cases the genus classes show rather similar probabilities for different genus classes. In Figure 28C, for instance, there is a rather high probability for both the Pinus and Picea class and both the Quercus and other broadleaf classes. Such overlap in probabilities can be caused by two main reasons. First, the model has difficulties discriminating genus classes in case of an overlapping feature space and is therefore not certain about the predicted class. Second, a pixel may consist of multiple genus classes, causing a mixed spectral signal and therefore also mixed probabilities for the genus classes. Hence, as also discussed and proposed by Hermosilla et al. (2022), looking only at the class with the highest probability does not provide

all the information. In cases where the multiple genus classes reach similar probabilities, it is recommended to account for the probabilities associated with the genus classes to get more detailed information on tree genus assemblages and the model confidence.

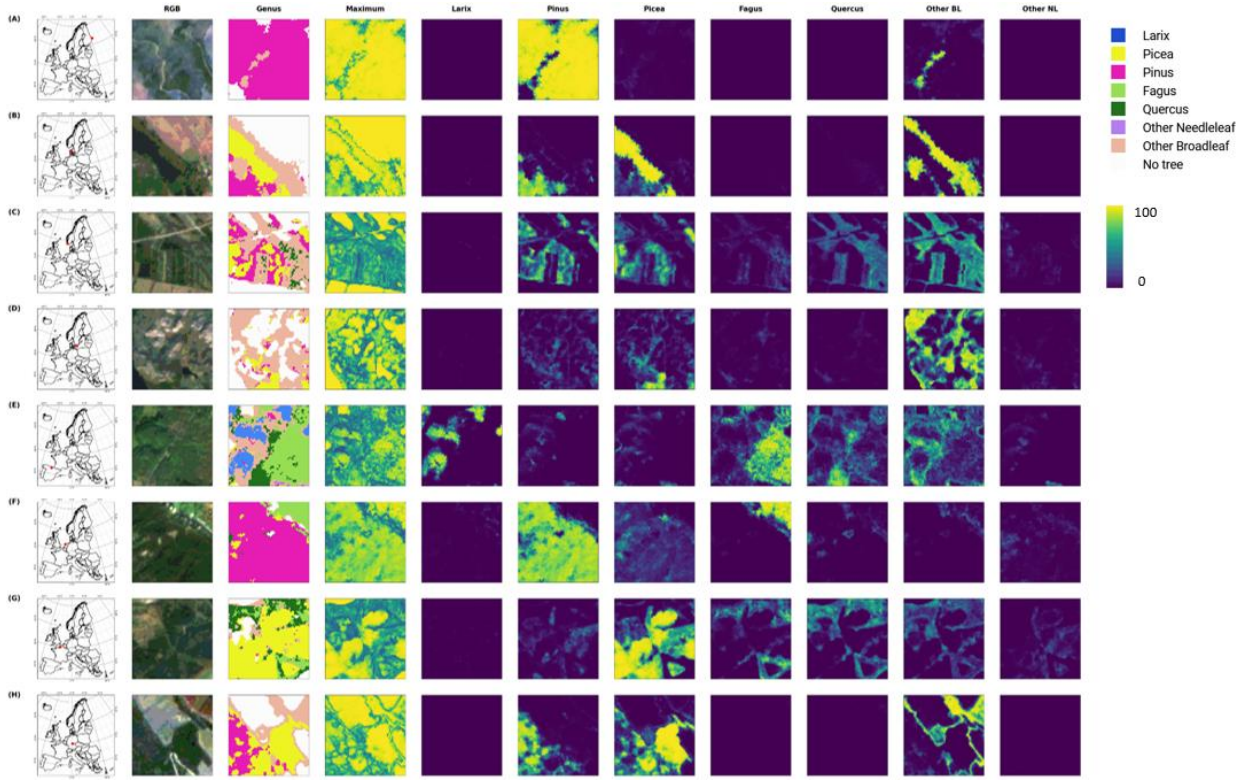

Figure 28: Examples of genus predictions and their associated probabilities. From left to right: location of the example patch, RGB image based on an annual Sentinel-2 composite, predicted genus class, the maximum probability over all classes, the probability of Larix, Picea, Pinus, Fagus, Quercus, other needleleaf, other broadleaf, and the no woody vegetation class.

4.3.2 Literature comparison

Overall, the geographical distribution of the relative probability per genus tends to coincide with the distribution that was mapped using a combination of compositional kriging and a multiple logistic regression model at 1 km resolution over Europe by Brus et al. (2012) and the relative probability of presence mapped at 1 km resolution by De Rigo et al. (2016) (Figure 29). The probability of Quercus is however higher than the other distribution maps. Also, the comparison of our results with tree species maps at regional to national scales (e.g. Wallonia or Germany) shows similar distribution patterns (Bolyn et al., 2022; Welle et al., 2022).

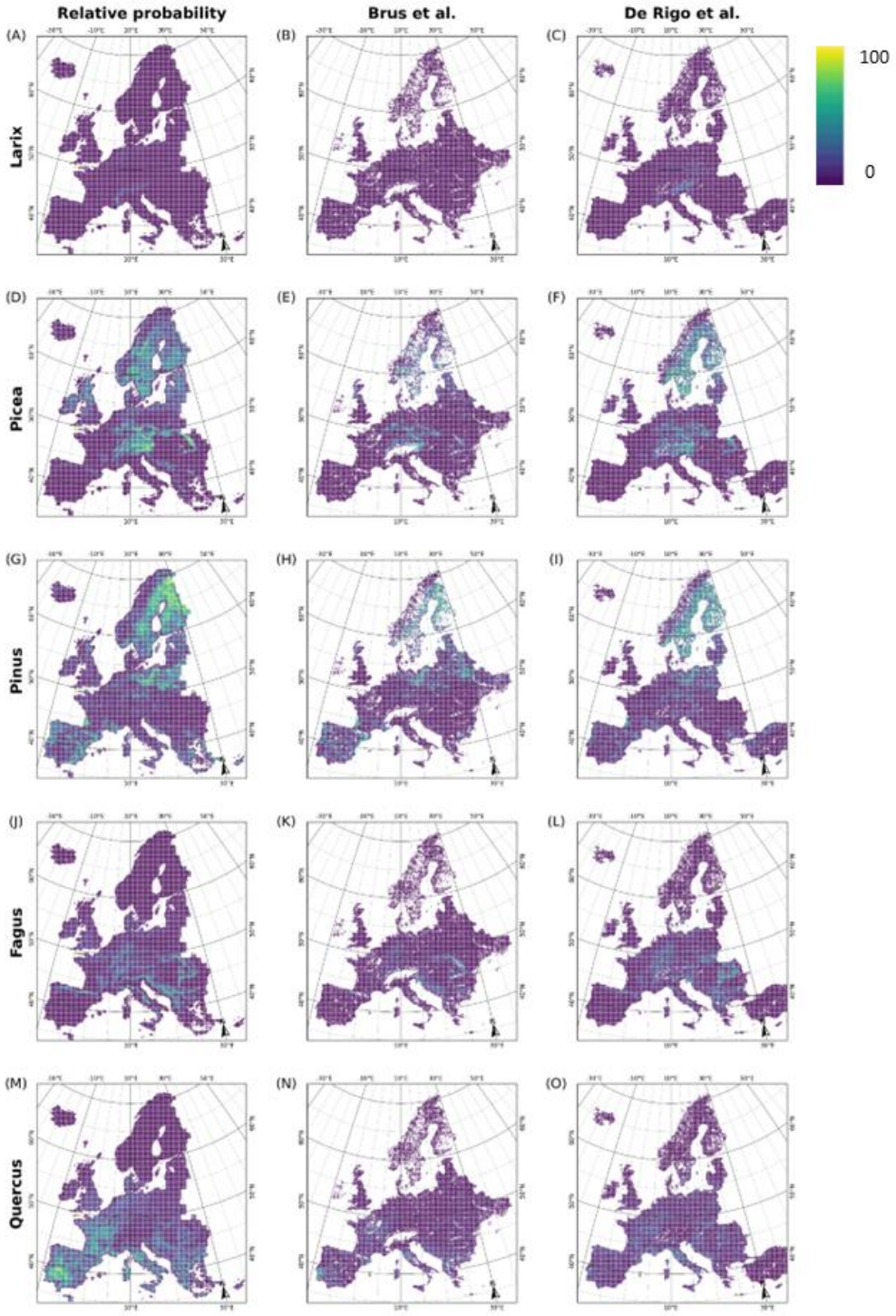

Figure 29: Relative probability maps (A, D, G, J, M), probabilities of Brus et al. (2012) (B, E, H, K, N), and relative probability of presence of De Rigo et al. (2016) (C, F, I, L, O) per genus: Larix (A-C), Picea (D-F), Pinus (G-I), Fagus (J-L), and Quercus (M-O).

4.3.3 Evaluation over the test set

The CatBoost model reached an overall accuracy of 93% on the hold-out set. Most classes reach relatively high F1-scores, where the F1-scores of the needleleaf classes tend to be higher than the broadleaf classes. The former reaches values ranging between 0.78 and 0.89, while the latter ranges between 0.58 and 0.74 (Table 9). The recall, precision, and F1-scores for each of the modelled genus classes show that Quercus was the genus that had the lowest accuracy, followed by Fagus and other broadleaf. Indeed, the model shows some confusion between these classes, where about 30% of the Quercus observations were mapped as other broadleaf (Table 10).

Although single validation statistics have been provided over Europe, it is expected that the classification accuracy may vary spatially due to the availability of training data or changes in environmental characteristics. Given that the training and validation data are not equally distributed over Europe (e.g. few samples could be collected in Balkan countries), it should be noted that the validation mostly represents the areas where data was available.

Finally, the quality of the training data may depend on the data source. In general, plot inventories, such as those available via NFI data, are the most reliable source of species information. Hence, further improvements in accuracy of the genus map could be expected with the increased availability of high-quality genus information over areas that are currently sparsely covered. While inventory data with exact coordinates could be collected (sometimes indirectly) for six countries, expanding these data over all European countries with exact coordinates is challenging (Päivinen et al., 2023).

Table 9: Precision, recall, F1-score and support of the model. The overall accuracy of the model is 93%.

Genus	Precision	Recall	F1-score	Support
Larix	0.88	0.77	0.82	514
Picea	0.82	0.85	0.84	2552
Pinus	0.90	0.89	0.89	4481
Fagus	0.69	0.80	0.74	1587
Quercus	0.66	0.52	0.58	2309
Other needleleaf	0.70	0.87	0.78	411
Other broadleaf	0.74	0.74	0.74	3919
No woody	0.99	1.00	1.00	32814

Table 10: Relative confusion matrix

		Predicted							
		Larix	Picea	Pinus	Fagus	Quercus	Other needleleaf	Other broadleaf	Not woody
Reference	Larix	77.2	7.8	8.9	1	0	0.2	2.9	1.9
	Picea	0.4	84.8	9.3	0.5	0	0.8	2.8	1.4
	Pinus	0.6	6.7	89.4	0.2	0.5	0.7	1.1	0.8
	Fagus	0.4	0.6	0.2	79.6	7.8	0.3	11	0.2
	Quercus	0.1	0.7	2.7	11.8	51.6	2.3	28.8	2
	Other needleleaf	0.2	5.4	6.1	0.2	1	86.6	0.5	0
	Other broadleaf	0.2	1.7	2.1	7.2	11.5	0.9	74.2	2.1
	Not woody	0	0	0	0	0.1	0	0.1	99.7

4.4 Conclusions

The tree genus was modelled and mapped over Europe at 10 m resolution using Sentinel-1 and Sentinel-2 data and localizing features. Eight classes were mapped, including Larix, Picea, Pinus, Fagus, Quercus, other needleleaf, other broadleaf, and no forest.

While the model showed relatively high f1-scores for the genus classes and the overall geographic distribution of the genus classes seems to correspond with maps found in the literature, more pronounced probabilities were found for Quercus. In addition, the model sometimes shows confusion between classes, where class probabilities may give more detailed information. The class confusion may be caused by spectral overlap between classes or pixels that are covered by a mixture of tree genus classes. The Quercus class showed the lowest f1-score, where the model showed confusion with Fagus and other broadleaf.

5 Acknowledgements

We would like to acknowledge and thank the contact persons at the National Forest Inventories of Flanders, Germany, Sweden, the Netherlands, and Spain and at the Finnish Forest Centre and the European Forest Institute for preparing and sharing the tree species information.

6 References

Abdalati, W., Zwally, H. J., Bindschadler, R., Csatho, B., Farrell, S. L., Fricker, H. A., Harding, D., Kwok, R., Lefsky, M., & Markus, T. (2010). The ICESat-2 laser altimetry mission. Proceedings of the IEEE, 98(5), 735-751.

Adam, M., Urbazaev, M., Dubois, C., & Schmullius, C. (2020). Accuracy assessment of GEDI terrain elevation and canopy height estimates in European temperate forests: Influence of environmental and acquisition parameters. Remote Sensing, 12(23), 3948.

Abrams, M., Crippen, R., & Fujisada, H. (2020). ASTER Global Digital Elevation Model (GDEM) and ASTER Global Water Body Dataset (ASTWBD). Remote Sensing, 12(7), 1156. https://doi.org/10.3390/rs12071156

Adam, M., Urbazaev, M., Dubois, C., & Schmullius, C. (2020). Accuracy assessment of GEDI terrain elevation and canopy height estimates in European temperate forests: Influence of environmental and acquisition parameters. Remote Sensing, 12(23), 3948.

Alberdi, I., Bombín, R. V., González, J. G. Á., Ruiz, S. C., Ferreiro, E. G., García, S. G., Mateo, L. H., Jáuregui, M. M., Pita, F. M., & de Oliveira Rodríguez, N. (2017). The multi-objective Spanish national forest inventory. Forest systems, 26(2), 14.

Álvarez-González, J. G., Canellas, I., Alberdi, I., Gadow, K. V., & Ruiz-González, A. (2014). National Forest Inventory and forest observational studies in Spain: Applications to forest modeling. Forest Ecology and Management, 316, 54-64.

ANF. (2023). Bosinventaris. https://www.natuurenbos.be/beleid-wetgeving/natuurbeheer/bosinventaris

Banskota, A., Kayastha, N., Falkowski, M. J., Wulder, M. A., Froese, R. E., & White, J. C. (2014). Forest Monitoring Using Landsat Time Series Data: A Review. Canadian Journal of Remote Sensing, 40(5), 362-384. https://doi.org/10.1080/07038992.2014.987376

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Blickensdörfer, L., Oehmichen, K., Pflugmacher, D., Kleinschmit, B., & Hostert, P. (2024). National tree species mapping using Sentinel-1/2 time series and German National Forest Inventory data. Remote Sensing of Environment, 304, 114069.

Bolyn, C., Lejeune, P., Michez, A., & Latte, N. (2022). Mapping tree species proportions from satellite imagery using spectral–spatial deep learning. Remote Sensing of Environment, 280, 113205.

Bonannella, C., Hengl, T., Heisig, J., Parente, L., Wright, M. N., Herold, M., & De Bruin, S. (2022). Forest tree species distribution for Europe 2000–2020: mapping potential and realized distributions using spatiotemporal machine learning. PeerJ, 10, e13728.

Breidenbach, J., Waser, L. T., Debella-Gilo, M., Schumacher, J., Rahlf, J., Hauglin, M., Puliti, S., & Astrup, R. (2021). National mapping and estimation of forest area by dominant tree species using Sentinel-2 data. Canadian Journal of Forest Research, 51(3), 365-379.

Breidenbach, J., Ellison, D., Petersson, H., Korhonen, K. T., Henttonen, H. M., Wallerman, J., Fridman, J., Gobakken, T., Astrup, R., & Næsset, E. (2022). Harvested area did not increase abruptly—How advancements in satellite-based mapping led to erroneous conclusions. Annals of Forest Science, 79(1), 2. https://doi.org/10.1186/s13595-022-01120-4

Brus, D., Hengeveld, G., Walvoort, D., Goedhart, P., Heidema, A., Nabuurs, G., & Gunia, K. (2012). Statistical mapping of tree species over Europe. European Journal of Forest Research, 131, 145-157.

Buma, B. (2015). Disturbance interactions: Characterization, prediction, and the potential for cascading effects. Ecosphere, 6(4), 1-15. https://doi.org/10.1890/ES15-00058.1

Cardille, J. A., Perez, E., Crowley, M. A., Wulder, M. A., White, J. C., & Hermosilla, T. (2022). Multi-sensor change detection for within-year capture and labelling of forest disturbance. Remote Sensing of Environment, 268, 112741. https://doi.org/10.1016/j.rse.2021.112741

Ceccherini, G., Duveiller, G., Grassi, G., Lemoine, G., Avitabile, V., Pilli, R., & Cescatti, A. (2020). Abrupt increase in harvested forest area over Europe after 2015. Nature, 583(7814), 72-77. https://doi.org/10.1038/s41586-020-2438-y

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321-357. https://doi.org/10.1613/jair.953

Chazdon, R. L., Brancalion, P. H. S., Laestadius, L., Bennett-Curry, A., Buckingham, K., Kumar, C., Moll-Rocek, J., Vieira, I. C. G., & Wilson, S. J. (2016). When is a forest a forest? Forest concepts and definitions in the era of forest and landscape restoration. Ambio, 45(5), 538-550. https://doi.org/10.1007/s13280-016-0772-y

Ciais, P., Schelhaas, M. J., Zaehle, S., Piao, S. L., Cescatti, A., Liski, J., Luyssaert, S., Le-Maire, G., Schulze, E.-D., Bouriaud, O., Freibauer, A., Valentini, R., & Nabuurs, G. J. (2008). Carbon accumulation in European forests. Nature Geoscience, 1(7), 425-429. https://doi.org/10.1038/ngeo233

Cohen, W. B., Yang, Z., & Kennedy, R. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync — Tools for calibration and validation. Remote Sensing of Environment, 114(12), 2911-2924. https://doi.org/10.1016/j.rse.2010.07.010

Cohen, W., Healey, S., Yang, Z., Stehman, S., Brewer, C., Brooks, E., Gorelick, N., Huang, C., Hughes, M., Kennedy, R., Loveland, T., Moisen, G., Schroeder, T., Vogelmann, J., Woodcock, C., Yang, L., & Zhu, Z. (2017). How Similar Are Forest Disturbance Maps Derived from Different Landsat Time Series Algorithms? Forests, 8(4), 98. https://doi.org/10.3390/f8040098

Coops, N. C., Shang, C., Wulder, M. A., White, J. C., & Hermosilla, T. (2020). Change in forest condition: Characterizing non-stand replacing disturbances using time series satellite imagery. Forest Ecology and Management, 474, 118370. https://doi.org/10.1016/j.foreco.2020.118370

Crist, E. P. (1985). A TM Tasseled Cap equivalent transformation for reflectance factor data. Remote Sensing of Environment, 17(3), 301-306. https://doi.org/10.1016/0034-4257(85)90102-6

d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H. I., Joebges, C., Lemoine, G., & Van Der Velde, M. (2020). Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union. Scientific Data, 7(1), 352. https://doi.org/10.1038/s41597-020-00675-z

De Rigo, D., Caudullo, G., Houston Durrant, T., & San-Miguel-Ayanz, J. (2016). The European Atlas of Forest Tree Species: modelling, data and information on forest tree species. European atlas of forest tree species, 40-45.

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., & Luthcke, S. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Science of remote sensing, 1, 100002.

Dutrieux, Loïc., Senf, C., Feret, J.-B., Malenovský, Zbyněk., & Beck, P. S. A. (2023). JRC Expert Meeting on Remote Sensing: Based bark beetle outbreak detection and mapping : held 22 November 2022 online. Publications Office of the European Union.

European Commission, Joint Research Centre, Dutrieux, L., Beck, P., Herold, M., Olofsson, P., & Tsendbazar, N. (2023). JRC expert meeting on accuracy assessment and comparison of forest disturbance products. Publications Office of the European Union. https://doi.org/10.2760/881978

EEA. (2020). Forest Type 2018 (raster 10 m), Europe, 3-yearly, Oct. 2020 https://doi.org/https://doi.org/10.2909/59b0620c-7bb4-4c82-b3ce-f16715573137

FAO. (2020). Global Forest Resources Assessment 2020: Main report.

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

Fassnacht, F. E., White, J. C., Wulder, M. A., & Næsset, E. (2024). Remote sensing in forestry: Current challenges, considerations and directions. Forestry: An International Journal of Forest Research, 97(1), 11-37. https://doi.org/10.1093/forestry/cpad024

Fayad, I., Baghdadi, N., Alcarde Alvares, C., Stape, J. L., Bailly, J. S., Scolforo, H. F., Cegatta, I. R., Zribi, M., & Le Maire, G. (2021). Terrain slope effect on forest height and wood volume estimation from GEDI data. Remote Sensing, 13(11), 2136.

Ferretti, M. (2024). Europe needs a joined-up approach for monitoring and protecting its forests. Nature, 626(8001), 954-954.

Ferretti, M., Gessler, A., Cools, N., Fleck, S., Guerrieri, R., Jakovljević, T., Nicolas, M., Nieminen, T. M., Pitar, D., Potočić, N., Raspe, S., Schaub, M., Schwärzel, K., Timmermann, V., Vejpustková, M., Vesterdal, L., Vanninen, P., Waldner, P., Zimmermann, L., & Sanders, T. G. (2024). Perspectives: Resilient forests need joint forces for better inventorying and monitoring. Forest Ecology and Management, 561, 121875. https://doi.org/10.1016/j.foreco.2024.121875

FOREST EUROPE. (2020). State of Europe's Forests. Ministerial Conference on the Protection of Forests in Europe. https://foresteurope.org/wp-content/uplo ads/2016/08/SoEF_2020.pdf

Forman, G., & Scholz, M. (2010). Apples-to-apples in cross-validation studies: Pitfalls in classifier performance measurement. ACM SIGKDD Explorations Newsletter, 12(1), 49-57. https://doi.org/10.1145/1882471.1882479

Forzieri, G., Dutrieux, L. P., Elia, A., Eckhardt, B., Caudullo, G., Taboada, F. Á., Andriolo, A., Bălăcenoiu, F., Bastos, A., Buzatu, A., Dorado, F. C., Dobrovolný, L., Duduman, M., Fernandez-Carrillo, A., Hernández-Clemente, R., Hornero, A., Ionuț, S., Lombardero, M. J., Junttila, S., …

Beck, P. S. A. (2023). The Database of European Forest Insect and Disease Disturbances: DEFID2. Global Change Biology, 29(21), 6040-6065. https://doi.org/10.1111/gcb.16912

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., … Beck, P. S. A. (2020). A spatially explicit database of wind disturbances in European forests over the period 2000–2018. Earth System Science Data, 12(1), 257-276. https://doi.org/10.5194/essd-12-257-2020

Francini, S., Hermosilla, T., Coops, N. C., Wulder, M. A., White, J. C., & Chirici, G. (2023). An assessment approach for pixel-based image composites. ISPRS Journal of Photogrammetry and Remote Sensing, 202, 1-12. https://doi.org/10.1016/j.isprsjprs.2023.06.002

Francini, S., McRoberts, R. E., D'Amico, G., Coops, N. C., Hermosilla, T., White, J. C., Wulder, M. A., Marchetti, M., Mugnozza, G. S., & Chirici, G. (2022). An open science and open data approach for the statistically robust estimation of forest disturbance areas. International Journal of Applied Earth Observation and Geoinformation, 106, 102663. https://doi.org/10.1016/j.jag.2021.102663

Francini, S., McRoberts, R. E., Giannetti, F., Marchetti, M., Scarascia Mugnozza, G., & Chirici, G. (2021). The Three Indices Three Dimensions (3I3D) algorithm: A new method for forest disturbance mapping and area estimation based on optical remotely sensed imagery. International Journal of Remote Sensing, 42(12), 4693-4711. https://doi.org/10.1080/01431161.2021.1899334

Francini, S., Schelhaas, M.-J., Vangi, E., Lerink, B.-J., Nabuurs, G.-J., McRoberts, R. E., & Chirici, G. (2024). Forest species mapping and area proportion estimation combining Sentinel-2 harmonic predictors and national forest inventory data. International Journal of Applied Earth Observation and Geoinformation, 131, 103935.

Frantz, D. (2019). FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. Remote Sensing, 11(9), 1124. https://doi.org/10.3390/rs11091124

Frantz, D., Roder, A., Stellmes, M., & Hill, J. (2016). An Operational Radiometric Landsat Preprocessing Framework for Large-Area Time Series Applications. IEEE Transactions on Geoscience and Remote Sensing, 54(7), 3928-3943. https://doi.org/10.1109/TGRS.2016.2530856

Fridman, J., Holm, S., Nilsson, M., Nilsson, P., Ringvall, A. H., & Ståhl, G. (2014). Adapting National Forest Inventories to changing requirements–the case of the Swedish National Forest Inventory at the turn of the 20th century. Silva Fennica, 48(3).

García, M. L., & Caselles, V. (1991). Mapping burns and natural reforestation using Thematic Mapper data. Geocarto International, 6(1), 31-37.

Gómez, C. (s. f.). 4 Remote sensing for the Spanish forests in the 21st century: A review of 5 advances, needs, and opportunities. 6 Authors and affiliations.

Govaere, L. (2019). Protocol En Handleiding Derde Bosinventarisatie Vlaams Gewest. Brussels, Belgium

Govaere, L., Van de Kerckhove, P., Roelandt, B., Sannen, P., & Schrey, L. (2009). Handleiding Tweede Bosinventarisatie Vlaams Gewest. Werkdocument ANB.

Griffiths, P., Van Der Linden, S., Kuemmerle, T., & Hostert, P. (2013). A Pixel-Based Landsat Compositing Algorithm for Large Area Land Cover Mapping. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6(5), 2088-2101. https://doi.org/10.1109/JSTARS.2012.2228167

Grünig, M., Seidl, R., & Senf, C. (2022). Increasing aridity causes larger and more severe forest fires across Europe. Global Change Biology, 29(6), 1648-1659. https://doi.org/10.1111/gcb.16547

Hammond, W. M., Williams, A. P., Abatzoglou, J. T., Adams, H. D., Klein, T., López, R., Sáenz-Romero, C., Hartmann, H., Breshears, D. D., & Allen, C. D. (2022). Global field observations of tree die-off reveal hotter-drought fingerprint for Earth's forests. Nature Communications, 13(1), 1761. https://doi.org/10.1038/s41467-022-29289-2

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. Science, 342(6160), 850-853. https://doi.org/10.1126/science.1244693

Healey, S., Cohen, W., Zhiqiang, Y., & Krankina, O. (2005). Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. Remote Sensing of Environment, 97(3), 301-310. https://doi.org/10.1016/j.rse.2005.05.009

Heisig, J., & Hengl, T. (2020). Harmonized Tree Species Occurrence Points for Europe (0.2). https://doi.org/https://doi.org/10.5281/zenodo.5524611

Hermosilla, T., Wulder, M. A., White, J. C., & Coops, N. C. (2019). Prevalence of multiple forest disturbances and impact on vegetation regrowth from interannual Landsat time series (1985–2015). Remote Sensing of Environment, 233, 111403. https://doi.org/10.1016/j.rse.2019.111403
Hermosilla, T., Wulder, M. A., White, J. C., & Coops, N. C. (2022a). Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes. Remote Sensing of Environment, 268, 112780. https://doi.org/10.1016/j.rse.2021.112780

Hermosilla, T., Bastyr, A., Coops, N. C., White, J. C., & Wulder, M. A. (2022b). Mapping the presence and distribution of tree species in Canada's forested ecosystems. Remote Sensing of Environment, 282, 113276.

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., & Hobart, G. W. (2015). Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using

Landsat-derived time-series metrics. Remote Sensing of Environment, 170, 121-132. https://doi.org/10.1016/j.rse.2015.09.004

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., & Hobart, G. W. (2017). Updating Landsat time series of surface-reflectance composites and forest change products with new observations. International Journal of Applied Earth Observation and Geoinformation, 63, 104-111. https://doi.org/10.1016/j.jag.2017.07.013

Hirschmugl, M., Gallaun, H., Dees, M., Datta, P., Deutscher, J., Koutsias, N., & Schardt, M. (2017). Methods for Mapping Forest Disturbance and Degradation from Optical Earth Observation Data: A Review. Current Forestry Reports, 3(1), 32-45. https://doi.org/10.1007/s40725-017-0047-2

Huang, C., Goward, S. N., Masek, J. G., Thomas, N., Zhu, Z., & Vogelmann, J. E. (2010). An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. Remote Sensing of Environment, 114(1), 183-198. https://doi.org/10.1016/j.rse.2009.08.017

Hughes, M., Kaylor, S., & Hayes, D. (2017). Patch-Based Forest Change Detection from Landsat Time Series. Forests, 8(5), 166. https://doi.org/10.3390/f8050166

IGN. (2016). BD Forêt Version 2.0. January 2016

Kacic, P., Thonfeld, F., Gessner, U., & Kuenzer, C. (2023). Forest structure characterization in Germany: novel products and analysis based on GEDI, sentinel-1 and sentinel-2 data. Remote Sensing, 15(8), 1969.

Kennedy, R. E., Cohen, W. B., & Schroeder, T. A. (2007). Trajectory-based change detection for automated characterization of forest disturbance dynamics. Remote Sensing of Environment, 110(3), 370-386. https://doi.org/10.1016/j.rse.2007.03.010

Kennedy, R. E., Yang, Z., & Cohen, W. B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. Remote Sensing of Environment, 114(12), 2897-2910. https://doi.org/10.1016/j.rse.2010.07.008

Lang, N., Jetz, W., Schindler, K., & Wegner, J. D. (2023). A high-resolution canopy height model of the Earth. Nature Ecology & Evolution, 7(11), 1778-1789.

Lecina-Diaz, J., Senf, C., Grünig, M., & Seidl, R. (2024). Ecosystem services at risk from disturbance in Europe's forests. Global Change Biology, 30(3), e17242. https://doi.org/10.1111/gcb.17242

Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., Seidl, R., Delzon, S., Corona, P., Kolström, M., Lexer, M. J., & Marchetti, M. (2010). Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. Forest Ecology and Management, 259(4), 698-709. https://doi.org/10.1016/j.foreco.2009.09.023

Liu, S., Brandt, M., Nord-Larsen, T., Chave, J., Reiner, F., Lang, N., Tong, X., Ciais, P., Igel, C., Pascual, A., Guerra-Hernandez, J., Li, S., Mugabowindekwe, M., Saatchi, S., Yue, Y., Chen, Z., & Fensholt, R. (2023). The overlooked contribution of trees outside forests to tree cover and woody biomass across Europe. SCIENCE ADVANCES.

Mandl, L., Viana-Soto, A., Seidl, R., Stritih, A., & Senf, C. (2024). Unmixing-based forest recovery indicators for predicting long-term recovery success. Remote Sensing of Environment, 308, 114194. https://doi.org/10.1016/j.rse.2024.114194

Maroschek, M., Seidl, R., Poschlod, B., & Senf, C. (2024). Quantifying patch size distributions of forest disturbances in protected areas across the European Alps. Journal of Biogeography, 51(3), 368-381. https://doi.org/10.1111/jbi.14760

Matasci, G., Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., & Zald, H. S. (2018). Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. Remote Sensing of Environment, 209, 90-106.

Mauri, A., Strona, G., & San-Miguel-Ayanz, J. (2017). EU-Forest, a high-resolution tree occurrence dataset for Europe. Scientific data, 4(1), 1-8.

McDowell, N. G., Allen, C. D., Anderson-Teixeira, K., Aukema, B. H., Bond-Lamberty, B., Chini, L., Clark, J. S., Dietze, M., Grossiord, C., Hanbury-Brown, A., Hurtt, G. C., Jackson, R. B., Johnson, D. J., Kueppers, L., Lichstein, J. W., Ogle, K., Poulter, B., Pugh, T. A. M., Seidl, R., … Xu, C. (2020). Pervasive shifts in forest dynamics in a changing world. Science, 368(6494), eaaz9463. https://doi.org/10.1126/science.aaz9463

Miguel, S., Ruiz-Benito, P., Rebollo, P., Viana-Soto, A., Mihai, M. C., García-Martín, A., & Tanase, M. (2024). Forest disturbance regimes and trends in continental Spain (1985- 2023) using dense Landsat time series. Environmental Research, 119802. https://doi.org/10.1016/j.envres.2024.119802

Moisen, G. G., Meyer, M. C., Schroeder, T. A., Liao, X., Schleeweis, K. G., Freeman, E. A., & Toney, C. (2016). Shape selection in Landsat time series: A tool for monitoring forest dynamics. Global Change Biology, 22(10), 3518-3528. https://doi.org/10.1111/gcb.13358

Morresi, D., Maeng, H., Marzano, R., Lingua, E., Motta, R., & Garbarino, M. (2024). High-Dimensional Detection of Landscape Dynamics: A Landsat Time Series-Based Algorithm for Forest Disturbance Mapping and Beyond [Preprint]. SSRN. https://doi.org/10.2139/ssrn.4703613

Nabuurs, G.-J., Harris, N., Sheil, D., Palahi, M., Chirici, G., Boissière, M., Fay, C., Reiche, J., & Valbuena, R. (2022). Glasgow forest declaration needs new modes of data ownership. Nature Climate Change, 12(5), 415-417. https://doi.org/10.1038/s41558-022-01343-3

Neuenschwander, A., Pitts, K., Jelley, B., Robbins, J., Markel, J., Popescu, S., Nelson, R., Harding, D., Pederson, D., & Klotz, B. (2019). Ice, Cloud, and Land Elevation Satellite 2

(ICESat-2) algorithm theoretical basis document (ATBD) for land-vegetation along-track products (ATL08). National Aeronautics and Space Administration: Washington, DC, USA.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment, 148, 42-57. https://doi.org/10.1016/j.rse.2014.02.015

Orsi, F., Ciolli, M., Primmer, E., Varumo, L., & Geneletti, D. (2020). Mapping hotspots and bundles of forest ecosystem services across the European Union. Land Use Policy, 99, 104840. https://doi.org/10.1016/j.landusepol.2020.104840

Palahí, M., Valbuena, R., Senf, C., Acil, N., Pugh, T. A. M., Sadler, J., Seidl, R., Potapov, P., Gardiner, B., Hetemäki, L., Chirici, G., Francini, S., Hlásny, T., Lerink, B. J. W., Olsson, H., González Olabarria, J. R., Ascoli, D., Asikainen, A., Bauhus, J., … Nabuurs, G.-J. (2021). Concerns about reported harvests in European forests. Nature, 592(7856), E15-E17. https://doi.org/10.1038/s41586-021-03292-x

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., Nagel, T. A., Nikinmaa, L., Panyatov, M., Bercak, R., Seidl, R., Ostrogović Sever, M. Z., … Schelhaas, M. (2023). Significant increase in natural disturbance impacts on European forests since 1950. Global Change Biology, 29(5), 1359-1376. https://doi.org/10.1111/gcb.16531

Pauls, J., Zimmer, M., Kelly, U. M., Schwartz, M., Saatchi, S., Ciais, P., Pokutta, S., Brandt, M., & Gieseke, F. (2024). Estimating Canopy Height at Scale. arXiv preprint arXiv:2406.01076.

Päivinen, R., Astrup, R., Birdsey, R. A., Breidenbach, J., Fridman, J., Kangas, A., Kauppi, P. E., Köhl, M., Korhonen, K. T., & Johannsen, V. K. (2023). Ensure forest-data integrity for climate change studies. Nature Climate Change, 13(6), 495-496.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

Pflugmacher, D., Rabe, A., Peters, M., & Hostert, P. (2019). Mapping pan-European land cover using Landsat spectral-temporal metrics and the European LUCAS survey. Remote Sensing of Environment, 221, 583-595. https://doi.org/10.1016/j.rse.2018.12.001

Pucher, C., Neumann, M., & Hasenauer, H. (2022). An Improved Forest Structure Data Set for Europe. Remote Sensing, 14(2), 395.

Pugh, T. A. M., Arneth, A., Kautz, M., Poulter, B., & Smith, B. (2019). Important role of forest disturbances in the global biomass turnover and carbon sinks. Nature Geoscience, 12(9), 730-735. https://doi.org/10.1038/s41561-019-0427-2

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil, 7(1), 217-240.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., & Silva, C. E. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sensing of Environment, 253, 112165.

Riedel, T. (2017). Die dritte Bundeswaldinventur: BWI 2012; Inventur-und Auswertungsmethoden.

Reinosch, E., Backa, J., Adler, P., Deutscher, J., Eisnecker, P., Hoffmann, K., Langner, N., Puhm, M., Rüetschi, M., Straub, C., Waser, L. T., Wiesehahn, J., & Oehmichen, K. (2024). Detailed validation of large-scale Sentinel-2-based forest disturbance maps across Germany. Forestry: An International Journal of Forest Research, cpae038. https://doi.org/10.1093/forestry/cpae038

Roy, D. P., Kovalskyy, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., & Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. Remote Sensing of Environment, 185, 57-70. https://doi.org/10.1016/j.rse.2015.12.024

Ruiz-Benito, P., Vacchiano, G., Lines, E. R., Reyer, C. P., Ratcliffe, S., Morin, X., Hartig, F., Mäkelä, A., Yousefpour, R., & Chaves, J. E. (2020). Available and missing data to model impact of climate change on European forests. Ecological Modelling, 416, 108870.

Saarikoski, H., Jax, K., Harrison, P. A., Primmer, E., Barton, D. N., Mononen, L., Vihervaara, P., & Furman, E. (2015). Exploring operational ecosystem service definitions: The case of boreal forests. Ecosystem Services, 14, 144-157. https://doi.org/10.1016/j.ecoser.2015.03.006

Sabatini, F. M., Burrascano, S., Keeton, W. S., Levers, C., Lindner, M., Pötzschner, F., Verkerk, P. J., Bauhus, J., Buchwald, E., Chaskovsky, O., Debaive, N., Horváth, F., Garbarino, M., Grigoriadis, N., Lombardi, F., Marques Duarte, I., Meyer, P., Midteng, R., Mikac, S., … Kuemmerle, T. (2018). Where are Europe's last primary forests? Diversity and Distributions, 24(10), 1426-1439. https://doi.org/10.1111/ddi.12778

San-Miguel-Ayanz, J., De Rigo, D., Caudullo, G., Houston Durrant, T., & Mauri, A. (2016). European Atlas of forest tree species.− Publication Office of the European Union. European Commission, Luxembourg. doi, 10, 038466.

Schelhaas, M.-J., Clerkx, S., & Lerink, B. (2022). Zevende Nederlandse Bosinventarisatie: 2017-2021 (2667-1263).

Sebald, J., Senf, C., & Seidl, R. (2021). Human or natural? Landscape context improves the attribution of forest disturbances mapped from Landsat in Central Europe. Remote Sensing of Environment, 262, 112502. https://doi.org/10.1016/j.rse.2021.112502

Seidl, R. (2014). The Shape of Ecosystem Management to Come: Anticipating Risks and Fostering Resilience. BioScience, 64(12), 1159-1169. https://doi.org/10.1093/biosci/biu172

Seidl, R., & Senf, C. (2024). Changes in planned and unplanned canopy openings are linked in Europe's forests. Nature Communications, 15(1), 4741. https://doi.org/10.1038/s41467-024-49116-0

Senf, C. (2022). Seeing the System from Above: The Use and Potential of Remote Sensing for Studying Ecosystem Dynamics. Ecosystems, 25(8), 1719-1737. https://doi.org/10.1007/s10021-022-00777-2

Senf, C., Pflugmacher, D., Zhiqiang, Y., Sebald, J., Knorn, J., Neumann, M., Hostert, P., & Seidl, R. (2018). Canopy mortality has doubled in Europe's temperate forests over the last three decades. Nature Communications, 9(1), 4978. https://doi.org/10.1038/s41467-018-07539-6

Senf, C., Sebald, J., & Seidl, R. (2021). Increasing canopy mortality affects the future demographic structure of Europe's forests. One Earth, 4(5), 749-755. https://doi.org/10.1016/j.oneear.2021.04.008

Senf, C., & Seidl, R. (2021a). Mapping the forest disturbance regimes of Europe. Nature Sustainability, 4(1), 63-70. https://doi.org/10.1038/s41893-020-00609-y

Senf, C., & Seidl, R. (2021b). Storm and fire disturbances in Europe: Distribution and trends. Global Change Biology, 27(15), 3605-3619. https://doi.org/10.1111/gcb.15679

Stahl, A. T., Andrus, R., Hicke, J. A., Hudak, A. T., Bright, B. C., & Meddens, A. J. H. (2023). Automated attribution of forest disturbance types from remote sensing data: A synthesis. Remote Sensing of Environment, 285, 113416. https://doi.org/10.1016/j.rse.2022.113416

Stritih, A., Senf, C., Seidl, R., Grêt-Regamey, A., & Bebi, P. (2021). The impact of land-use legacies and recent management on natural disturbance susceptibility in mountain forests. Forest Ecology and Management, 484, 118950. https://doi.org/10.1016/j.foreco.2021.118950

Thom, D., & Seidl, R. (2016). Natural disturbance impacts on ecosystem services and biodiversity in temperate and boreal forests. Biological Reviews, 91(3), 760-781. https://doi.org/10.1111/brv.12193

Tolan, J., Yang, H.-I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., Spore, J., Majumdar, S., Haziza, D., & Vamaraju, J. (2024). Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. Remote Sensing of Environment, 300, 113888.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. Remote sensing of Environment, 8(2), 127-150.

Turner, M. G., & Seidl, R. (2023). Novel Disturbance Regimes and Ecological Responses. Annual Review of Ecology, Evolution, and Systematics, 54(1), 63-83. https://doi.org/10.1146/annurev-ecolsys-110421-101120

Turubanova, S., Potapov, P., Hansen, M. C., Li, X., Tyukavina, A., Pickens, A. H., Hernandez-Serna, A., Arranz, A. P., Guerra-Hernandez, J., Senf, C., Häme, T., Valbuena, R., Eklundh, L., Brovkina, O., Navrátilová, B., Novotný, J., Harris, N., & Stolle, F. (2023). Tree canopy extent and height change in Europe, 2001–2021, quantified using Landsat data archive. Remote Sensing of Environment, 298, 113797. https://doi.org/10.1016/j.rse.2023.113797

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. Remote Sensing of Environment, 114(1), 106-115. https://doi.org/10.1016/j.rse.2009.08.014

Viana-Soto, A., & Senf, C. (2024). European Forest Disturbance Atlas (2.1.1.) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.13333034

Vihervaara, P., Auvinen, A.-P., Mononen, L., Törmä, M., Ahlroth, P., Anttila, S., Böttcher, K., Forsius, M., Heino, J., & Heliölä, J. (2017). How essential biodiversity variables and remote sensing can help national biodiversity monitoring. Global Ecology and Conservation, 10, 43-59.
Welle, T., Aschenbrenner, L., Kuonath, K., Kirmaier, S., & Franke, J. (2022). Mapping dominant tree species of German forests. Remote Sensing, 14(14), 3330.

Wulder, M. A., Hermosilla, T., White, J. C., Bater, C. W., Hobart, G., & Bronson, S. C. (2024). Development and implementation of a stand-level satellite-based forest inventory for Canada. Forestry: An International Journal of Forest Research, cpad065. https://doi.org/10.1093/forestry/cpad065

Wulder, M. A., Roy, D. P., Radeloff, V. C., Loveland, T. R., Anderson, M. C., Johnson, D. M., Healey, S., Zhu, Z., Scambos, T. A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C. J., Masek, J. G., Hermosilla, T., White, J. C., Belward, A. S., Schaaf, C., Woodcock, C. E., … Cook, B. D. (2022). Fifty years of Landsat science and impacts. Remote Sensing of Environment, 280, 113195. https://doi.org/10.1016/j.rse.2022.113195

G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., & Arino, O. (2022). ESA WorldCover 10 m 2021 v200 (Versie v200). https://doi.org/https://doi.org/10.5281/zenodo.7254221

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.-E., . . . Arino, O. (2021). ESA WorldCover 10 m 2020 v100 (Versie v100). https://doi.org/10.5281/zenodo.5571936

Zhao, K., Wulder, M. A., Hu, T., Bright, R., Wu, Q., Qin, H., Li, Y., Toman, E., Mallick, B., Zhang, X., & Brown, M. (2019). Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. Remote Sensing of Environment, 232, 111181. https://doi.org/10.1016/j.rse.2019.04.034

Zhu, Z. (2017). Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. ISPRS Journal of Photogrammetry and Remote Sensing, 130, 370-384. https://doi.org/10.1016/j.isprsjprs.2017.06.013

Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sensing of Environment, 118, 83-94. https://doi.org/10.1016/j.rse.2011.10.028
Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. Remote Sensing of Environment, 144, 152-171. https://doi.org/10.1016/j.rse.2014.01.011