OXFORD

Institute of Chartered Foresters

Forestry: An International Journal of Forest Research, 2024, 1–17

https://doi.org/10.1093/forestry/cpae038

Original Article

Detailed validation of large-scale Sentinel-2-based forest disturbance maps across Germany

Eike Reinosch[1,*], Julian Backa[2], Petra Adler[3], Janik Deutscher[4], Philipp Eisnecker[3], Karina Hoffmann[2], Niklas Langner[5], Martin Puhm[4], Marius Rüetschi[6], Christoph Straub[1], Lars T. Waser[6], Jens Wiesehahn[7], Katja Oehmichen[5]

[1]Department of Information Technology, Bavarian State Institute of Forestry (LWF), Hans Carl-von-Carlowitz-Platz 1, Freising, Bavaria 85354, Germany
[2]Forest GIS, Mapping, Surveying, Competence Centre for Wood and Forestry, Public Enterprise Sachsenforst, Bonnewitzer-Straße 34, OT Graupa, Pirna, Saxony 01796, Germany
[3]Biometrics and Information Technology, Forest Research Institute Baden-Württemberg, Wonnhaldestraße 4, Freiburg, Baden-Württemberg 79100, Germany
[4]Remote Sensing and Geoinformation Joanneum Research, Steyrergasse 17 Graz, Styria 8010, Austria
[5]Forest Resources and Climate Protection, Thünen Institute of Forest Ecosystems, Alfred-Möller-Straße 1, Eberswalde, Brandenburg 16225, Germany
[6]Swiss Federal Institute for Forest, Snow, and Landscape Research WSL, Zürcherstrasse 111, Birmensdorf, Zurich 8903, Switzerland
[7]Remote Sensing and GIS, Northwest German Forest Research Institute, Graetzelstrasse 2, Göttingen, Lower Saxony 37079, Germany

*Corresponding author. Department of Information Technology, Bavarian State Institute of Forestry (LWF), Freising, Bavaria 85354, Germany.
E-mail: ereinosch@web.de

Abstract

Monitoring forest areas with satellite data has become a vital tool to derive information on disturbances in European forests at large scales. An extensive validation of generated maps is essential to evaluate their potential and limitations in detecting various disturbance patterns. Here, we present the validation results of forest disturbance maps generated for four study areas in Germany using Sentinel-2 data from 2018 to 2022. We apply a time series filtering method to map annual forest disturbances larger than 0.1 ha based on spectral clustering and annual change magnitude. The presented method is part of a research study to design a precursor for a national German forest disturbance monitoring system. In this context, annual forest change areas are used to estimate affected timber volume and related economic losses. To better understand the thematic accuracies and the reliability of the area estimates, we performed an independent and extensive validation of the annual product using 20 validation sets embedded in our four study areas and comprising a total of 11 019 sample points. The collected reference datasets are based on an expert interpretation of high-resolution aerial and satellite imagery, including information on the dominant tree species, disturbance cause, and disturbance severity level. Our forest disturbance map achieves an overall accuracy of 99.1 ± 0.1% in separating disturbed from undisturbed forest. This is mainly indicative of the accuracy for undisturbed forest, as that class covers 97.2% of the total forest area. For the disturbed forest class, the user's accuracy is 84.4 ± 2.0% and producer's accuracy is 85.1 ± 3.4% for 2018 to 2022. The similar user's and producer's accuracies indicate that the total disturbance area is estimated accurately. However, for 2022, we observe an overestimation of the total disturbance extent, which we attribute to the high drought stress in that year leading to false detections, especially around forest edges. The accuracy varies widely among validation sets and seems related to the disturbance cause, the disturbance severity, and the disturbance patch size. User's accuracies range from 31.0 ± 8.4% to 98.8 ± 1.3%, while producer's accuracies range from 60.5 ± 37.3% to 100.0 ± 0.0% across the validation sets. These variations highlight that the accuracy of a single local validation set is not representative of a region with a large diversity of disturbance patterns, such as Germany. This emphasizes the need to assess the accuracies of large-scale disturbance products in as many different study areas as possible, to cover different patch sizes, disturbance severities, and disturbance causes.

Introduction

Forests provide a myriad of ecological, economic, and climatic benefits, all of which are influenced by both natural and climate-driven disturbances. Natural disturbances, such as wild fires, insect infestations, and windthrows, are an integral part of ecosystem dynamics in forests around the globe (Turner, 2010). Climate-driven changes in disturbance regimes and a decreasing adaptive capacity of forests, both occurring at an accelerating rate, have been observed worldwide (Hoegh-Guldberg *et al.*, 2018, Senf and Seidl, 2021a). According to Patacca *et al.* (2022), a substantial increase in the occurrence and severity of disturbances has been observed in European forest ecosystems since the 1950s. Collecting reliable and ongoing information on forest disturbances is key for future ecological pathways of forest management and increased forest resilience (Senf and Seidl, 2021a).

Ground-based data are considered the gold standard for detecting forest disturbances, but its availability is often temporally and spatially limited (Bowman *et al.*, 2013). Hence, remote sensing data and technologies have been widely used to detect forest disturbances in the last decades across different scales at high spatial and temporal resolutions (Stahl *et al.*, 2023). Reliable and operational methods for mapping forest disturbances over large areas have lately become increasingly important for sustainable forest management (Hirschmugl *et al.*, 2017). Nowadays, besides Landsat, an increasing number of new satellite constellations, such as Sentinel-1 and -2, PlanetScope, and WorldView-3, meet the necessary requirements by providing higher spatial, temporal, and spectral resolutions compared with those of sensors from earlier decades (Ustin and Middleton, 2021). Moreover, improvements in computing performance and machine learning algorithms have

Handling editor: Prof. Teja Kattenborn
Received 23 August 2023. **Revised** 17 May 2024. **Accepted** 28 June 2024

facilitated advancements in the process of detecting forest disturbances (Zhu et al., 2020). Recently, the open-access Copernicus Sentinel data have been used extensively to detect and monitor forest disturbances in Europe, e.g. for multiple disturbance agents (Francini et al., 2022; Candotti et al., 2022; Senf and Seidl 2021b, Schiefer et al., 2023), windthrows (Rüetschi et al., 2019; Puhm et al., 2020; Laurin et al., 2021; Giannetti et al., 2021), and bark beetle infestations (Fernandez-Carrillo et al., 2020; Bárta et al., 2021; Dalponte et al., 2022).

A variety of algorithms for different aspects of optical forest monitoring were developed in recent years, including: BFAST (Verbesselt et al. 2010), LandTrendr (Kennedy et al. 2010), Vegetation Change Tracker (Huang et al., 2010), CCDC (Zhu and Woodcock, 2014), BEAST (Zhao et al., 2019), CODED (Bullock et al., 2022), EWMACD (Brooks et al. 2014), Edyn (Brooks et al., 2017), AVOCADO (Decuyper et al., 2022), NRT-MoNITOR (Shang et al., 2022), ForWarn II (e.g. Norman and Christie, 2020), and FORDEAD (Dutrieux et al., 2021a). An overview of forest monitoring methods is also provided in Hirschmugl et al. (2017) and Gao et al. (2020). Hirschmugl et al. (2017) highlight the importance of preprocessing the satellite data, such as cloud detection and radiometric correction, which we also emphasize in this study.

International forest disturbance products of different scales include the Global Forest Watch (2023) datasets, which are either derived from Landsat (GLAD alerts; Hansen et al., 2013) or from Sentinel-2 data (GLAD-S2 alerts) and are combined with the Sentinel-1-based RADD alerts (Reiche et al., 2021). Many other national and regional forest monitoring products are available such as the SAR-based DETER-R (Doblas et al., 2022a). For details on other SAR-based near-real-time operational forest disturbance detection systems produced by various research agencies (INPE, in Brazil, CESBIO, in France, JAXA, in Japan, and Wageningen University, in the Netherlands), we refer to a recent study comparing various algorithm results with the GLAD-S2 mapping performance (Doblas et al., 2022b).

In the last few years, Germany was hit hard by widespread forest disturbances due to storms, prolonged periods of drought, and insect pests. Between 2018 and 2022 alone, an estimated area of 490 km^2 was affected, resulting in 255 million m^3 of damaged timber (BMEL, 2022). The estimated economic loss to the German forestry sector exceeded 12.7 billion euros between 2018 and 2020 (Möhring et al., 2021). Consequently, various actors in the field of remote sensing in the research, private, and commercial sectors are currently engaged in assessing forest disturbances. Several nationwide products for forest vitality and disturbance detection exist, most of which can be accessed online: (i) annual forest disturbances between 1986 and 2020 are available for all of Germany as part of the European forest disturbance map based on Landsat data (Senf and Seidl, 2021a). (ii) The European Forest Condition Monitor (2023) is intended to visualize the condition of forests in Germany and Europe by determining the greenness of the vegetation and comparing it to long-term observations using MODIS data (Buras et al., 2021). (iii) Nationwide tree canopy loss information between January 2018 and April 2021 was provided by Thonfeld et al. (2022). Their analysis was based on monthly image composites of the disturbance index derived from Sentinel-2 and Landsat-8 time series. (iv) Forestwatch-DE (2023) reports annual changes in the vitality of forests as a service. Their analysis is based on vegetation indices derived from Sentinel-2 data and compared with the reference year 2017. (v) Waldmonitor Deutschland (2023) provides nationwide information on forest damage areas (2017–23), dominant tree species, forest vitality trends between 2016 and 2021, and drought stress indicators based on Sentinel-2 data. A very detailed overview of earth observation based forest monitoring activities in German is given in Holzwarth et al. (2023).

Undoubtedly, these products are valuable resources to derive annual to intra-annual information on disturbances at the national level. However, some limitations exist. Besides the limited spatial coverage of the products, their limitations are mainly related to the transparency of their quality assessment. All products suffer in some way from incomplete documentation, the absence of different disturbance categories, the use of different forest definitions, or the absence of independent validation procedures.

A uniform, timely, and up-to-date reporting of forest disturbances at the national level is necessary for optimal crisis management of forest disturbances and the associated measures of politics, economy, and society for the adaptation of forest ecosystems in Germany. So far, forest disturbance data are collected by the Federal States and compiled by the Federal Ministry of Food and Agriculture at the national level, due to the lack of a corresponding operational national system. Due to different procedures of the actors involved, collecting complete, uniform, and reliable information about forest disturbances remains a challenging and often slow process. To overcome these problems, the Federal Ministry of Food and Agriculture commissioned the development of a novel, highly automated approach for the nationwide monitoring of forest disturbances. As a result, the Fernerkundungsbasiertes Nationales Erfassungssystem für Waldschäden (FNEWS, https://www.fnews-wald.de/en/) project was initiated with the aim of timely detection of disturbed areas in order to quantify the salvage timber volume across Germany based on open remote sensing data.

Our annual forest disturbance product uses a straightforward approach based on freely available multispectral Sentinel-2 data at a 10-m spatial resolution. Our definition of forest disturbance includes both natural causes, resulting from windthrow, drought, insect pests, and non-natural causes, resulting from salvage logging or regular forest practices. Our method generates a filtered time series (Puhm et al., 2020), removes low-quality pixels, and detects forest disturbances based on the changes in reflectance accompanied by reduced vitality. We acquired high-resolution reference imagery for four representative study areas. From this imagery we created a detailed set of training data of different forest disturbances. Moreover, we generated 20 validation sets from this imagery, covering many different forest types and disturbance patterns. We validated and analyzed the product in diverse parts of our study areas, and transparently documented the validation process.

The main objective of this article is to present a reliable annual forest disturbance monitoring approach for Germany and to assess the thematic accuracy and spatial estimates of the annual product. Due to the high environmental heterogeneity of Germany, the system is likely to be transferable to other central European countries.

Materials and Methods

Study areas

We conducted the study in four typical forest areas of Germany: (A) southern Lower Saxony, (B) Saxony, (C) south-western Baden-Wuerttemberg, and (D) northern Bavaria (Fig. 1). These areas cover a wide range of tree species, age classes, forest structures, and site conditions (Table 1). Between January 2018 and August 2022, storms, drought, and heavy bark beetle infestations led to persistent forest disturbances. In total, the study areas cover

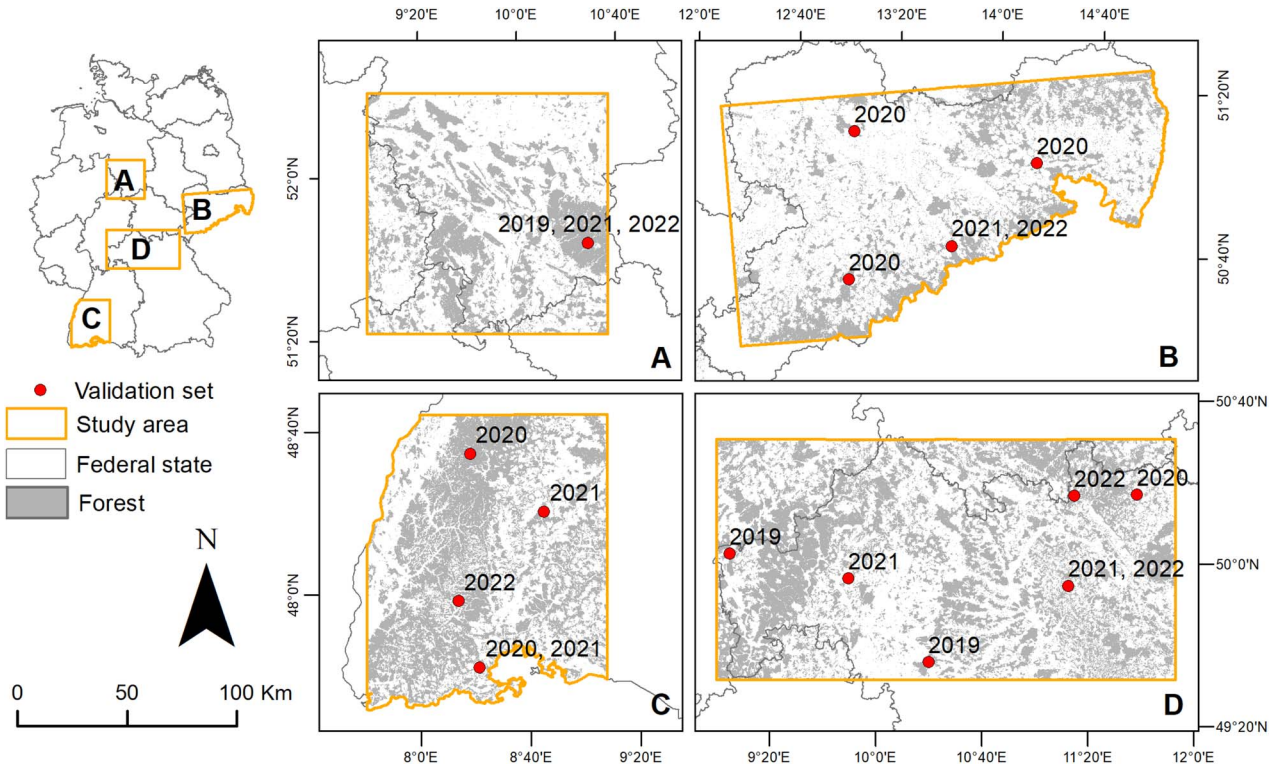

Figure 1. We used four study areas covering up to two Sentinel-2 tile footprints: (A) southern Lower Saxony, (B) Saxony, (C) south-western Baden-Wuerttemberg, and (D) northern Bavaria. Each study area consisted of three to seven validation sets, shown here with their validation year. Each validation set contained several hundred sample points.

Table 1. Summarized characteristics of the four study areas. The annual precipitation represents the mean from 1981 to 2010 (DWD); the forest cover is based on Langner *et al.* (2022), and the disturbance type is based on our reference imagery.

	Study area	Elevation [m a.s.l.]	Annual precipitation [mm]	Forest cover [%]	Disturbance types
A	Southern Lower Saxony	5–927	585–1760	37	Windthrow, bark beetle
B	Saxony	73–1215	562–1242	28	Windthrow, bark beetle
C	South-western Baden-Wuerttemberg	119–1493	508–2015	77	Windthrow, bark beetle
D	Northern Bavaria	102–1045	530–1338	40	Windthrow, bark beetle, complex disturbance

~2 407 000 ha (22%) of the national stocked forest area in Germany. We performed both training and validation inside these study areas. The following information regarding disturbance types and tree species composition originate from the state institutes of forestry of each state including (A) the Northwest German Forest Research Institute (NW-FVA), (B) the Competence Centre for Wood and Forestry in Saxony (SBS), (C) the Forest Research Institute Baden-Wuerttemberg (FVA-BW) and (D) the Bavarian State Institute of Forestry (LWF).

In study area (A) in southern Lower Saxony, homogeneous stands of Norway spruce (*Picea abies*) occur predominantly at higher elevations in the Harz mountains, which is partly protected as a National Park. Forests in this region were profoundly hit by storm Friederike (18 January 2018), and spruce stands were heavily affected by bark beetle (*Scolytinae*) infestations in recent years.

In study area (B) in Saxony, the predominant tree species in the region are Scots pine (*Pinus sylvestris*) in the north and Norway spruce (*P. abies*) in the south, which thrive in both mixed forests and pure stands. The study area includes the Saxon Switzerland National Park, where forest management is strictly limited to roads and hiking trails, to preserve the park's natural state. Saxony experienced notable consequences of the winter storm Friederike (18 January 2018) and the autumn storm Fabienne (23 September 2018), with scattered damage rather than widespread devastation. Bark beetle (*Scolytinae*) infestations in subsequent years led to both scattered damage and, in certain instances, substantial disturbance events.

Most forest in study area (C) in south-western Baden-Wuerttemberg is commercially managed. Exceptions include the Black Forest National Park. Norway spruce (*P. abies*) and silver fir (*Abies alba*) are the dominant species in the Black Forest, whereas European beech (*Fagus sylvatica*) is dominant in most other parts of the study area. Further species include oak (*Quercus* spp.) and *Douglas fir* (*Pseudotsuga menziesii*). Most of the forest damage in this region is due to bark beetles (*Scolytinae*). The combination of multiple dry summers and severe storms, including Eleanor (3 January 2018) and Friederike (18 January 2018), weakened the forest, which led to infestation by this pest. Scattered small-scale forest disturbances throughout the state are the consequence with some differences in distribution patterns across the state due to tree species composition and geography.

The dominant tree species in study area (D) in northern Bavaria are European beech (*F. sylvatica*) and oak (*Quercus* spp.) in the center, Scots pine (*P. sylvestris*) in the south, and Norway spruce (*P. abies*) in the east. This study area has seen a multitude of disturbance types, including windthrow from the summer storms Fabienne (23 September 2018) and Bernd (18 August 2019), complex disturbance on beech stands in 2019, gypsy moth (*Lymantria dispar*) infestation on oak stands in 2018 and 2019, and continual bark beetle (*Scolytinae*) infestation of spruce since 2019. The disturbance severity ranges from scattered complex disturbance to large-scale windthrows and disturbances related to bark beetle infestation.

Data and methods

Our disturbance product builds on three Sentinel-2 bands with a 10-m spatial resolution, namely, bands 3, 4, and 8 corresponding to the green, red, and near-infrared spectra. The value of band 2 (blue) for detecting forest disturbances is limited because of its high sensitivity regarding atmospheric artefacts and therefore unfavorable signal-to-noise ratio (SNR). We tested the integration of other Sentinel-2 bands with lower spatial resolution in the development phase, but ultimately decided to use a set of bands with homogeneous spatial resolution for the final product.

The disturbance analysis only considers forest pixels based on the stocked forest area map (Langner *et al.*, 2022). This map is based on Sentinel-2 imagery from 2018 and displays stocked forest areas in Germany of at least 0.25 ha and a coverage of at least 50%, with an overall accuracy of 96.5%. In the following, we refer to this stocked forest area map as the forest map. To generate the annual disturbance product, we evaluate a dense time series of Sentinel-2 data using a fully automated processing chain. The processing chain is divided into three parts: preprocessing, time series filtering, and disturbance map updating.

The first part is image preprocessing, in which a cloud mask is first calculated for all available top-of-atmosphere (TOA) images using the Fmask algorithm (Qiu *et al.*, 2019). Further image preprocessing continues only if a defined minimum percentage (e.g. 10%) of valid pixels is present after cloud masking. We consider pixels classified as cloud, cloud shadow, or snow as invalid. When a sufficient fraction of valid pixels is present, atmospheric correction follows using the sen2cor software provided by the European Space Agency (Louis *et al.*, 2016). The result is data of processing level L2A, which radiometrically corresponds to a calibrated bottom-of-atmosphere (BOA) reflection. After atmospheric correction, the next step is to perform an improvement of the geometric positional accuracy of the images with respect to each other. For this purpose, a cloud-free reference image is selected for each tile. A coregistration process adjusts each image of the time series stack in its position to the reference image. For a given image pair, the coregistration process evaluates the cross-correlation of the red spectral band at regularly spaced sample locations (spacing: 1 km) using a search window of 7 by 7 pixels. Sample locations are discarded where either the cross-correlation maximum is below 0.7 or the back-matching distance exceeds one pixel. For all remaining samples, we use the sub-pixel offsets of the cross-correlation maximum in x and y direction to estimate the coefficients of an eight-parameter projective transformation polynomial through least squares adjustment. Finally, we apply the estimated transformation to all image bands using bicubic interpolation. In the next step of preprocessing, invalid pixels are masked out based on the Fmask classification result. Finally, the Minnaert method (Gallaun *et al.*, 2007) is used to perform a topographic correction. The underlying terrain model is the Copernicus Digital Elevation Model (COP-DEM), which is available for Europe in 30-m resolution (European Space Agency, 2021). As part of the topographic correction, a BRDF correction (bidirectional reflectance distribution function) is also performed, which compensates for systematic effects due to different viewing angles depending on the satellite orbit (Roy *et al.*, 2017).

For time series filtering, we use the method described by Puhm *et al.* (2020), which builds on a structural time series model and Kalman filtering (KF, Harvey, 1989) for use with Sentinel-2 surface reflectance signals. Structural time series models provide the means to decompose an input signal into additive components such as trend and seasonality. The operation of the KF is recursive, meaning that the time series is processed one image at a time. Each recursion consists of two steps: the time update and the measurement update. In the time update, a model prediction of the expected reflectance at the time of acquisition of the current image is made, starting from the date of the last processed image. In the measurement update, the model prediction (and with it the model parameters) is corrected by integrating the current data. For the correct representation of seasonal signal components, it is necessary to process a multi-year historical time series to train the model.

After the training phase, KF provides two outputs for each new image available in the monitoring phase, which are used for the later derivation of the products in the last part of the processing chain. On the one hand, these are the so-called innovations, which represent statistically normalized differences between the model predictions and the actually observed signal values. On the other hand, filtered versions of the input images are produced, which we deem to have favourable properties for annual large-scale forest disturbance mapping. Gaps in the original images caused by cloud masking are filled by the model predictions, and the remaining signal noise is reduced. Through measurement updates, sustained signal changes caused by forest disturbances are also visible in the filtered images with a certain delay. However, a challenging property of the signals used here is that they contain not only normally distributed Gaussian noise but also outliers due to unmasked clouds, cloud shadows, or snow as well as systematic influences of smaller magnitudes due to poor atmospheric conditions (haze, heightened aerosol levels, adjacent clouds).

When the full near-real-time signal-tracking capability of the KF or comparable methods is unlocked, the existence of these signal artefacts creates a dilemma. At filter runtime, both artefacts and changes of the ground state, e.g. phenology shifts or disturbances, manifest themselves in the form of larger-than-normal innovations. Therefore, advanced signal tracking and noise reduction requires a highly reliable assessment of each innovation's trustworthiness. On the one hand, all innovations corresponding to signal artefacts should be ignored or down-weighted, but on the other hand, innovations corresponding to changes of the ground state should be used with full weight in the measurement update. In our experience, semi-transparent artefacts from clouds or haze are often not accounted for at all by traditional cloud masks. However, semi-transparent clouds and haze occur quite frequently in Germany.

We therefore developed a new approach to quantify signal quality in order to increase the robustness of the time series filtering and the products based on it against these effects. The intention is to produce a pixel-wise continuous quality index instead of a binary mask derived from typical hard classification approaches. In its current state of development, the concept is tailored to forest applications and therefore makes use of

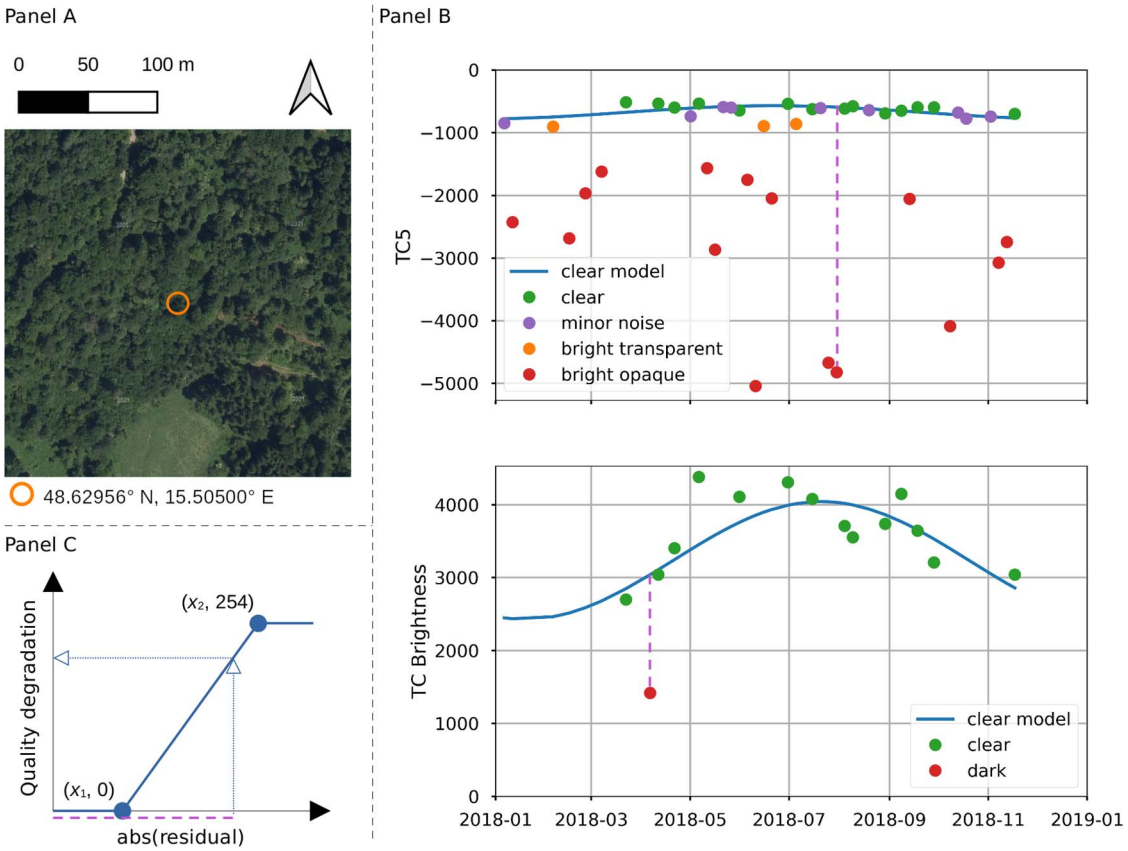

Figure 2. Panel A shows a true-color aerial image (Orthofoto Österreich, 2023) with a coordinate marker in the center. In Panel B, temporal profiles of the tasselled cap (TC) transformation features TC5 and brightness are plotted for the broadleaved forest pixel marked in Panel A. We show all Sentinel-2 observations acquired in 2018 for tile 33UWP, relative orbit 122. The blue lines represent harmonic first-order models based on clear observations. Vertical dashed lines represent examples of model residuals, which are mapped to a quality degradation index using a piecewise linear function indicated in Panel C. We refer the reader to the text for a detailed explanation of the observation labels in Panel B and the parameters of the function in Panel C.

simplified assumptions. We propose to group signal artefacts into 'bright' (atmospheric, snow) and 'dark' (shadows) categories and claim that the specific properties of two features of the tasselled cap transformation (Crist, 1985) are particularly useful for detecting signal interference caused by them. It is assumed that signal interference from cloud shadows correlates with a drop in observed Tasselled Cap Brightness (TCB) and that the normal TCB level of a forest pixel is approximated sufficiently accurate by a first-order harmonic model. Large negative model residuals correspond to an increased quality degradation. The second assumption is that atmospheric artefacts or snow correlate with a drop in the fifth feature of the tasselled cap transformation (TC5). Fig. 2 gives examples of both assumptions by showing time series of TC5 and TCB over 1 year including a thematic label of each observation in Panel B. The respective pixel location is described by Panel A. The interpretation of TC5 is less intuitive than that of TCB, but the high sensitivity of this feature can be explained by the high negative weighting of the blue spectral band in the transformation. Moreover, it is a low-order feature, which means that most of the normal non-artefact signal dynamics of forest pixels are already contained in the higher-order features. Thus, the normal TC5 level is practically invariant to phenology and can be captured very well by a simple first-order harmonic model. This is a key advantage over using the blue band itself and enables to reliably down-weight pixels affected by transparent artefacts. As with TCB, negative model residuals correspond to increased quality degradation. Fig. 2C describes the mapping of model residuals to quality degradation values between 0 (best quality) and 254 (very bad quality) using a piecewise linear function with parameters x_1 and x_2 set differently for TC5 and TCB. The derived degradation values have no particular interpretation, but simply reflect our choice of using 8-bit unsigned integer as output data type. We incorporate the computed quality degradation index into the observation model of the KF (Puhm *et al.*, 2020). Absolute residual values below x_1 will not cause any down-weighting of corresponding observations in the measurement update, while absolute residual values equal and above x_2 cause corresponding observations to be disregarded entirely. This extension improves the signal tracking ability of the filter considerably, allowing us to produce filtered (i.e. radiometrically stabilized) versions of the input images. A simplified flowchart of the recursive filtering process is given in Fig. 3.

We sequentially evaluate the KF colour-infrared images (CIR; Sentinel-2 bands B08, B04, B03) at the end of August to update the disturbance map using the workflow depicted in Fig. 3. There are a few reasons why this date was considered as a good compromise for updating the annual disturbance map: (i) most of the new bark beetle calamities of the current season are already visible, (ii) the chance of acquiring high-quality imagery close before the target date is high, and (iii) sun elevation is still high, which increases the SNR. The filtered CIR imagery with reduced colour bit depth are published by Oehmichen *et al.* (2024) to complement the disturbance products. Our main underlying assumption is that, through the filtering process, the information of all images

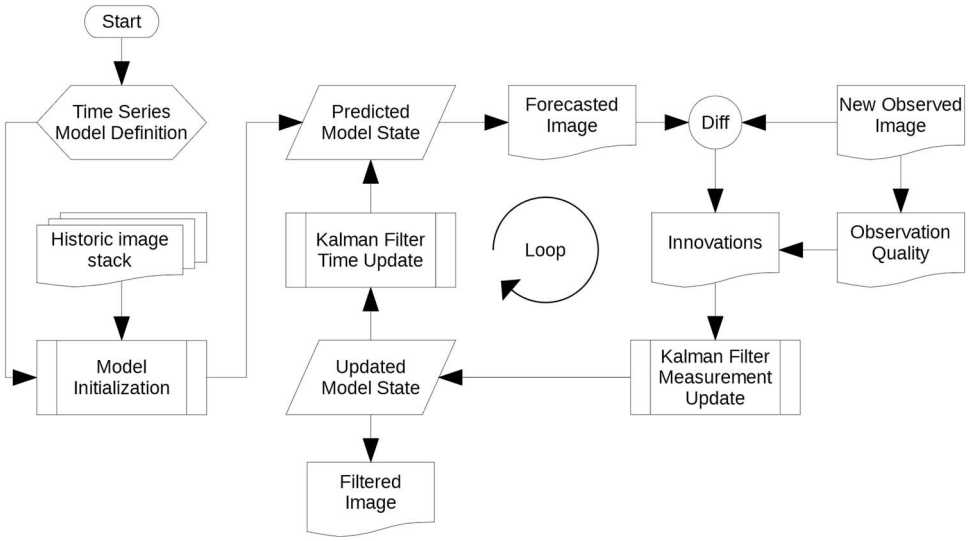

Figure 3. Flowchart of the recursive Kalman filter (KF) operation. The observation quality index is used to weight the innovations before the measurement update.

available up to the respective target date has been considered to derive the best estimate of the current ground state in a near-real-time manner.

In the first step of the workflow, feature values are extracted at predefined sampling locations representing broadleaved and coniferous forest. In each study area, a regularly sampled grid of points using a spacing of 1 km is defined and reference forest type labels are extracted from the respective Copernicus High Resolution Layer (Copernicus Land Monitoring Service, 2020). Using the extracted samples for each class, we set up three statistical models. In Fig. 4, they are labelled M1, M2, and M3. For M2, the median Normalized Difference Vegetation Index (NDVI) provides a simple univariate decision boundary to label pixels with lower vitality, i.e. below the median NDVI of the respective forest class at the target date. For M1 and M3, we assume that both the characteristic reflectance of forest (M1) and spectral differences compared with the last year (M3) can be modelled as multivariate Gaussian distributions with the respective location and covariance estimated for each class using the robust Minimum Covariance Determinant (MCD) algorithm (Rousseeuw, 1984). These statistical models provide the means to label forest and nonforest pixels, as well as pixels with unusually large spectral differences, based on certain levels of significance α given in Fig. 4. The feature space we work in has three dimensions; hence, each class is modelled by an ellipsoid in the feature space. The location and covariance estimated using the MCD method define the center of the ellipsoid, its proportions, and the orientation of its axes; the parameter alpha adjusts its size. Points inside the ellipsoid belong to the class, so the value of α tunes the decision boundary. The various decision boundaries of the models M1, M2, and M3 allow us to infer if a monitored pixel has a certain property or not. All pixels sharing a certain property form a set. However, one pixel can be an element of multiple sets. The legend in the bottom part of Fig. 4 summarizes how sets are represented in the illustration and which properties are inferred.

In the second step of the workflow (lower half of Fig. 4), the initially empty (i.e. all pixels healthy) disturbance map is updated. Pixels may be moved not only from the set of healthy pixels to the set of disturbed pixels, but also the other way around (recovery). We isolate disturbance and recovery candidates by combining the different property labels derived in step one through set intersections. The intersection of two or more sets is defined as the set containing all elements that are members of every input set. The 'flower-like' elements of Fig. 4 are the graphical representations of the intersection operation, with the intersection result coloured in orange.

For example, the properties isolating a new coniferous disturbance candidate are therefore the following:

1) The pixel must be an element of the set of coniferous pixels according to the reference classification. This is denoted by the property label 'Con'.
2) The pixel must be an element of the set of healthy pixels in the previous epoch, denoted by the property label 'H_{prv}'.
3) The pixel must be an element of the set of nonforest pixels according to M1 (property label 'nF'). This means that the pixel does not belong any more to the coniferous or the deciduous class based on a level of significance $\alpha = 1\%$. The quality of M1 in separating coniferous from deciduous forest is therefore of low importance as long as the overall approximation of the spectral appearance of healthy forest is acceptable.
4) The pixel must be an element of the set of pixels with lower vitality according to M2 (property label 'LoV_c'). This means the pixel has an NDVI value lower than the median of all sampled coniferous reference pixels. By itself, M2 has the lowest power in isolating disturbances compared to the others, but it effectively complements M1 and M3 by restricting the disturbance direction.
5) The pixel must be an element of the set of pixels showing significant and anomalous spectral differences with respect to the previous epoch according to M3 (property label '$Hi\Delta_c$'). Recalling that normal spectral differences are modelled per forest class as multivariate Gaussian distribution, this means that the pixel's difference vector fails a χ^2 outlier test with level of significance $\alpha = 5\%$.

To reduce the number of false positive detections in the product and avoid having individual pixels show up as disturbance patches, we define a Minimum Mapping Unit (MMU) of 0.1 ha or

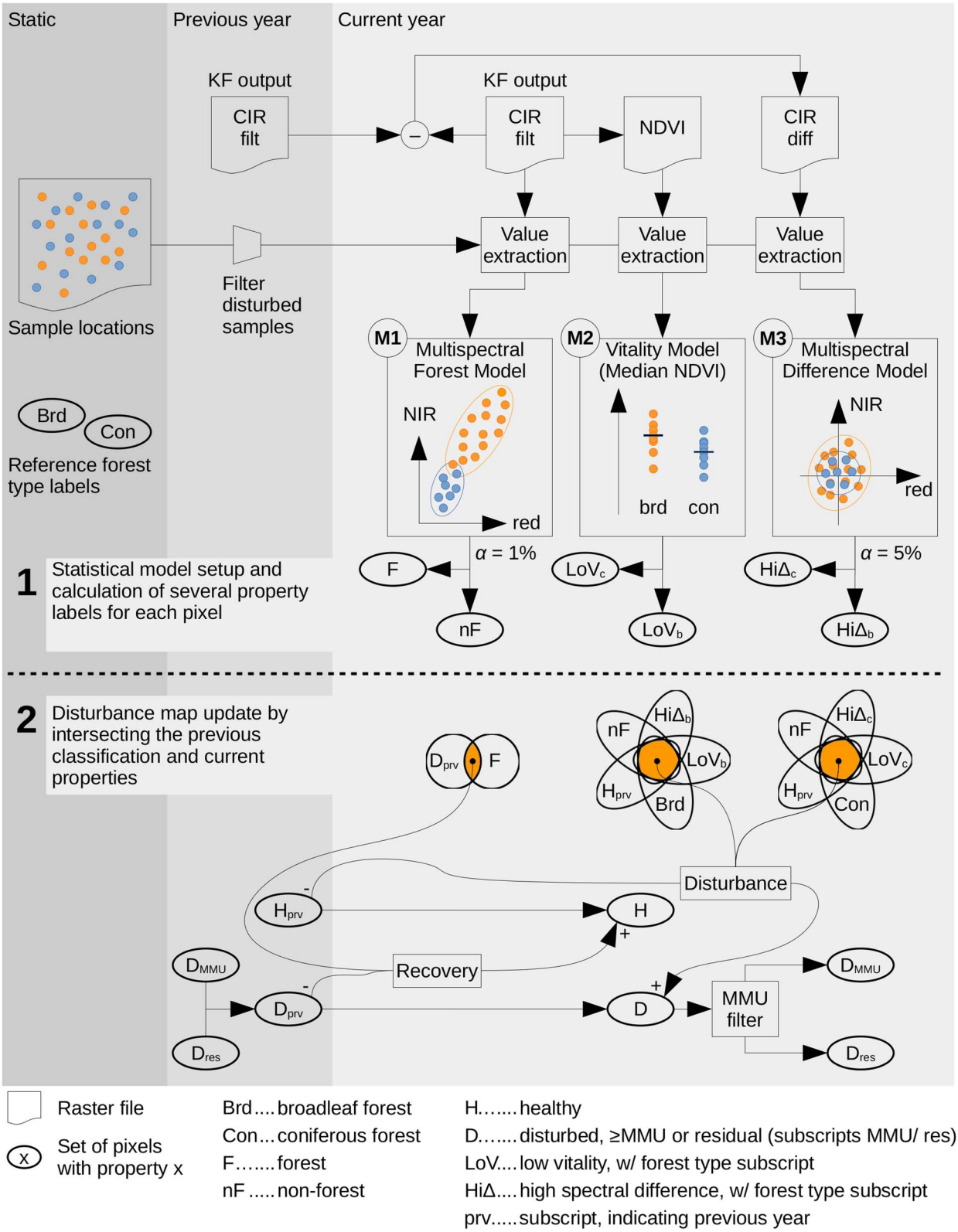

Figure 4. Sequential evaluation of Kalman Filtered (KF) colour-infrared images (CIR, Sentinel-2 bands B08, B04, B03) at the end of August, used to update the disturbance map. In step one, statistical models are set up using feature samples extracted at predefined locations. The spectral (M1) and difference (M3) models use multivariate Gaussians with the respective location and covariance estimated for each class using the robust Minimum Covariance Determinant (MCD) algorithm. Step two illustrates the update process performed within the initial forest map, where pixels may be moved (indicated by arrows with −/+ signs) from the set of healthy pixels to the set of disturbed pixels, but also the other way around (recovery).

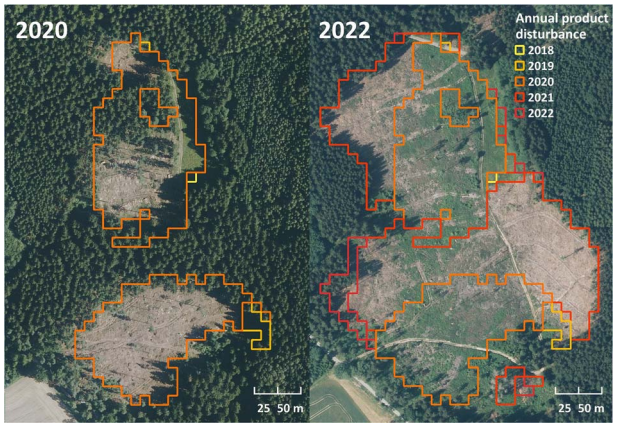

Figure 5. Example of growing disturbance patches. Disturbances detected between 2018 and 2020 are shown on the left side, and disturbances of the same area detected between 2018 and 2022 on the right side. Aerial images from 2020 and 2022 are displayed in the background (© GeoSN).

10 adjacent pixels. A disturbance patch is therefore only included in the disturbance product if at least 10 adjacent pixels have been classified as disturbed by the algorithm (Fig. 5). However, the MMU can be reached over multiple years because the state of residual disturbance pixels is saved, allowing us to also detect small, persistently growing disturbance patches. The individual classification date is recorded per pixel. Therefore, disturbances smaller than the MMU can contribute to the disturbance area of any given year in historic statistical evaluations. The map still does not include small and scattered disturbances never reaching the required MMU, but the spatial resolution of Sentinel-2 is also not lost entirely.

The outlined workflow to derive the annual products is mostly data-driven and requires few user-defined parameters. We determined appropriate values for the significance levels α by evaluating the performance with training data, as described in the following section. We used a single set of parameters for the entire processing chain for all study areas together. While our product update rate is only annual, there are also two important near-real-time aspects involved. Firstly, the workflow design allows quick product generation after the target date, as the computationally expensive processes (Pre-processing, KF operation) can already happen incrementally over the year. Secondly, the advanced signal tracking implementation allows the product to be highly up-to-date at the time of publishing.

Reference imagery

We used high-resolution aerial orthoimages and satellite imagery as a reference to collect training and validation data. The orthoimages have a spatial resolution between 0.1 and 0.2 m. The satellite imagery was acquired by the sensors Pléiades-1B (0.5 m), WorldView-2 (0.5 m), and PlanetScope (3 m). Additionally, to enhance the level of detail and information, we used canopy height models with spatial resolutions between 0.5 and 1 m. Image enhancement techniques, i.e. contrast and histogram adaptation, were applied to guarantee optimal visual interpretation. Position accuracies of the reference imagery were 5 m or better. For validation, we selected reference imagery as close as possible to 31 August to ensure good agreement with the disturbance product target date. Imagery between end of July and beginning of October was available for 17 of the 20 validation sets. For the remaining three validation sets, we used imagery of mid-June and mid-October.

Training data

The training data rely on reference imagery from 2018 to 2021. We created training polygons with a minimum area of 0.1 ha and with a minimum width of 20 m. For the annual product, we used the training data primarily to tune the statistical model parameters (e.g. in Fig. 2). We additionally used the training data for a separability analysis of different disturbance types. We do not present and discuss the results of this separability analysis in this manuscript, but this additional study explains the very detailed training data nomenclature, which is more detailed than would be required for a simple detection of annual forest change areas. The selected disturbance types for training the product are 'windthrow', 'bark beetle', 'gypsy moth', 'complex disturbance', 'indeterminate disturbances', and 'undisturbed' areas (examples in Fig. 6). Complex disturbance refers to damage caused by drought stressor a combination of drought stress and pest infestation. The term 'indeterminate' represents disturbances of indeterminate origin and trees that have already been removed from the forest for unidentified reasons, including salvage logging and regular forest management practices. Additionally, we classified the disturbances into 'scattered' and 'complete' categories. 'Scattered' denotes areas with 30%–90% canopy cover disturbance, and 'complete' refers to training polygons with >90% disturbance (Table 2).

Validation

For the validation of our product, we adhered to the best practice guidelines for assessing the accuracy of land cover change described by Olofsson *et al.* (2014) and elaborated on by Stehman and Foody (2019). The validation workflow consisted of the following steps: stratification (i), random sampling (ii), independent validation (iii), plausibility check (iv), and accuracy calculation (v).

We stratified the forest area into three strata prior to distributing the validation samples, to ensure a sufficient sample size in all areas where we expected different product accuracies. We expected the lowest accuracy in the transition area between disturbed and undisturbed forest, as this was previously observed in similar studies (e.g. Arévalo *et al.*, 2019; Francini *et al.*, 2022). This area is defined as a buffer stratum, which was analysed separately to accurately assess the omission and commission error of the disturbed area. It is comparable to the buffer stratum discussed in Olofsson *et al.* (2020) and used in similar studies (Francini *et al.*, 2022; Thonfeld *et al.*, 2022). The other two strata are the undisturbed stratum and the disturbed stratum. We expected the highest product accuracy in the undisturbed stratum and an intermediate accuracy in the disturbed stratum. The three strata were derived from the disturbed areas (i.e. the patches) identified by the product. The buffer stratum was equivalent to a 10-m buffer inside and outside the circumference of the disturbed area. The disturbed stratum represented the remaining parts of the disturbed area, and the undisturbed stratum represented the remaining parts of the undisturbed forest area (Fig. 7).

Based on the available reference imagery, we created a diverse collection of 20 datasets, which we refer to as validation sets. Validation sets either cover different years or different regions and are located in regions that contain sufficient disturbance pixels for validation. They include various disturbance characteristics, such as different disturbance types, intensities, and disturbance years, to ensure a representative validation data set. The formula to calculate the necessary sample size for the random sampling

Figure 6. Field photos of different disturbance types in our study areas. (A) Complex disease on European beech (Fagus sylvatica) in Bavaria, (B) gypsy moth infestation in June on oak (Quercus spp.) in Bavaria, (C) cleared area, red and grey attack state of bark beetle infestation in Lower Saxony, and (D) windthrow in Saxony (© Philip Beckschäfer, Falkhardt Dau, Hannes Lemme, Christoph Straub).

Table 2. Different disturbance types covered in the training data based on the reference imagery. Information on the years covered, dominant forest tree species, and the extent of both scattered and complete disturbance for each disturbance type. Complete disturbance refers to areas with >90% disturbance.

Primary disturbance type	Study areas	Years	Dominant tree species	Disturbance extent [ha]	
				Scattered	Complete
Windthrow	A–D	2018–20	Spruce and pine	3927	3207
Bark beetle	A–D	2018–21	Spruce	2014	1328
Complex disturbance	C, D	2019–21	Beech	348	2
Gypsy moth	D	2019	Oak	453	306
Indeterminate	A, C, D	2018–20	Spruce	698	3612
Undisturbed	A–D	2018–21	Broadleaf and conifer	4581 (undisturbed)	

per validation set (Formula 1) was adapted from Olofsson *et al.* (2014).

$$n = \left(\frac{\sum W_i \sqrt{U_i (1 - U_i)}}{S\left(\hat{O}\right)} \right)^2 \quad (1)$$

where i = stratum, $S(\hat{O})$ = standard error of the overall accuracy (defined as 0.01), W_i = fraction of the area of the stratum relative to the total area, and U_i = expected accuracy in the stratum.

The sample size depends on the relative sizes of the strata and the expected user's accuracy of each stratum. We expected a user's accuracy of 90% in the disturbed stratum, 75% in the buffer stratum, and 95% in the undisturbed stratum. These expectations were slightly higher than suggested by Olofsson *et al.* (2014) and were based on our experience with similar forest disturbance products. The disturbed and buffer strata often made up <5% of the total forest area. A proportional distribution of the sample points would therefore result in only very few points located within these strata and lead to large standard errors for their accuracies. Instead, we followed the recommendation of Olofsson *et al.* (2014) to randomly distribute a minimum of 100 samples in each stratum. Whenever possible, we generated points with a minimum distance of 50 m to each other to avoid clustering. In total, we created 20 independent validation sets for 2019 to 2022. They cover either different years, different regions or both. In total, we validated 11 019 sample points, ranging from 486 to 732 samples per validation set (Fig. 8).

To assess the product performance, we not only used the same reference imagery that was used for training but also incorporated additional reference imagery from 2022, which we used exclusively for validation purposes. Multiple interpreters manually evaluated whether the Sentinel-2 pixel intersecting the sample point was disturbed on >50% of its area and whether the total disturbance size reached the defined MMU. If both factors were met, we considered the point disturbed; otherwise, we considered it undisturbed. We discarded points in cases where the pixel was not stocked, the pixel was already disturbed in older reference imagery, the MMU was difficult to assess, or the reference imagery was of insufficient quality due to clouds or shadows. The interpreters were not aware of the product prediction at this

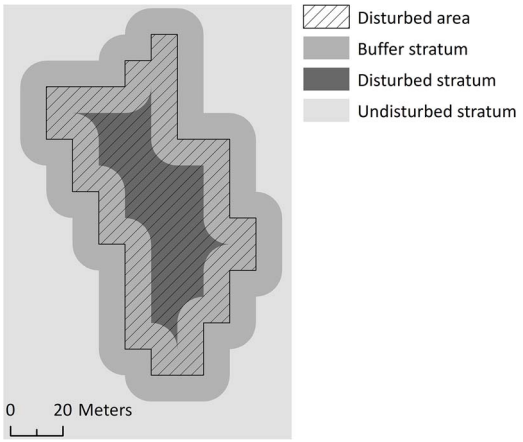

Figure 7. Schematic figure of the three strata used for stratified random sampling. The disturbed area shows the product's classification result, and the buffer stratum is equivalent to a 10 m buffer inside and outside its circumference. The disturbed and undisturbed strata represent the remaining forest areas inside and outside the disturbed area, respectively.

stage of the validation, to ensure an independent assessment. We primarily evaluated the product accuracy with an MMU of 0.1 ha, but for comparison, we also evaluated the effect of an increased MMU of 0.25 ha within five selected validation sets. This allowed us to evaluate the influence of the MMU on the accuracy of the product.

It was sometimes difficult to determine whether a pixel was sufficiently disturbed, due to scattered disturbance patterns, partially disturbed pixels, or low quality of the reference imagery. We therefore created an extensive validation guideline for all photo interpreters to ensure the interpretation was as comparable as possible. We therefore also performed an additional plausibility check, where we looked again at sample points for which the map class and the blind interpretation did not agree. We considered these sample points correctly classified if the product class was plausible. To that end, we assessed the reference imagery and the product together. We considered sample points with partial disturbance plausible, especially those adjacent to large disturbed areas. The geometric accuracy of Sentinel-2 pixels is in the range of 10–12 m (Gascon *et al.*, 2017) which is further improved during coregistration. We therefore evaluated whether the product prediction would be correct if the pixel was moved by a few metres. Accuracies given in the results section include the plausibility check, whilst accuracies without the plausibility check are presented in the Supplementary Material 2.

To estimate the product accuracies, we generated area proportional error matrices according to Stehman (2014). We used the overall accuracy (OA), the user's accuracy (UA), the producer's accuracy (PA), and the F1 score to describe the product's reliability. The OA describes the percentage of the area classified correctly by the product (Formula 2). The UA, also known as precision, represents the percentage of the area classified as the disturbed area that was confirmed to be disturbed through comparison with the reference imagery (Formula 3). It therefore gives us information about the commission error or where the product overestimates the disturbance. The PA, also known as recall, represents the percentage of disturbed area correctly classified by the product (Formula 4). It provides information about the omission error or where the product underestimates the disturbance. The F1 score represents the harmonic mean of UA and PA (Formula 5) We always used UA, PA, and F1 in reference to the disturbed area and never in reference to the undisturbed area. The area weighted analysis allowed us to calculate accuracy metrics for individual validation sets, as well as summarized metrics for varying groups (e.g. by year, study area, or stratum) and all validation sets combined.

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$UA = \frac{TP}{TP + FP} \quad (3)$$

$$PA = \frac{TP}{TP + FN} \quad (4)$$

$$PA = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (5)$$

where TP = true positive, FP = false positive, FN = false negative, and TN = true negative.

Results

The purpose of our annual forest disturbance product is to reliably detect and map forest disturbances with a uniform method for all forests in Germany. The diverse disturbance patterns in German forests necessitated an equally varied and spatially distributed validation setup to ensure a representative validation result with reliable accuracies. To that end, we extensively validated our product with 20 separate validation sets spread across four study areas (A–D) in different parts of Germany (see Fig. 1 for locations). For all four study areas combined, our disturbance product achieved an overall accuracy (OA) of 99.1 ± 0.1%. Disturbed areas were detected with a user's accuracy (UA) of 84.4 ± 2.0% and a producer's accuracy (PA) of 85.1 ± 3.4% resulting in an F1-score (F1) of 84.7% (Fig. 9). All values for UA, PA, and F1 refer only to the accuracy of the disturbed area and not the undisturbed area. All values, including the OA, are area-weighted accuracies. Accuracies were highest for the undisturbed stratum and lowest for the buffer stratum. On average, OA was 93.4 ± 1.1% in the disturbed stratum, 84.8 ± 2.0% in the buffer stratum, and 99.7 ± 0.1% in the undisturbed stratum. Without the plausibility checks, the OA was 98.6 ± 0.1%, the UA was 74.9 ± 2.4%, the PA was 76.6 ± 3.7%, and the F1 was 75.7%, for all validation sets combined.

Accuracies in the four study areas

Comparing the study areas, we observed an OA of at least 95.8 ± 1.1% in all cases. UAs of the disturbed area ranged from 75.2 ± 5.5% for study area (C) to 92.4 ± 3.2% for study area (A). PAs ranged from 76.3 ± 11.2% for study area (C) to 96.2 ± 2.8% for study area (D). Study areas (B) and (C) had similar UAs and PAs. Differences were larger between the accuracies in study areas (A) and (D). In the following, we refer to individual validation sets by their location with their validation year in parentheses. A list of all validation sets and their corresponding accuracies is included in the Supplementary Material 1.

Study area (A) consists of three validation sets, all conducted within the same area of the Harz National Park in the years 2019, 2021, and 2022. Most disturbances were caused by continual bark beetle infestation (NW-FVA, Fig. 10A). These validation sets had the lowest OA, i.e. 93.5 ± 2.1% to 97.8 ± 1.4%. The disturbances covered 7.5 ± 0.2% to 33.0 ± 0.4% of the forest area, which was the largest percentage of all validation sets. UA ranged from 76.5 ± 11.5% to 93.8 ± 4.8%, PA from 75.2 ± 13.7% to 95.9 ± 5.8%, and F1 from. In Harz (2019, 2021), the UA exceeded the PA by ~18%, indicating an underestimation of the actual disturbance extent.

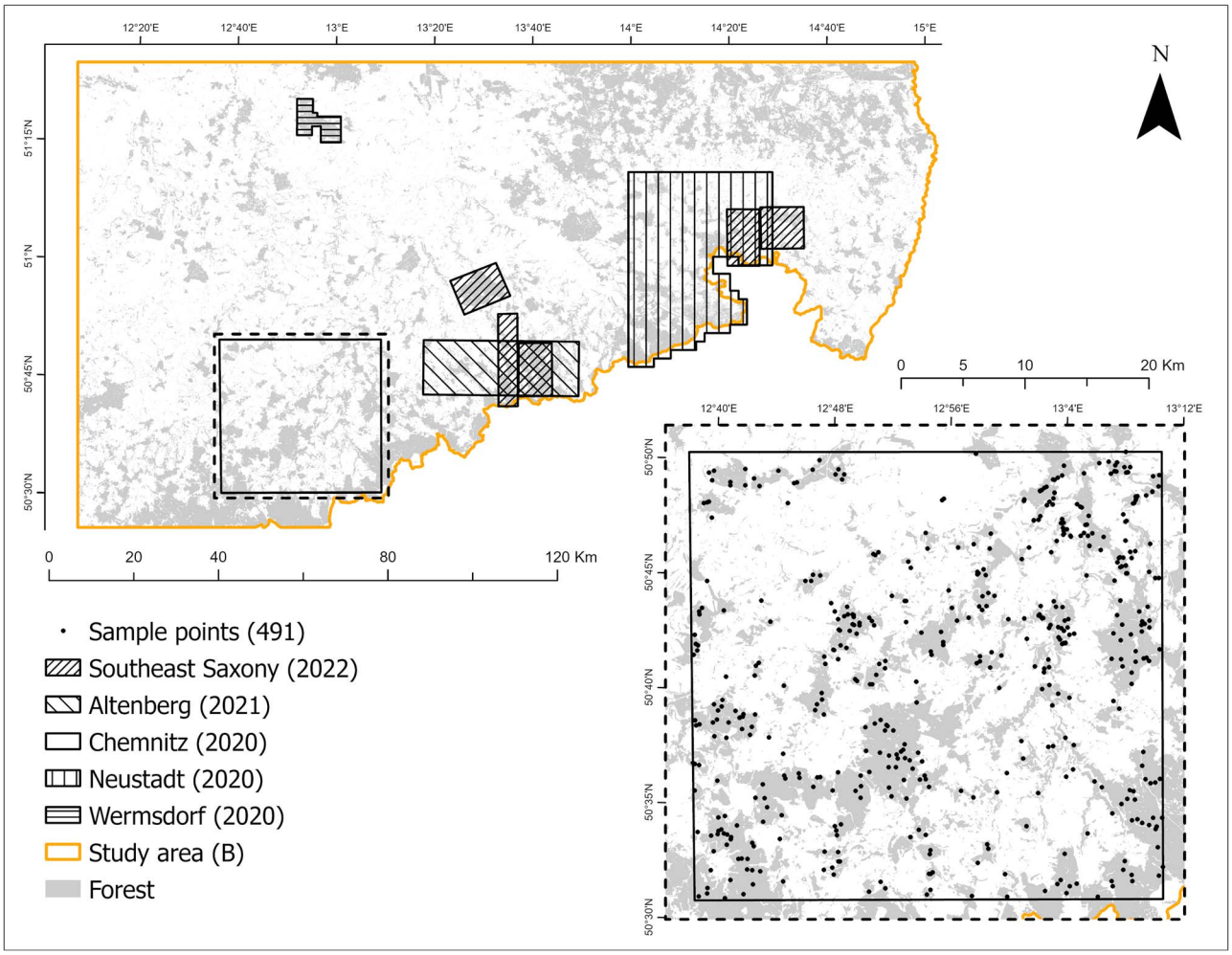

Figure 8. Extents of the five validation sets in study area B (acquisition year in parenthesis). The sample points (total number in parenthesis) are given for the Chemnitz subset.

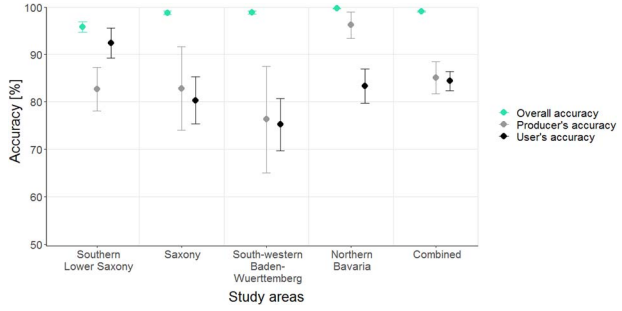

Figure 9. Overall accuracy for both disturbed and undisturbed forest, as well as user's and producer's accuracies of disturbed forest in the four study areas, and the combined accuracies of all four study areas together. The error bars represent the 95% confidence intervals.

In Harz (2022), the PA exceeded the UA by 19.4%, indicating an overestimation.

Study area (B) included three validation sets from 2020 and one validation set each from 2021 and 2022. The validation sets were spread across Saxony. They included both small, isolated disturbance patches of indeterminate origin and large, steadily growing patches related to bark beetle infestation (SBS, Fig. 10B). UA ranged from 70.4 ± 10.9% for the validation set for Southeast Saxony (2022) to 93.6 ± 5.5% for the validation set for Chemnitz (2020). In both cases, most disturbances were bark-beetle related, but Chemnitz (2020) mainly featured large disturbance patches. PA were high, i.e. 84.0 ± 10.1% to 100.0 ± 0.0% for all validation sets, except for Chemnitz (2020) where the PA was only 68.9 ± 31.0%.

Study area (C) consisted of five validation sets from 2020 to 2022 in south-western Baden-Wuerttemberg, including the Black Forest. The disturbance patches were generally smaller than in the other study areas, scattered, and often bark-beetle related (FVA-BW, Fig. 10C). This study area had the lowest overall UA and PA. UA ranged from 51.7 ± 11.6% to 84.0 ± 7.5% and PAs from 60.5 ± 37.3% to 85.5 ± 24.5%. The two lowest UA in this study area were for the Black Forest (2020) and Baar (2021) validation sets, with 51.7 ± 11.6% and 59.4 ± 13.0%, respectively. They included 385 ± 33 ha and 78 ± 13 ha of disturbed area, representing only 1.2% and 0.1% of the forest area, respectively. The highest UA (84.0 ± 7.5%) was for the Waldshut (2020) validation set, which was also where the most disturbance occurred, covering 2968 ± 73 ha.

Study area (D) included seven validation sets from 2019 to 2022 in northern Bavaria. Disturbance patterns and causes were very diverse and even included temporary disturbances due to gypsy moth infestations (LWF, Fig. 10D). UA were high, i.e. 82.5 ± 9.7% to 96.6 ± 4.1%, except for the Lower Franconia (2019) validation set. There, the disturbances consisted of scattered complex disturbance and large-scale gypsy moth infestation. The corresponding UA was only 31.0 ± 8.4%, which was the lowest of

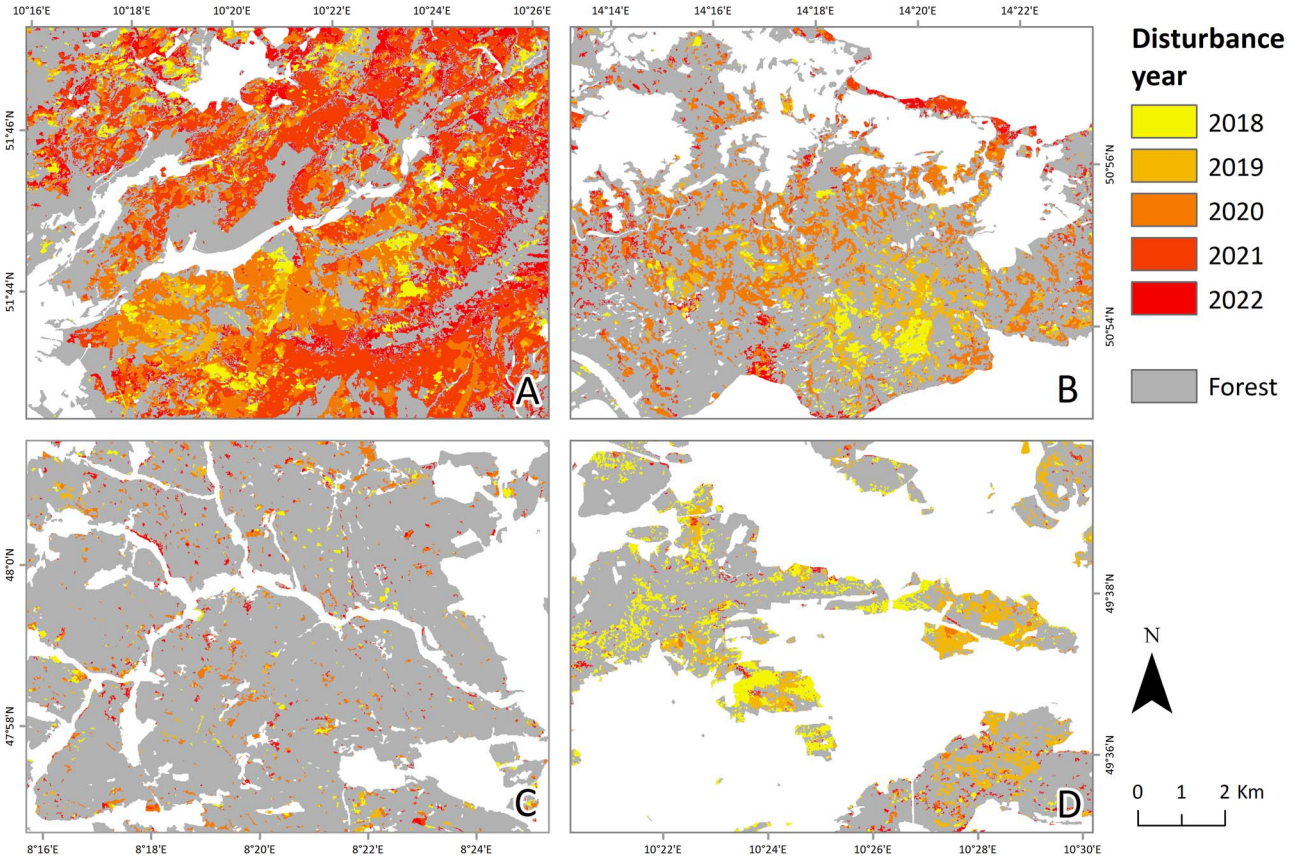

Figure 10. Examples of different spatial and temporal patterns of disturbance in the four study areas. The letters A to D refer to the study areas displayed in (Fig. 1). A: Extreme bark beetle infestation of spruce stands in the Harz mountains (NW-FVA). B: Heavy bark beetle infestation in Saxon Switzerland National Park (SBS). C: Scattered disturbance of mostly indeterminate origin in the southern Black Forest (FVA-BW). D: Complex disease and gypsy moth infestation in broadleaf stands in northern Bavaria in 2018 and 2019 (LWF).

all validation sets. PA ranged from 85.9 ± 24.0% to 100.0 ± 0.0% across all validation sets, resulting the highest combined PA among all study areas. The four validation sets with large-scale bark beetle disturbances had high accuracies, as in the other study areas. Notably, the Franconia (2021) validation set, with mainly small patches of scattered disturbances, also scored very high, with a UA of 88.5 ± 7.5% and a PA of 96.4 ± 4.9%.

Mapped and actual disturbance extent

Besides assessing the accuracy of our product when mapping individual disturbance patches, we also evaluated how reliably it mapped the total disturbance extent. The total disturbance extent represents the sum of all disturbance patches for a chosen region and period. For the forest area covered by the validation sets, our product mapped 27 582 ha of forest disturbance during the entire observation period of September 2018 to August 2022. This represented 2.86% of the total forest area. Using our validation sets, we estimated the error-adjusted actual disturbance extent at 26 847 ± 201 ha, or 2.78 ± 0.02% of the total forest area. Overall, our product therefore slightly overestimates the actual disturbance extent.

We also evaluated the disturbance extent for individual years to assess inter-annual variations (Table 3). The mapped disturbance extent was outside the confidence interval of the actual disturbance extent for all years. For 2020 and 2021, the product underestimates the disturbance extent by 9.1% and 5.8%, respectively. For 2019 and 2022, we observe overestimations of 11.1% and 22.9%. This overestimation in 2022 is consistent across all five validation sets.

Comparing minimum mapping units

Applying a larger MMU resulted in notably higher accuracies for three of the five validation sets, at the cost of reducing the total detected disturbed area (Table 4). The validation set for Southeast Saxony (2020) had a F1 of 84.2% with an MMU of 0.1, but an increased F1 of 98.3 % with an MMU of 0.25. The F1 of the Franconian Forest (2020) validation set also increased from 87.6% to 94.1% when we increased the MMU. For Baden-Wuerttemberg (2021), we similarly observed an increase of 68.8% to 84.7% for the F1. The remaining two validation sets did not show notable changes in their accuracies. Naturally, increasing the MMU reduced the total disturbed area detected by the product. This reduction in the total disturbed area ranged from 4.9% for the Harz (2021) validation set up to 30.9% for Southeast Saxony (2020).

Discussion

Overall, we observe high UA (84.4 ± 2.0%), PA (85.1 ± 3.4%), and F1 (84.8%) when evaluating all validation sets together. However, we also note large site-specific variations in the accuracies of our product for the different validation sets. Hence, individual validation sets at a local scale were not representative of a region with diverse disturbance patterns, such as Germany. These findings were only possible because we evaluated a larger number of

Table 3. Accuracy and disturbance extent of validation sets, with their 95% confidence intervals grouped by year and with all years combined. User's and producer's accuracies refer only to disturbed forest, while overall accuracy also includes undisturbed forest. Disturbance extent is the sum of all disturbance patches of the respective year. Mapped disturbance refers to the disturbances mapped by our product. Error-adjusted disturbance describes the actual disturbance extent corrected with the validation sets.

Year	Accuracy [%]			Disturbance extent [%]		
	Overall accuracy	User's accuracy	Producer's accuracy	F1-Score	Mapped disturbance	Error-adjusted disturbance
2019	97.4 ± 0.8	75.6 ± 4.1	79.1 ± 11.3	77.3	6.21	5.59 ± 0.14
2020	98.8 ± 0.4	82.0 ± 4.3	77.4 ± 9.4	79.6	2.71	2.98 ± 0.05
2021	99.6 ± 0.1	94.0 ± 1.7	84.6 ± 4.1	89.1	1.80	1.91 ± 0.02
2022	98.4 ± 0.4	78.0 ± 5.0	94.7 ± 4.7	85.5	6.22	5.06 ± 0.08
Combined	99.1 ± 0.1	84.4 ± 2.0	85.1 ± 3.4	84.7	2.86	2.78 ± 0.02

Table 4. Comparison of product accuracies for MMUs of 0.1 and 0.25 ha. Validation set refers to the validated sets of sample points, with the validation year in parentheses. Study area indicates in which of the four study areas displayed in Fig. 1 we performed the validation. The overall accuracy, user's accuracy, producer's accuracy, median size of a disturbance patch, and total disturbed area are compared.

Validation set	Study area	MMU [ha]	Overall accuracy [%]	User's accuracy [%]	Producer's accuracy [%]	F1-Score [%]	Median disturbance patch size [ha]	Total disturbance extent [ha]
Harz (2021)	A	0.1	93.5 ± 2.1	98.8 ± 1.3	81.1 ± 5.4	89.1	0.32	6683 ± 78
		0.25	93.7 ± 2.1	96.9 ± 2.2	82.6 ± 5.4	89.2	0.77	6356 ± 72
Southeast Saxony (2020)	B	0.1	98.2 ± 0.7	84.3 ± 6.8	84.0 ± 10.1	84.1	0.28	2399 ± 63
		0.25	99.9 ± 0.1	97.7 ± 3.2	98.8 ± 2.3	98.2	0.54	1658 ± 42
Baden-Wuerttemberg (2021)	C	0.1	99.9 ± 0.0	59.4 ± 13.0	81.7 ± 12.9	68.8	0.20	641 ± 36
		0.25	99.8 ± 0.1	73.5 ± 9.9	100.0 ± 0.0	84.7	0.44	470 ± 29
Franconian Forest (2020)	D	0.1	99.7 ± 0.1	87.2 ± 8.8	87.9 ± 8.9	87.5	0.18	252 ± 20
		0.25	99.9 ± 0.1	91.1 ± 5.9	97.2 ± 3.9	94.1	0.40	186 ± 16
Franconian Forest (2021)	D	0.1	99.9 ± 0.1	96.6 ± 4.1	95.7 ± 4.7	96.1	0.24	1990 ± 49
		0.25	99.9 ± 0.1	96.6 ± 3.8	95.5 ± 4.3	96.0	0.49	1830 ± 41

validation sets in different regions and with different disturbance types, patch sizes, and disturbance severity. We cover the most frequent causes of disturbance in Germany including windthrow, bark beetle infestation, and complex disease (Bolte *et al.*, 2022). The study areas also include disturbances impacting the four main tree species in Germany: Norway spruce, Scots pine, European beech, and oak (Thuenen-Institute, 2012). Together with the large number of samples considered, we are therefore confident that the accuracies we present are reliable estimates for most German forests.

However, we could not cover all forested regions within our study areas, mostly due to a lack of high-quality reference imagery. Also, neither the high mountain areas of the European Alps, at Germany's southern border, nor the flat lowlands of Germany's northern states were part of the FNEWs study areas. Furthermore, our sample sizes are not balanced regarding the causes of disturbance. Bark beetle infestation, windthrow, and indeterminate disturbances make up the vast majority of both our training and validation datasets. Complex disturbance and gypsy moth infestation are present in larger numbers only in one validation set. On a local scale, the accuracy of our product may therefore differ from the summarized accuracy of all our validation sets. In addition, we originally chose the FNEWs project study areas based on previously known forest disturbances and these sites thus represent forest regions that were more severely affected by disturbances than the nationwide average.

Applying an MMU of 0.1 ha to our disturbance detection helps to reduce noise and false positive detection especially near forest edges where shadows cause variations in the spectral reflectance of pixels. The noticeably lower accuracy we observed for smaller disturbance patches, like in study area (C), suggests that the MMU is necessary as smaller disturbances are not detected reliably. Scattered small-scale disturbances are underestimated, as they rarely reach the required 10 adjacent pixels to be included in our product. Damages due to drought stress may affect only individual trees, which our product does not detect. While these small-scale disturbances are not the focus of our product, underestimating scattered disturbances does limit the application for forest practitioners requiring more detailed disturbance detection.

OA is lowest in the buffer stratum with 84.8%, whereas values were 93.4% in the disturbed stratum and 99.7% in the undisturbed stratum. This confirms our expectation that the product's accuracy is lowest for the buffer stratum and it highlights the importance of the buffer stratum when evaluating a forest disturbance product. Of the 11 019 sample points, 10.2% could not be evaluated. This was the case where the size or extent of a disturbance patch was unclear, the reference data was insufficient (e.g. clouds in imagery), or the forest map was inaccurate. The forest map that we apply to define the forest area in which we detect the changes has an OA of 96.5% (Langner *et al.*, 2022). We observe errors in the forest map most frequently near (forest) roads or where the extent of a prior disturbance is not detected perfectly. As these points were not evaluated further (neither correct nor incorrect), our accuracy metrics and error-adjusted area estimations are unaffected by errors in the forest mask. However, mapped disturbances and pixel-based area estimations

may have included these regions and suffered from these masking errors. This highlights the importance of a reliable forest map for all disturbance studies.

Regional variations

OA tends to be lower in the validation sets with higher disturbance proportions, such as Harz (2019, 2020 and 2022). This is to a large degree due to the fact that undisturbed forest areas are classified with higher certainty than areas with disturbance. Due to the weighting by stratum size, a higher percentage of undisturbed forest results in a higher OA. Consequently, OA is mostly a representation of the accuracy in the undisturbed forest and is by itself a poor metric to evaluate the product's accuracy in detecting disturbance. Together with the high accuracy in the disturbed stratum, the very high OA suggests, however, that only a few completely incorrect disturbance patches were detected. UA and PA of the disturbed class provide better information than OA for assessing how accurately disturbance was detected. Validation sets with many small disturbance patches, scattered disturbance, and a small total number of disturbed areas tend to have a lower UA and PA. We observe this particularly in study area (C) for the Black Forest (2020) and Baar (2021) validation sets. Small disturbance patches were almost entirely located within the buffer stratum, where the accuracy is lowest. The accuracies of validation sets with many small disturbance patches and a larger buffer stratum area percentage were therefore lower.

The lowest UA (31.0%) corresponds to the Lower Franconia (2019) validation set. It consisted of disturbances from gypsy moth infestation and scattered complex disturbance (Fig. 10C). The low UA indicated a greatly overestimated disturbance extent. Reference imagery showed that the gypsy moth infestation peaked at the end of June leading to widespread defoliation. By the end of August, at the reference date of the annual product, the affected forest stands were largely foliated again. Due to the still reduced vitality of the forest, the product still identified large disturbed areas. This example also illustrates the limits of the annual assessment of our monitoring system. In this case, the filtered imagery used to create the product still reflected a more severe state of disturbance than what existed in real time, because of a temporal lag of the model adaptation introduced by the time-series filtering process. During the manual validation, these areas were marked as undisturbed by the interpreter, as their damage had only been temporary and peaking in June, but by end of August, the forest stands had largely recovered. This disagreement leads to a very low UA for gypsy moth disturbances. Temporary defoliating disturbances seem to lead to an overestimation of disturbance extent. However, in the long term, this temporary overestimation is unproblematic for our monitoring system, as these temporary disturbances are reintegrated automatically into the forest mask over the following 2 years.

Mapped disturbance extent

Overall, the total mapped disturbance extent provides a good estimate of the actual disturbance extent. This is also supported by the similar value ranges of omission and commission errors. Accurately estimating the total disturbed area is very important for decision makers to assess the extent to which the forest is affected and to distribute the required funds to forest managers. For the validation sets from 2022, we observe a noticeable overestimation of the actual disturbance area, by ~23%. This was likely in part caused by the very dry spring and summer and subsequent drought stress in 2022 (Bissolli *et al.*, 2022). Grasslands responded more quickly to drought stress than forests during the drought year of 2018 (Reinermann *et al.*, 2019). Sentinel-2 pixels at forest edges or in forests with sparse canopy were often covered by both trees and the surrounding landscapes, such as grassland. The rapid response of the surrounding vegetation may have reduced the perceived vegetation vitality in a pixel sufficiently to register a forest disturbance, even if the trees themselves appeared largely unaffected in the reference imagery. Furthermore, during the manual validation, it was often difficult to assess whether damage from drought stress was permanent or temporary. Another likely reason for the observed overestimation is that 2022 imagery was not included as training data during model parameterization.

Comparison to similar studies

There is an increasing number of local and regional studies focusing on forest disturbance mapping and monitoring with satellite data (e.g. Laurin *et al.*, 2021; Dalponte *et al.*, 2022; Francini *et al.*, 2022; Thonfeld *et al.*, 2022). Here, we compare the validation results for our product with those from three disturbance product studies similar to ours in terms of scale, location, methodology, and rigorous validation. Thonfeld *et al.* (2022) used monthly composites of Sentinel-2 and Landsat-8 imagery to assess forest disturbance in Germany between 2018 and 2020, with an MMU of 0.03 ha. Their monthly temporal resolution was therefore higher than that of the annual product presented in the present study. Their validation approach is similar to ours, as they also include a buffer stratum for their random sampling. They achieve a UA of 71%, compared with our 84.4%, and a PA of 92%, compared with our 85.1%. Their UA was much lower than their PA, indicating a noticeable overestimation of the actual disturbance area in their product. It is difficult to assess whether their product accuracies are representative for all of Germany, as they do not specify where the validation samples were located. They also validate only 1538 sample points, compared with our 11 019, and distributed the points differently among the three strata. Furthermore, the width of their buffer stratum is larger than ours and only drawn on the outside of the product shape. A direct comparison of the accuracies is therefore difficult.

The product of Francini *et al.* (2022) shows forest disturbances in Italy between August 2017 and August 2018, with an MMU of 0.05 ha. Their validation approach is very similar to ours, though their buffer stratum extends 10 m farther from disturbed areas. They provide no overall accuracy values but instead present the accuracies for their three strata, based on 19 300 sample points overall. They reported a UA of 94.2% for the disturbed stratum, compared to our 93.4% for the same stratum. For the buffer stratum, their product achieves a UA of 80.2% and a PA of 46.6%, suggesting that it greatly underestimated the disturbance in that stratum. In the same stratum, our product achieved a UA of 78.2% and a PA of 89.7%, indicating that it overestimated the disturbed area slightly. As with Thonfeld *et al.* (2022), a direct comparison of our results to the accuracies reported by Francini *et al.* (2022) is not meaningful, as decisive parameters were chosen differently or are not documented in detail in their publication.

The open-source python package 'Fordead' by Dutrieux *et al.* (2021a) was developed as a consequence of the intense bark beetle crisis in spruce trees in northeastern France from 2018 to 2020. A validation was only conducted for bark beetle damage on coniferous stands, utilizing 619 classified reference areas from the BDforet dataset (Dutrieux *et al.*, 2021b). The observed frequencies of the three classes (healthy, damaged, and cut) did not align proportionally with the distribution in the study area, but are based on their occurrence in the dataset. The detection of small-scale anomalies remains unevaluated due to missing reference

areas smaller than 0.5 ha in the dataset. Their PA of damaged area was 85% with a UA of 90.1%. A direct comparison to our product is difficult as they used more classes and validated in much smaller and less diverse forest areas, but their accuracies are similar to ours for validation sets with mostly bark beetle–related disturbances.

The documentation of our thorough validation process is of great importance. Covering 22% of the stocked forest area of Germany, the size and distribution of our study areas make our summarized accuracies a very reasonable accuracy estimate for all of Germany. Although other studies have involved a smaller MMU, the MMU of this study was chosen consciously to balance a relevant disturbance patch size with a trustworthy accuracy. Thus, our product is a valuable contribution, due to its high accuracy, detailed validation, and readiness for a uniform and continuous monitoring system of Germany's forests.

Conclusion

The forest disturbance product we present here scores high values for UA (84.4 ± 2.0%), PA (85.1 ± 3.4%) and F1 (84.8%) when all 11 019 sample points were evaluated together. The total disturbance area is estimated accurately for most years. The exception was the year 2022, when drought stress may have led to an overestimation of the disturbance area around forest edges. Our emphasis on the validation of this product highlights several issues. First, site-specific conditions, such as the cause of the disturbance, the size of disturbance patches, and the disturbance severity, can lead to large variations in the product's accuracy. It is therefore difficult to assess the extent to which a validation of individual sites is representative of a region with diverse disturbance patterns, such as Germany. Our study also emphasizes the importance of choosing many different and diverse sites to achieve a more representative result for the validation. Second, the optimal MMU depends on the planned further application of the product. A larger MMU will produce better overall accuracies, as small disturbance patches that are generally detected at a lower accuracy will be excluded. However, this comes at the cost of a lower accuracy when estimating the total disturbed area. The smaller patches, between 0.1 and 0.25 ha, represented up to 30.9% of the total disturbed area, which is why we chose a smaller MMU of 0.1 ha for the final annual product. Lastly, it is important to choose a suitable stratification to distribute the sample points. The disturbance area is much smaller than the undisturbed area in most cases, and stratification is necessary to ensure that enough sample points lie within the disturbed area. Furthermore, we recommend a buffer stratum between the disturbed and undisturbed area, as our accuracy was lowest there (OA 84.8 ± 2.0%). The buffer should extend into both the disturbed and the undisturbed area to facilitate assessments of both omission and commission errors.

Considering the large dimensions of forest disturbances, and the more severe weather phenomena expected in the future, automated monitoring systems of forest areas play a central role in their protection and management. Our product is an important step towards making uniform large-scale forest disturbance analysis available to forest managers and decision makers. Furthermore, the validation methodology presented here could potentially be used as a template for future accuracy analyses of large-scale geodata, regardless of its thematic content.

Acknowledgements

We thank the FNEWS team for their tremendous effort to make this monitoring system possible. In particular, we thank Lea Henning for her support concerning the validation data base and Philip Beckschäfer for his feedback. We also thank Dr Melissa Dawes for professional language editing. Furthermore, we thank the editors of Forestry for their help and timely review process and the four reviewers for their time and detailed feedback.

Author contributions

Eike Reinosch (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft), Julian Backa (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft), Petra Adler (Conceptualization, Data curation, Funding acquisition, Resources, Supervision), Janik Deutscher (Conceptualization, Methodology, Project administration, Supervision, Writing—review & editing), Philipp Eisnecker (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Writing—original draft), Karina Hoffmann (Conceptualization, Data curation, Funding acquisition, Resources, Supervision), Niklas Langner (Conceptualization, Data curation, Methodology, Project administration, Visualization, Writing—original draft), Martin Puhm (Conceptualization, Data curation, Investigation, Methodology, Resources, Software, Visualization, Writing—original draft), Marius Rüetschi (Conceptualization, Writing—original draft), Christoph Straub (Conceptualization, Data curation, Funding acquisition, Supervision), Lars Waser (Conceptualization, Supervision, Writing—original draft), Jens Wiesehahn (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing—original draft), and Katja Oehmichen (Conceptualization, Funding acquisition, Project administration, Supervision, Writing—original draft).

Supplementary data

Supplementary data are available at *Forestry* online.

Conflict of interest statement

None declared.

Funding

This work was supported by funds from the German Federal Ministry of Food and Agriculture (BMEL), based on a decision made by the Parliament of the Federal Republic of Germany via the Fachagentur Nachwachsende Rohstoffe e. V. (FNR) under the funding code 2220NR013.

Data availability

The training polygons are available at https://atlas.thuenen.de/layers/geonode_data:geonode:referenzdaten_fnews_3_0 and the disturbance product for the four study areas is available at https://atlas.thuenen.de/layers/geonode_data:geonode:fnews_jp_18_22. The sample points we used to validate the disturbance product are available from Thünen Institute upon reasonable request at katja.oehmichen@thuenen.de.

References

Arévalo, P., Olofsson, P. and Woodcock, C.E. 2020. Continuous monitoring of land change activities and post-disturbance dynamics

from Landsat time series: a test methodology for REDD+ reporting. *Remote Sens Environ* **238**, 111051. https://doi.org/10.1016/j.rse.2019.01.013.

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

Bissolli, P., Janssen, J., Ziese, M., Imbery, F., Friedrich, K., Paxian, A.. *et al.* 2022. *Trockenheit in Europa 2022*. Deutscher Wetter Dienst. https://www.dwd.de/DE/leistungen/besondereereignisse/duerre/download_tabelle.html, accessed: 11-07-2023.

BMEL (Bundesministerium für Ernährung und Landwirtschaft – Federal Ministry of Food and Agriculture) 2022. *Massive Schäden – Einsatz für die Wälder*. BMEL. https://www.bmel.de/DE/themen/wald/wald-in-deutschland/wald-trockenheit-klimawandel.html, accessed: 22-08-2023.

Bolte, A., Sanders, T. and Wellbrock, N. 2022. *Waldschäden durch Trockenheit und Hitze*. Thünen Institute for Forest Ecosystems. https://www.thuenen.de/de/themenfelder/waelder/forstliches-umweltmonitoring-mehr-als-nur-daten/waldschaeden-durch-trockenheit-und-hitze. accessed: 11-07-2023.

DMJS, B., RJW, B., Gloor, E.. *et al.* 2013. Detecting trends in tree growth: not so simple. *Trends Plant Sci* **18**, 11–17. https://doi.org/10.1016/j.tplants.2012.08.005.

Brooks, E., Yang, Z., Thomas, V.. *et al.* 2017. Dynamic Signaling of changes to forests using exponentially weighted moving average charts. *Forests* **8**, 304. https://doi.org/10.3390/f8090304.

Brooks, E.B., Wynne, R.H., Thomas, V.A.. *et al.* 2014. On-the-fly massively multitemporal change detection using statistical quality control charts and Landsat data. *IEEE Trans Geosci Remote Sens* **52**, 3316–3332. https://doi.org/10.1109/TGRS.2013.2272545.

Bullock, E.L., Healey, S.P., Yang, Z.. *et al.* 2022. Timeliness in forest change monitoring: a new assessment framework demonstrated using Sentinel-1 and a continuous change detection algorithm. *Remote Sens Environ* **276**, 113043. https://doi.org/10.1016/j.rse.2022.113043.

Buras, A., Rammig, A. and Zang, C.S. 2021. The European forest condition monitor: using remotely sensed forest greenness to identify hot spots of forest decline. *Front Plant Sci* **12**, 689220. https://doi.org/10.3389/fpls.2021.689220.

Candotti, A., De Giglio, M., Dubbini, M.. *et al.* 2022. Sentinel-2 based multi-temporal monitoring framework for wind and bark beetle detection and damage mapping. *Remote Sens* **14**, 6105. https://doi.org/10.3390/rs14236105.

Crist, E.P. 1985. A TM Tasseled cap equivalent transformation for reflectance factor data. *Remote Sens Environ* **17**, 301–306. https://doi.org/10.1016/0034-4257(85)90102-6.

Dalponte, M., Solano-Correa, Y.T., Frizzera, L.. *et al.* 2022. Mapping a European spruce bark beetle outbreak using Sentinel-2 remote sensing data. *Remote Sens* **14**, 3135. https://doi.org/10.3390/rs14133135.

Decuyper, M., Chávez, R.O., Lohbeck, M.. *et al.* 2022. Continuous monitoring of forest change dynamics with satellite time series. *Remote Sens Environ* **269**, 112829. https://doi.org/10.1016/j.rse.2021.112829.

Copernicus Land Monitoring Service, European Environment Agency (EEA) 2020. *Forest Type 2018*. Copernicus. https://land.copernicus.eu/pan-european/high-resolution-layers/forests/forest-type-1/status-maps/forest-type-2018. accessed: 14-08-2023.

Doblas, J., Reis, M.S., Belluzzo, A.P.. *et al.* 2022a. DETER-R: an operational near-real time tropical forest disturbance warning system based on Sentinel-1 time series analysis. *Remote Sens* **14**, 3658. https://doi.org/10.3390/rs14153658.

Doblas, J., Lima, L., Mermoz, S.. *et al.* 2022b. Inter-comparison of optical and SAR-based forest disturbance warning systems in the Amazon shows the potential of combined SAR-optical monitoring. *Int J Remote Sens* **44**, 59–77. https://doi.org/10.1080/01431161.2022.2157684.

Dutrieux, R., Féret, J.-B., Ose, K.. *et al.* 2021a. Package Fordead. *Recherche Data Gouv* **V1**. https://doi.org/10.15454/4TEO6H.

Dutrieux, R., Féret, J.B. and Ose, K. 2021b. Mise au point d'une méthode reproductible pour le suivi généralisé des dégats de scolytes par télédétection satellitaire. *ONF Rend Techn* **69–70**, 37–44.

DWD (Deutscher Wetterdienst – German Weather Service) Perennial precipitation 1981–2010. https://opendata.dwd.de/climate_environment/CDC/. accessed: 16-05-2023.

European Space Agency 2021. *Copernicus DEM - Global and European Digital Elevation Model (COP-DEM)*. ESA. https://doi.org/10.5270/ESA-c5d3d65

Fernandez-Carrillo, A., Patočka, Z., Dobrovolný, L.. *et al.* 2020. Monitoring bark beetle forest damage in Central Europe. A remote sensing approach validated with field data. *Remote Sens* **12**, 3634. https://doi.org/10.3390/rs12213634.

Forest Condition Monitor (Waldzustandsmonitor) 2023. http://interaktiv.waldzustandsmonitor.de/. accessed: 10-07-2023.

Forestwatch-DE 2023. https://forestwatch.lup-umwelt.de/app/. accessed: 10-07-2023.

Francini, S., McRoberts, R.E., D'Amico, G.. *et al.* 2022. An open science and open data approach for the statistically robust estimation of forest disturbance areas. *Int J Appl Earth Obs Geoinf* **106**, 102663. https://doi.org/10.1016/j.jag.2021.102663.

Gallaun, H., Schardt, M. and Linser, S. 2007. Remote sensing based forest map of Austria and derived environmental indicators. *Proceed ForestSAT Conf* **2007**, 5–7.

Gao, Y., Skutsch, M., Paneque-Gálvez, J.. *et al.* 2020. Remote sensing of forest degradation: a review. *Environ Res Lett* **15**, 103001. https://doi.org/10.1088/1748-9326/abaad7

Gascon, F., Bouzinac, C., Thépaut, O.. *et al.* 2017. Copernicus sentinel-2A calibration and products validation status. *Remote Sens* **9**, 584. https://doi.org/10.3390/rs9060584.

Giannetti, F., Pecchi, M., Travaglini, D.. *et al.* 2021. Estimating VAIA windstorm damaged Forest area in Italy using time series Sentinel-2 imagery and continuous change detection algorithms. *Forests* **12**, 680. https://doi.org/10.3390/f12060680.

Global Forest Watch 2023. www.globalforestwatch.org. accessed: 10-07-2023.

Hansen, M.C., Potapov, P.V., Moore, R.. *et al.* 2013. High-resolution global maps of 21st-century Forest cover change. *Science* **342**, 850–853. https://doi.org/10.1126/science.1244693.

Harvey, A.C. 1989. Forecasting, structural time series models and the Kalman filter. *Cambridge University Press*, 100–233. https://doi.org/10.1017/cbo9781107049994.

Hirschmugl, M., Gallaun, H., Dees, M.. *et al.* 2017. Methods for mapping forest disturbance and degradation from optical earth observation data: a review. *Curr For Rep* **3**, 32–45. https://doi.org/10.1007/s40725-017-0047-2.

Hoegh-Guldberg, O., Jacob, D., Taylor, M.. *et al.* 2018. Impacts of 1.5°C of global warming on natural and human systems. *Intergov Pan Clim Chan IPCC* **138**, 186–196.

Holzwarth, S., Thonfeld, F., Kacic, P.. *et al.* 2023. Earth-observation-based monitoring of forests in Germany—recent progress and research Frontiers: a review. *Remote Sens* **15**, 4234. https://doi.org/10.3390/rs15174234.

Huang, C., Goward, S.N., Masek, J.G.. *et al.* 2010. An automated approach for reconstructing recent forest disturbance history

using dense Landsat time series stacks. *Remote Sens Environ* **114**, 183–198. https://doi.org/10.1016/j.rse.2009.08.017.

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

Langner, N., Oehmichen, K., Henning, L.. *et al.* 2022. Bestockte Holzbodenkarte 2018. https://doi.org/10.3220/DATA20221205151218.

Laurin, G.V., Francini, S., Luti, T.. *et al.* 2021. Satellite open data to monitor forest damage caused by extreme climate-induced events: a case study of the Vaia storm in northern Italy. *Forestry* **94**, 407–416. https://doi.org/10.1093/forestry/cpaa043.

Louis, J., Debaecker, V., Pflug, B.. *et al.* 2016. Sentinel-2 Sen2Cor: L2A processor for users. *Proceed Liv Plan Sympos* **2016**, 1–8.

Möhring, B., Bitter, A.W., Bub, G.. *et al.* 2021. Schadenssumme insgesamt 12,7 Mrd. Euro: Abschätzung der ökonomischen Schäden der Extremwetterereignisse der Jahre 2018 bis 2020 in der Forstwirtschaft. *Hol-Zen* **9**, 155–158.

Norman, S.P. and Christie, W.M. 2020. Chapter 7 - satellite-based evidence of forest stress and decline across the conterminous United States for 2016, 2017, and 2018. *Gen Tech Rep SRS-250* 151–166.

Oehmichen, K., Puhm, M., Adler, P.. *et al.* 2024. FNEWs Kalman-gefilterte Sentinel-2 Bilder 2017 bis 2022 (1.0). Zenodo. https://doi.org/10.5281/zenodo.11208078.

Olofsson, P., Foody, G.M., Herold, M.. *et al.* 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sens Environ* **148**, 42–57. https://doi.org/10.1016/j.rse.2014.02.015.

Olofsson, P., Arévalo, P., Espejo, A.B.. *et al.* 2020. Mitigating the effects of omission errors on area and area change estimates. *Remote Sens Environ* **236**, 111492. https://doi.org/10.1016/j.rse.2019.111492.

Orthofoto Österreich, 2023. *Geoland.at* http://data.europa.eu/88u/dataset/254757be-69ef-4a6c-a4c1-1432815d7522. (Original work published 2015). accessed: 10-01-2024.

Patacca, M., Lindner, M., Lucas-Borja, M.E.. *et al.* 2022. Significant increase in natural disturbance impacts on European forests since 1950. *Glob Chang Biol* **29**, 1359–1376. https://doi.org/10.1111/gcb.16531.

Puhm, M., Deutscher, J., Hirschmugl, M.. *et al.* 2020. A near real-time method for forest change detection based on a structural time series model and the Kalman filter. *Remote Sens* **12**, 3135. https://doi.org/10.3390/rs12193135.

Qiu, S., Zhu, Z. and He, B. 2019. Fmask 4.0: improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sens Environ* **231**, 111205. https://doi.org/10.1016/j.rse.2019.05.024.

Reiche, J., Mullissa, A., Slagter, B.. *et al.* 2021. Forest disturbance alerts for the Congo Basin using Sentinel-1. *Environ Res Lett* **16**, 024005. https://doi.org/10.1088/1748-9326/abd0a8.

Reinermann, S., Gessner, U., Asam, S.. *et al.* 2019. The effect of droughts on vegetation condition in Germany: an analysis based on two decades of satellite earth observation time series and crop yield statistics. *Remote Sens* **11**, 1783. https://doi.org/10.3390/rs11151783.

Rousseeuw, P.J. 1984. Least median of squares regression. *J Am Stat Assoc* **79**, 871–880. https://doi.org/10.1080/01621459.1984.10477105.

Roy, D.P., Li, Z. and Zhang, H.K. 2017. Adjustment of Sentinel-2 multi-spectral instrument (MSI) red-edge band reflectance to nadir BRDF adjusted reflectance (NBAR) and quantification of red-edge band BRDF effects. *Remote Sens* **9**, 1325. https://doi.org/10.3390/rs9121325.

Rüetschi, M., Small, D. and Waser, L.T. 2019. Rapid detection of windthrows using Sentinel-1 C-band SAR data. *Remote Sens* **11**, 115. https://doi.org/10.3390/rs11020115.

Schiefer, F., Schmidtlein, S., Frick, A.. *et al.* 2023. UAV-based reference data for the prediction of fractional cover of standing deadwood from sentinel time series. *ISPRS Open Jour Photo Rem Sens* **8**, 100034. https://doi.org/10.1016/j.ophoto.2023.100034.

Senf, C. and Seidl, R. 2021a. Mapping the forest disturbance regimes of Europe. *Nat Sustain* **4**, 63–70. https://doi.org/10.1038/s41893-020-00609-y.

Senf, C. and Seidl, R. 2021b. Storm and fire disturbances in Europe: distribution and trends. *Glob Chang Biol* **27**, 3605–3619. https://doi.org/10.1111/gcb.15679.

Shang, R., Zhu, Z., Zhang, J.. *et al.* 2022. Near-real-time monitoring of land disturbance with harmonized Landsats 7–8 and Sentinel-2 data. *Remote Sens Environ* **278**, 113073. https://doi.org/10.1016/j.rse.2022.113073.

Stahl, A.T., Andrus, R., Hicke, J.A.. *et al.* 2023. Automated attribution of forest disturbance types from remote sensing data: a synthesis. *Remote Sens Environ* **285**, 113416. https://doi.org/10.1016/j.rse.2022.113416.

Stehman, S.V. 2014. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. *Int J Remote Sens* **35**, 4923–4939. https://doi.org/10.1080/01431161.2014.930207.

Stehman, S.V. and Foody, G.M. 2019. Key issues in rigorous accuracy assessment of land cover products. *Remote Sens Environ* **231**, 111199. https://doi.org/10.1016/j.rse.2019.05.018.

Thonfeld, F., Gessner, U., Holzwarth, S.. *et al.* 2022. A first assessment of canopy cover loss in Germany's forests after the 2018–2020 drought years. *Remote Sens* **14**, 562. https://doi.org/10.3390/rs14030562.

Thuenen-Institute 2012. *Third Federal Forest Inventory (Bundeswaldinventur) – Results Data Base*, https://bwi.info. accessed: 24-05-2023.

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

Verbesselt, J., Hyndman, R., Newnham, G.. *et al.* 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sens Environ* **114**, 106–115. https://doi.org/10.1016/j.rse.2009.08.014.

Ustin, S.L. and Middleton, E.M. 2021. Current and near-term advances in earth observation for ecological applications. *Ecol Process* **10**, 1–57. https://doi.org/10.1186/s13717-020-00255-4.

Waldmonitor Deutschland 2023. https://waldmonitor-deutschland.de/. accessed: 10-07-2023.

Zhao, K., Wulder, M.A., Hu, T.. *et al.* 2019. Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: a Bayesian ensemble algorithm. *Remote Sens Environ* **232**, 111181. https://doi.org/10.1016/j.rse.2019.04.034.

Zhu, Z. and Woodcock, C.E. 2014. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens Environ* **144**, 152–171. https://doi.org/10.1016/j.rse.2014.01.011.

Zhu, Z., Zhang, J., Yang, Z.. *et al.* 2020. Continuous monitoring of land disturbance based on Landsat time series. *Remote Sens Environ* **238**, 111116. https://doi.org/10.1016/j.rse.2019.03.009.