https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Evaluating the consistency of forest disturbance datasets in continental USA

Laura Eifler[1,2], Franziska Müller[1,2], and Ana Bastos[1,2]

[1]University Leipzig, Talstraße 35, 04103 Leipzig, DE
[2]Max-Planck Institut of Biogeochemistry, Hans-Knöll-Straße 10, 07745 Jena, DE

Correspondence: Laura Eifler l.eifler@studserv.uni-leipzig.de

Abstract. Forests play a crucial role in the Earth System, providing essential ecosystem services and sustaining biological diversity. However, forest ecosystems are increasingly impacted by disturbances, which are often integral to their dynamics but have been exacerbated by climate change. Despite the growing concern, there is currently a lack of globally consistent and temporally continuous data on forest disturbances to characterize changes in disturbance regimes. This gap hinders our ability to accurately assess and respond to these changes.

In this study, we focus on the continental United States and compare four datasets on forest disturbances to evaluate their consistency and reliability regarding their spatial and temporal characteristics and driven agents, when available. Our analysis reveals a moderate agreement across the datasets, with inventory-based comparisons demonstrating the highest level of consistency. In contrast, comparisons involving remote sensing data show lower alignment and a delayed detection of disturbances by satellite observations compared to ground-based inventories. Additionally, discrepancies were observed in the identification of disturbance agents in overlapping areas. Our findings underscore the importance of careful data quality assessment and consideration of their inherent uncertainty when utilizing them for further applications. This study highlights the need for improved data integration and accuracy to advance the understanding of forest disturbances.

1 Introduction

Forests play an important role in the Earth System as they provide ecosystem functioning and services, serve as a carbon sink and support biodiversity (Bonan, 2008; Lindner et al., 2010). These services are sustainable to society and maintain the biological diversity of forests (Thom and Seidl, 2016).

Forest disturbances such as forest fires or insect infestations, are integral parts of forest ecosystems and have significant impacts on forest functioning, structure, composition and dynamics (Turner, 2010). Thom and Seidl (2016) found that the impacts of disturbances on ecosystem services are generally negative across all categories of services. Given their impact on forest productivity, growth, mortality and composition, disturbances can feed-back to climate through changes in forest carbon balance (Bowman et al., 2009; Hicke et al., 2012).

The 2000s have seen an increase in hotter and prolonged droughts due to climate change (Allen et al., 2015; Masson-Delmotte et al., 2021), along with reported increases in climate-driven forest disturbances such as bark beetle outbreaks, storms

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

and fires, impacting forest ecosystems across all forested continents (Seidl et al., 2017; Patacca et al., 2023; Hartmann et al., 2022). Changes in the frequency, size, and severity of these disturbances driven by climate change can alter forest functioning, affect the provision of forest services (Turner, 2010; Meigs et al., 2017; Seidl et al., 2017; Kautz et al., 2017) and threaten forest stability (McDowell et al., 2020; Seidl and Turner, 2022).

Since disturbance impacts can compound in time under more frequent extremes (Bastos et al., 2021), quantifying the climate-change fingerprint on disturbance regime changes is crucial for predicting and mitigating the impacts of future changes on forest ecosystems (Bastos et al., 2023). However we currently lack globally consistent and temporally continuous data over periods sufficiently long to characterize changes in disturbance regimes, especially when it comes to biotic disturbances and wind (Kautz et al., 2017). Multiple datasets record and document the multiple natural and human disturbances, often in the same state/continent, but with different methods. The current evidence is based on sparse and discontinuous data (Kautz et al., 2017; FAO, 2010, 2015, 2020; Hammond et al., 2022), although more comprehensive datasets have been compiled for some regions (Hicke et al., 2012; Patacca et al., 2023; Forzieri et al., 2023; Senf and Seidl, 2021b, a).

Forest disturbance datasets currently available can be grouped into two groups: ground survey and inventory data (Forzieri et al., 2023; Patacca et al., 2021; Forest Service U.S. Department of Agriculture (IDS); Forest Service U.S. Department of Agriculture (FIA)) and remote-sensing based datasets (Hansen et al., 2013; Senf et al., 2020). Existing forest disturbance data differ in the attribution of disturbances, the details in information/records and the acquisition methods. Inventories provide detailed information about the disturbance location, extent, timing and agents, collected through aerial detection and ground surveys, but are sparse in space and time, and suffer from several uncertainties, e.g., due to differences in reporting methods, sampling strategies, human errors, etc. (Hammond et al., 2022; Coleman et al., 2018; Tinkham et al., 2018). Remote-sensing, in turn, offers the possibility to map large regions (up to the globe) in a spatially and temporally consistent manner, however, satellite-based disturbance map accuracy depend on the spatial-resolution of the dataset and on the intensity of the disturbances (McDowell et al., 2015), so that typically only stand-replacing disturbances are considered in large-scale datasets (Senf et al., 2020; Hansen et al., 2013) and attribution to specific agents is limited. Attribution of satellite-based disturbances to specific agents requires, however, high-quality ground data for calibration and validation, underscoring the importance of inventory data in supporting the development of disturbance classification and prediction models.

Characterizing the uncertainty in these different datasets is crucial given the wide range of applications of forest disturbance information, e.g. carbon cycle, forest productivity and growth, forest health, climate (Tinkham et al., 2018), and to develop machine learning models to extrapolate occurrence of different disturbance types across regions (Forzieri et al., 2021). For example, Hicke et al. (2020) used data from the Insect and Disease Survey (IDS) by the Forest Service of the U.S. Department of Agriculture (USDA) to characterize bark beetle outbreaks in the western United States. They highlighted that uncertainties and inconsistencies in the records arise due to variations in surveyor methods, survey locations, and flying conditions. They partially addressed these issues by incorporating specific assumptions into their analytical framework, but they could not account for example for the year-to-year variability. Therefore, further studies are necessary to refine and enhance aerial survey data, in order to improve the accuracy and reliability of the analysis. Coleman et al. (2018) investigated the accuracy of the IDS data by comparing aerial detection survey data with ground survey records and high-resolution satellite imagery

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

(WorldView-3). They utilized error matrices to assess accuracy across different categories. Their findings indicated variable accuracy among damage types and damage agent genera, with bark beetles, the most abundant genera, exhibiting some of the largest errors. Overall, they found that 75% of the damage was mapped correctly. However, the aerial detection survey polygons only overlapped with 44 ($\pm$0.06)% of ground observations.

In this study, we focus on the continental USA (CONUS + AK), a region where multiple datasets are available, to compare the temporal and spatial consistency of four different forest disturbance datasets. Specifically, we compare data from two systematic inventories, the IDS by USDA (Forest Service U.S. Department of Agriculture (IDS)) and the FIA by USDA(Forest Service U.S. Department of Agriculture (FIA)), tree mortality events based on literature review by the International Tree Mortality Network (ITMN, (Hammond et al., 2022)) and tree cover change from the Global Forest Watch (GFC, (Hansen et al., 2013)). Our goal is to quantify the robustness of the information on disturbance year and respective agents, and to identify advantages and potential shortcomings and inaccuracies of these datasets. In doing so, we aim to provide uncertainty estimates for disturbance extent, timing and agent, which can be used to guide the development of classification models in the future. To support these efforts, we compile a new spatially and temporally continuous forest disturbance dataset for the CONUS + AK region based on remote-sensing tree cover loss (Hansen et al., 2013), merged with information on annual disturbance occurrence, agents and hosts, and agreement with ground-data.

2 Data

2.1 Study area

This study focuses on the USA due to the availability of various forest disturbance datasets. Figure 1 presents a selection of data from four different sources used in this paper. The multi-layered nature of the IDS data is clearly visible in panel B.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

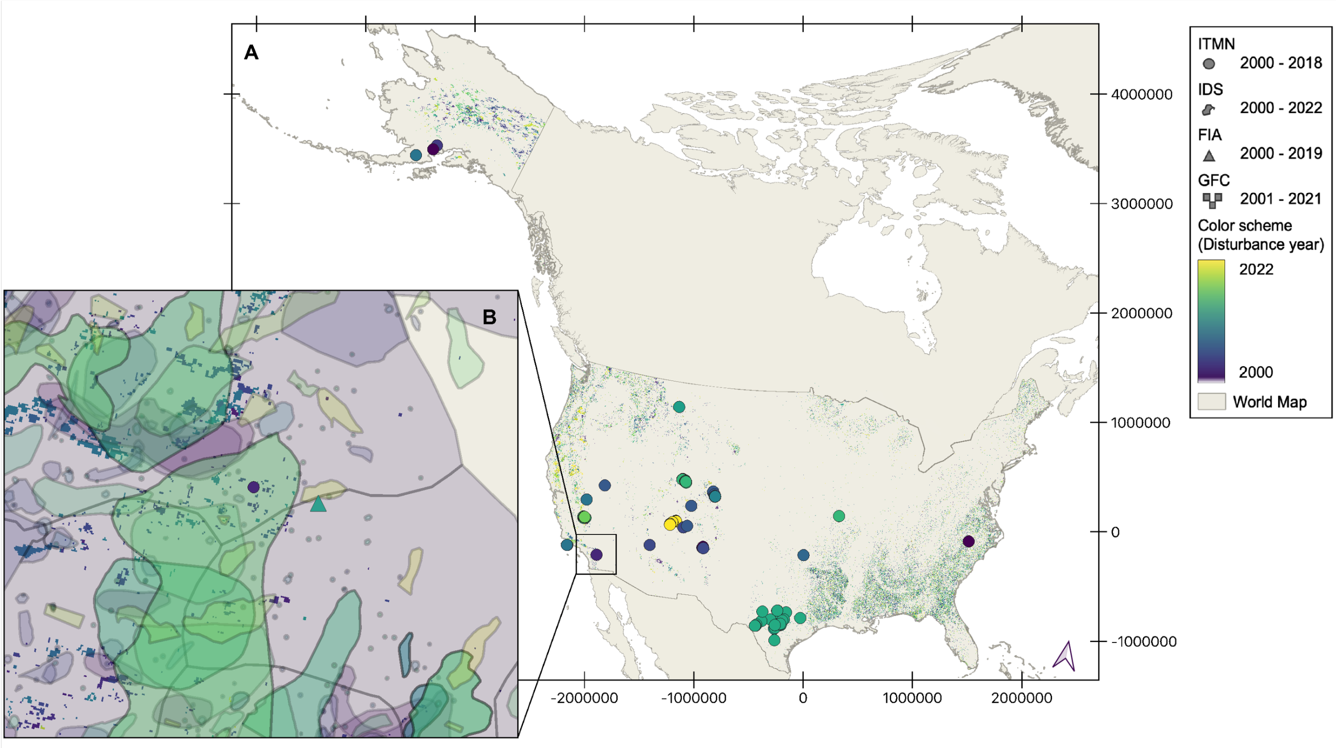

Figure 1. Map of the study area with a selection of data of each dataset. All datasets follow the same color scheme with earlier disturbance events (2000s) in purple and more recent events in green and yellow (2020s). ITMN is depicted as points, FIA as triangle, IDS is shown as polygons and GFC as raster data. Panel A shows the entire CONUS+AK region, and panels B shows the data with increasing detail.

We provide an overview of the datasets, their information, data format and useful additional comments on the data in Figure 1, and describe each in more detail in the sections below.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

	IDS	FIA	ITMN	GFC
Information	Tree mortality and damage in USA from insects and diseases	Tree mortality and damage in the USA on a plot level with further information on e.g. growth, biomass, ownership (to determine the extent, condition, volume, growth and use of trees of timber)	Observations of tree mortality due to drought and heat from literature review	Forest loss from 2000 - 2021, change from forest to non-forest state
Data	Geodatabase with data as grid cells, points and polygons from the insect and disease survey	Database tables as csv files	SuppData 1 and 2 as csv files	GeoTiff files with the information of the lossyear
Data basis	Surveys (aerial and ground) using applications like digital mobile sketch maps (DMSM)	Ground surveys	1303 plots from literature review collection	Results from time series analysis from Landsat 7 and Landsat 8 OLI images
Years	1997 - 2021, updated annually	Annually since 1999, before that on a periodic basis	1970 - 2018, non continuous	2000 - 2021, updated annually
Comments	-	Swapping and fuzzing plots on private ground (explained below)	Information on disturbance agents extracted by reading the papers by the author; Most plots in the Northern Hemisphere and limited to selected publications	-

Table 1. Description of the datasets as a general overview of the information, data and temporal availability.

2.2 Insect and Disease Detection Survey by USDA

The Insect and Disease Detection Survey by the Forest Service U.S. Department of Agriculture (IDS-USDA) is a database on forest damage and mortality due to different disturbance causing agents such as insects, diseases, fire and wind. In the following, we refer to this dataset simply as IDS (Insect and Disease Survey).

In IDS, the continental USA is divided in nine regions, were annual aerial and ground surveys are conducted. In these surveys the data collectors make use of applications like the Digital Mobile Sketch Mapping (DMSM) and the Southern Pine Beetle (SPB) Collector Map. The DMSM is a hardware and software used by the aerial surveyors, in light aircraft, as well as the ground observers to record forest disturbances and their causal agents by sketching aerial signatures of tree injuries and mortality in geo-referenced points or polygons. Additionally, they apply attributes to the points and polygons, such as disturbance agent, tree species, damage type (e.g. defoliation, mortality, discoloration) and severity of the disturbance. Surveyors incorporate different geo-referenced base layers maps into the software (e.g. aerial photographs, topographic maps, road and landscape maps, false color near-infrared based images). This assists the aerial detection efforts, tracks their position in the area and

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

limits duplicates if damage records already exist from previous years. The database contains both vector data as polygons and points of recorded disturbances. In this study, we focus on the polygons data.

2.3 Forest Inventory and Analysis (FIA)

The Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture (USDA) Forest Service Research and Development Branch, started in 1929 when data on the extent, condition, volume, growth and the use of trees in the Nation's forest land was collected on a periodic basis. In 1999 the program moved to an annualized inventory system. Since then, data is collected and compiled in a consistent tabular format, spanning all states and inventories. The current table structure is derived from the National Information Management System (NIMS) to process and store annual inventory data. The data is captured at various levels, including broad information such as location, tree species, invasive species and population, along with more detailed tables that provide specific information on conditions, plot and tree measurements, disturbances, agents, and ground cover (Burrill et al., 2021).

2.3.1 Plot design and location

The data is sampled in plots covering 1-acre area, where the standard plot consists of four 24.0-foot (7.3 m) radius subplots, where trees over a DBH of 5.0 inches (12.7 cm) are measured.

An important consideration is that the plot locations do not align with the precise locations of the plot centers in order to ensure the privacy of landowners and for other practical reasons. To address this, the FIA developed a method called "fuzzing and swapping" to approximate plot coordinates. Fuzzing adjusts plot coordinates to be within 1 mile of the exact location. However, since some landowners possess extensive land areas, plots could still potentially be identified. To enhance privacy, up to 20 percent of private plot coordinates are swapped with those of another similar private plot within the same county. According to FIA, this ensures enough uncertainty for the privacy requirements, but also that county summaries and for example the ownership class of a plot stay the same, since only the coordinates of the plots are swapped. Nevertheless, this introduces inherent uncertainty in the data, which can affect comparisons e.g. with remote-sensing information. In this study we evaluate if the swapping impacts the results and, due to the fuzzing, we compare FIA data with other datasets using buffers of different sizes (800 m, 1600 m).

2.3.2 Sampling procedure

The FIA data collection follows a two-phase sampling scheme (before 2012 a three-phase scheme).

Phase 1 - Stratification: Remote sensing is used to interpret remotely sensed samples and to classify the land into various classes. The purpose is to develop meaningful strata, which are groups of plots that have the same or similar classifications based on remote-sensing imagery.

Phase 2 - Plots are visited or photo-interpreted: The ground plot design is planned so that plots cover a 1-acre sample area, whereas not all trees on the ground are measured. New plots and/ or re-measured plots are visited and recorded.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Phase 3 - in a subset of Phase 2 plots: Additional health indication attributes are collected. Since 2012 Phase 3 is no longer a separate phase, all records are collected in Phase 2.

2.4 International Tree Mortality Network (ITMN)

The International Tree Mortality Network (ITMN) was created to advance the development of methods for combining, harmonizing, and integrating various data repositories and data types to provide comprehensive information on global tree mortality rates. This effort involves both terrestrial and remote sensing data sources. It therefore also relies on the contributions of other scientists to enrich the network with diverse data and expertise. The ITMN first compilation of heat and drought induced tree mortality is based on a literature review of 154 peer-reviewed studies since 1970 and data requests as described by (Hammond et al., 2022). The dataset has been published in a geo-referenced database with records of tree mortality events (International tree mortality network, 2022) and includes 1303 plots of recorded tree mortality from 1970 to 2018 as point data. For each event, the data includes additional details such as the location, the mortality year (sometimes provided as a range), a reference paper, the continent, the number of sites and plots reported in the reference, and biome information. The study notes that the data is inherently biased due to the use of only the available peer-reviewed studies showing drought- and heat-induced tree mortality in the field. Additionally, there is a strong bias to the Northern Hemisphere, which is explained by the fact that the global forested area is 78 % north of the Equator (Hammond et al., 2022). Therefore, some forest types and areas are still under-sampled.

2.5 Global Forest Change (GFC)

The Global Forest Change provides annual maps of global forest change (net, gains and loss) at approximately 30m spatial resolution since 2001 (Hansen et al., 2013). Based on analysis of time series from Landsat 7 and Landsat 8 OLI images, the year of tree loss and tree gain are identified for each pixel, and provided in georeferenced raster files organized in 10×10 degree lat/lon tiles. Their analysis was conducted using Google Earth Engine, beginning with the automated pre-processing of Landsat image data, which included tasks such as resampling, conversion to Top of Atmosphere (TOA) reflectance, quality assessment and image normalization. Following these preprocessing steps, time-series spectral metrics were calculated. The relationship between forest loss and other parameters was established through the use of a decision tree, linking the spectral metrics to the observed changes. Finally, forest loss was disaggregated to annual time scales by identifying the maximum annual decline in percent tree cover and the maximum annual decline in the minimum growing season Normalized Difference Vegetation Index (NDVI). Here, we focus on the loss year maps. The pixel values range from 0 to 21, indicating the year of forest loss, whereas 0 is no change (year 2000 as initial year), therefore the first year with changes can be 2001. Forest loss was defined as a stand-replacement disturbance or the complete removal of tree cover canopy at the Landsat pixel scale (Hansen et al., 2013). The causes for forest loss are wide spread, for example the tropics are predominantly affected by the prevalence of deforestation dynamics, which causes the greatest loss, while in extratropical regions (temperate, boreal), tree cover loss is determined by forestry, fires, logging, diseases and storms, at more moderate rates.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

3 Methods

The following flowchart (Figure 2) shows the steps per dataset from pre-processing to the final tables, which will be further explained below. All calculations were done in Python using the packages pandas, numpy, geopandas, scipy.stats. The code will be made publicly available upon acceptance.

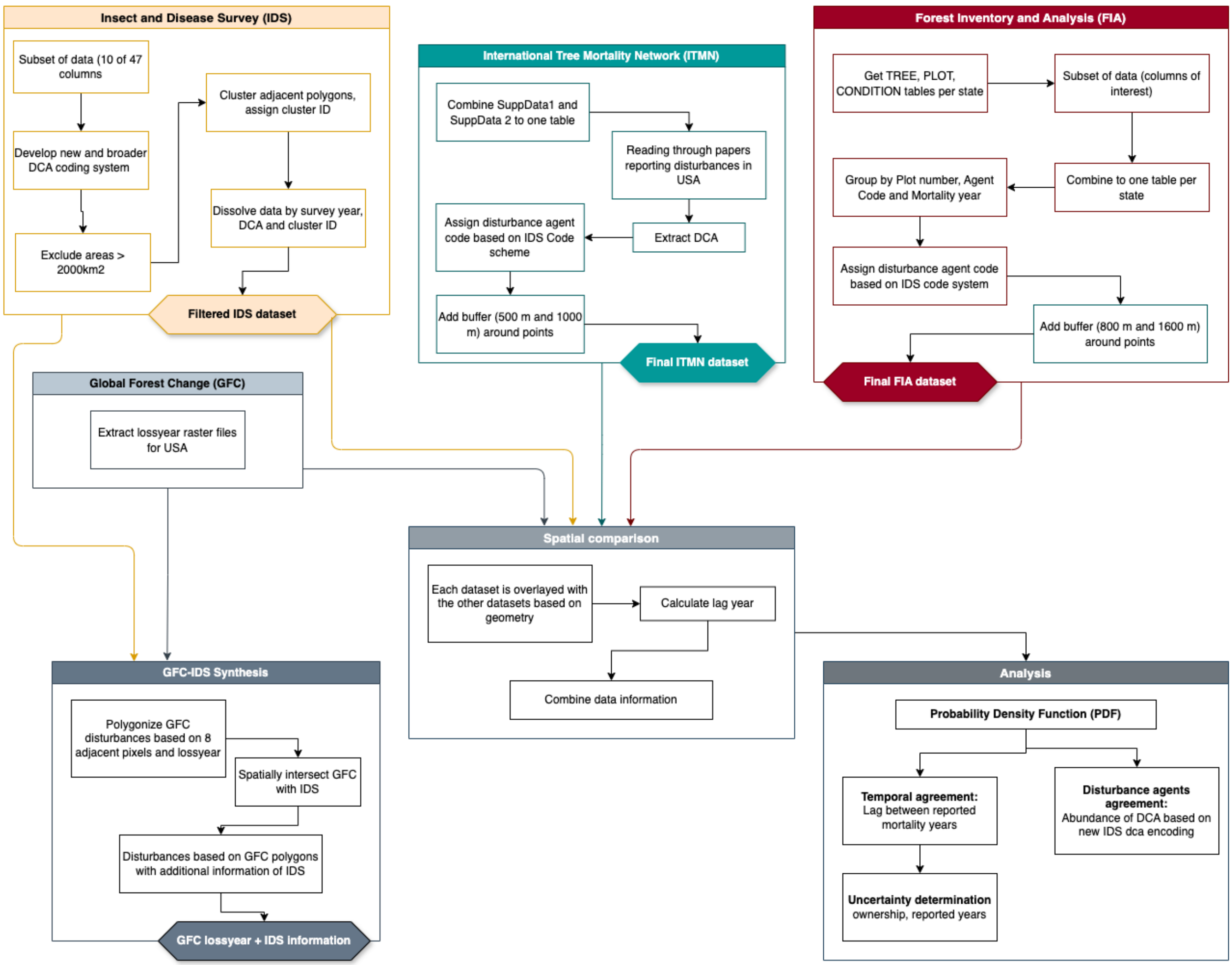

Figure 2. Flowchart of methods applied to each dataset.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

3.1 Data pre-processing

In order to be compared, the datasets had to be pre-processed individually, depending on the type, complexity and detail of the information within each dataset. Given that it is a georeferenced raster dataset with only one layer of information, the GFC dataset did not require any pre-processing.

3.1.1 IDS

The IDS data contain many different levels of information, including disturbance (type, intensity, percent affected), stand (tree species) and measurement details (survey year, month). The raw data table structured each event as a separate polygon with additional layers of information.

At first, the data was reduced from 47 layers to 11 key variables of interest, shown in Table B1. The IDS includes detailed information about the type of disturbance and species, following a disturbance agent (DCA) coding system including 26 different agent categories, see Table B2. The original IDS DCA_CODE is highly detailed, but since other datasets are not as precise in recording disturbance agents (major disturbance types only), we grouped the information from DCA_CODE into ten large disturbance agents (Table 2). We attributed new codes where 0 corresponds to No Data and each disturbance k is assigned a code $Code_k = 2^k$ (Table 2) to account for the fact that multiple agents can be detected in the same year and overlap in space (Figure 1).

Name	No data	Bark beetle	Defoliators	Drought	Fire	Wind	Multi damage	Other biotic	Other abiotic	Other
Code	0	1	2	4	8	16	32	64	128	256

Table 2. Coding system for major disturbance agent groups used in this study.

In some regions, IDS included irregular polygon shapes, encompassing extensive non-forest areas and covering unrealistically large extents. To address this, we calculated the area of each polygon and excluded those exceeding a threshold of 2000 km^2. The data was dissolved based on identical survey years and disturbance agents to create multipolygons recording the same year and agent.

3.1.2 FIA

The FIA data includes information on different levels, like location, tree, plot, invasive species, ground cover and down woody material. Each level contains multiple database tables in csv format per state with detailed information for example about the condition, tree biomass, the plots itself. In this study, the *Tree*, *Plot* and *Condition* tables per state were used. At first a subset of the tables with the columns of interest were generated, the columns of interest are shown in Table B3, Table B4 and Table B5. Then the three tables were merged based on the plot sequence number, which is a unique number to identify a plot record,

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

reported in each database table. The data was subset to only include accessible forest land (`COND_STATUS_CD` == 1). For a better comparison with other datasets, only events with an identified disturbance agent were selected, corresponding to 13% of the absolute number of records (varying from state to state from 4% to 33% of the absolute number of events.). In some cases, multiple records describe the same disturbance event, so that the data was aggregated based on the plot sequence number, the agent and the mortality year. While the FIA data should offer detailed information about disturbance agents, the level of detail varies across states. For that reason, only broad agent groups were assigned to the event in our analysis. Therefore, we mapped the FIA agent coding onto the IDS-USDA coding system, to generate a better comparability with the other datasets, shown in Table B6. Another issue that emerged during the analysis is the use of agent codes (*AGENTCD*) that should correspond to multiples of 10, yet some entries include integers between the full 10 numbers. It is unclear what type of disturbances these codes represent, and if they indeed signify disturbance codes. For the comparison of the agents, the mapping was further refined (Table B7).This ambiguity complicates the accurate identification of disturbance agents. The FIA data is provided as georeferenced point data, but the method of "fuzzing and swapping" due to privacy regulations described in Section 2.3 creates uncertainties to the locations of the plots. Therefore, two buffer sizes were applied - 800 m and 1600 m, representing the 0.5 mile and 1 mile fuzzing radii.

3.1.3 ITMN

The ITMN dataset was compiled from records of tree mortality attributed primarily to drought and heat in the literature and based on data requests (Hammond et al., 2022). However, tree mortality is a complex process, often caused by multiple factors which can, themselves be influenced by drought and heat (Allen et al., 2015). For example, biotic disturbances can benefit from heat and water stress, so that defoliator or bark-beetle outbreaks tend to occur during or following drought events (Allen et al., 2015). Such interactions between disturbances need to be considered when comparing with other datasets. To address this, we reviewed all the publications in the ITMN dataset that documented tree mortality in the USA to identify other possible disturbance agents. Specifically, we searched for additional disturbance agents reported in each original publication, and added this layer of information following our developed IDS coding system Table 2. Our compilation of disturbance agents based on the original literature from ITMN includes main five classes, namely *bark beetle*, *drought*, *bark beetle and drought*, *multi damage* and *drought and other biotic*. Since the ITMN data is point-based, two buffers of 500m and 1000m were applied around the points to account for potential inaccuracies when comparing with other datasets.

3.2 Spatial comparison

First, we examined the spatial overlap between pairs of datasets. Using geopandas.sjoin(), we compared two data tables to generate new tables that included the overlapping disturbance records and the attributes from both datasets. This method was applied to datasets from ITMN, IDS, and FIA.

GFC, imported as raster data, was compared with each individual event ID in each of the other three datasets (FIA and ITMN including the respective buffers, and the polygons in IDS). For spatial comparison, the polygon area from the inventory data served as a mask to crop the raster to only the overlapping pixels (which contain loss year information). We calculated

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

the fraction of GFC tree loss pixels within the disturbance polygons. We selected the two years with the highest overlap fraction (excluding 2000, which represents no loss) and added these to the joined data table. Because multiple *lossyears* occur proportionately in larger polygons, especially from IDS, we selected the two years with the highest fraction.

Using the disturbance/mortality year data from each pair of datasets, we can calculate their respective temporal agreement, given by the difference in detection timing (0 indicating perfect agreement). In the analysis, we focus on a subset of data where the disturbance years fall between 2000 and 2021. For the temporal comparison, we focus only on mortality events, therefore filtered the data to select events reported in IDS with the damage type *mortality*. FIA, GFC and ITMN already only report mortality as type of damage.

3.3 Uncertainty quantification

We then analyse the temporal differences in disturbance date between datasets by fitting Gaussian Probability Density Functions to the time lags across events The mean difference in mortality year detection between the two datasets provides an indication of potential systematic biases, while the standard deviation provides insight into the dispersion or variability in detection timing. Understanding these sources of uncertainty is essential, as it may prompt further investigation to enhance the reliability and consistency of the data.

3.4 Synthesis IDS and GFC

To evaluate the spatial consistency between the IDS and GFC datasets, we overlayed the GFC *lossyear* data on IDS polygons, by considering the patches of GFC with the same value and applying a buffer of the adjacent pixels of 30m in each direction. We considered continuous patches of GFC with the same tree loss year as a single disturbance. If a GFC patch spatially overlapped with the IDS data, we quantified the fraction of agreement between the two patches. In that case, we assumed that both datasets recorded the same disturbance, and added the attributes of the IDS dataset to the GFC patch and converting it into a GeoDataFrame. The new GeoDataFrame contains the geometry of the GFC disturbances as well as information from the IDS dataset, including the DCA and Host attributes, as well as metrics of agreement between the datasets, such as the respective temporal lag with IDS as a metric of temporal uncertainty. The data will be made publicly available as geopackage upon acceptance.

4 Results

4.1 Spatial agreement

In the study area of CONUS+AK, the ITMN dataset comprises a total of 388 forest plots that have recorded mortality events since the year 2000. By comparison, the FIA dataset includes 107318 forest plots with identified mortality agents covered in the same period. Additionally, the IDS contains an extensive 4033890 records of all reported forest damages since 2000, whereas 3234868 records report mortality as damage.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Table 3 indicates the total numbers of overlapping events when considering the smaller buffer sizes of 800m and 500m, for FIA and ITMN respectively. Since IDS can report multiple disturbances within the same area, e.g., if both drought and insects have occurred in a given place, the number of overlapping events will vary.

The FIA show the largest number of overlapping events with GFC (85105), followed by IDS (69538). In contrast, only 34 points overlap with six unique ITMN events, due to the point-scale nature of these datasets. Given the small overlap between FIA and ITMN, these datasets were not compared in the rest of the analysis. However, a larger sample of ITMN events overlap with IDS (126) and GFC (229), allowing for further analysis.

Both IDS and GFC report larger patches, so that we compare the mean proportion of overlapping GFC area with IDS disturbances. We find a median value of 10.8% overlap between GFC tree cover loss patches and IDS, with values ranging widely—from minimal overlap just above 0% to a maximum of 99.9%, and with an interquartile range of 1.9% to 37.1% (25–75 percentiles, respectively).

	FIA	IDS	GFC	ITMN
FIA	-	69538	85105	34
IDS	989166	-	10.8% (Q1 1.9, Q3 37.1)	1726
ITMN	6	126	229	-

Table 3. Total number of overlapping disturbance events between the dataset pairs. The comparison of IDS with GFC is given as the median value and the 25% (Q1) and 75% (Q3) quartile of the proportion of GFC area overlapping with IDS disturbance area per disturbance event.

4.2 Temporal agreement

We first compared pairs of datasets in terms of their temporal agreement of the recorded mortality/ disturbance year over all regions of the USA (Figure 3). For this, we analysed the distributions of the difference in disturbance timing between the two point-based datasets (FIA and ITMN) with the two spatially explicit datasets (IDS and GFC), as well as IDS with GFC.

All distributions show a shift of the mean towards negative values, which indicate that the point-based datasets (i.e., those based on ground surveys, FIA in red and ITMN in blue), tend to record disturbances earlier than IDS or GFC. Furthermore, the shift to the left in the comparison between the two spatially explicit datasets (yellow line) indicates that IDS tends to report disturbances earlier than GFC.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

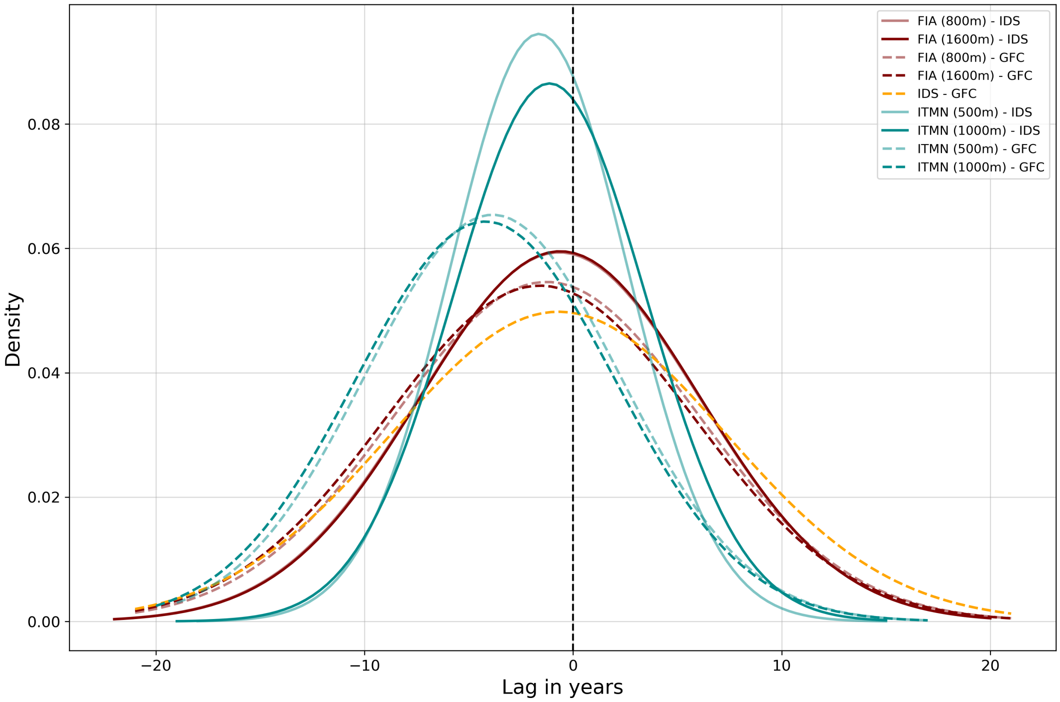

Figure 3. Probability density function of differences in mortality years across the four different the forest disturbance datasets in the USA. Comparisons with the GFC data are represented as dashed lines, comparisons with IDS as solid lines. Negative values indicate a earlier disturbance detection of the dataset which stands first, compared to the second one.

Datasets	Mean [years]	Standard deviation [years]
FIA (800m) - IDS	-0.7	6.7
FIA (1600m) - IDS	-0.6	6.7
FIA (800m) - GFC	-1.3	7.3
FIA (1600m) - GFC	-1.6	7.4
ITMN (500m) - IDS	-1.6	4.2
ITMN (1000m) - IDS	-1.1	4.6
ITMN (500m) - GFC	-3.9	6.1
ITMN (1000m) - GFC	-4.2	6.2
IDS - GFC	-0.7	8.0

Table 4. Mean and standard deviation values of the differences between pairs of datasets. For the point datasets (FIA, ITMN) two different buffer sizes have been considered to evaluate the role of uncertainty in the reported coordinates.

For the comparison of IDS and GFC with FIA, the distribution of the timing differences peak slightly negatively, with mean values between -0.6 (FIA-IDS, 1600m buffer) and -1.6 (FIA-GFC, 1600m buffer) years, and large standard deviation (ca. 7 years for both datasets), indicating that FIA records earlier than the compared data, but with a considerable variability. The best agreement is found for FIA with IDS with a buffer of 1600m, with a mean difference of -0.6 years and a standard deviation of 6.7 years. The consistency of FIA and GFC is lower, with a mean of -1.3 years for a 800 m of FIA (-1.6 years for 1600 m buffer). Additionally, the distribution of the differences is broader than for IDS, with standard deviation of 7.3 and 7.4 years for 800m and 1600m buffer, respectively. Using longer buffer length results in slightly better agreement for IDS but not for GFC (-1.3 vs. -1.5 years mean difference between 800m and 1600m), but does not affect the standard deviation of the differences.

We find the best agreement in timing of disturbance between ITMN and IDS, which show the narrowest curve of differences in reported disturbance timing, but also peaking in the negative range, with mean differences of -1.6 years for 500 m buffer to -1.1 years with 1000 m buffer. Contrary to the comparisons with FIA, the standard deviation of ITMN-IDS differences is smaller with the 500 m buffer than with the larger 1000m buffer, 4.2 years and 4.6 years, respectively.

The comparison of ITMN and GFC shows the largest differences of all datasets, with an average of -3.9 years for 500 m buffer and -4.2 years for 1000 m buffer, and standard deviation of 6.1 (500 m) and 6.2 (1000 m) years. The comparison of IDS with GFC shows similar differences as between FIA and GFC, with slightly negative mean differences of -0.7 years, but a slightly larger spread of the differences compared with other datasets, with standard deviation of 8 years.

Overall, most datasets show mean differences of circa 1 year and standard deviation of around 7 years, excepting ITMN-IDS comparison. However, we note than ITMN has a small sample size in the CONUS+AK region, so that the results might not be as robust as for the other datasets.

4.3 Disturbance agents

Next, we compare the agreement between datasets in terms of their reported disturbance agent. Given that GFC does not report cause of tree loss, it is not considered in this analysis. Therefore, we compared IDS with ITMN (Figure 4) and with FIA (Figure 5), using the smaller buffer size for ITMN and FIA, 500 m for ITMN and 800 m for FIA respectively, since larger distances do not significantly change the agreement between datasets (Section 4.2). After filtering the data to retain only overlapping events where both datasets reported agents, excluding zero and NaN values, the resulting dataset contained 1729 events for ITMN and IDS, and 989109 events for FIA and IDS.

We find low consistency between IDS and the other two datasets regarding the disturbance agents. The IDS dataset is primarily dominated by bark beetle-related disturbances, including mortality events only, whereas the other datasets mainly report *other agents*. For the events overlapping between IDS and ITMN, ITMN tends to attribute mortality to drought combined with other biotic disturbances, followed by drought combined with bark-beetle, whereas IDS predominantly reports bark-beetle only, followed by multi-damage and drought only. Both datasets report few events driven my multiple damages, which include biotic and other abiotic disturbances. Very few events coinciding with mortality reported in ITMN are identified as driven by defoliator (4), fire (3) disturbances by IDS.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

The majority of disturbances reported by the FIA are attribute to *other biotic* class, followed by *other abiotic*, *other* and *fire*. For the disturbances reported in the FIA dataset, the overlapping disturbances in IDS predominantly attributes these to bark-beetle, followed by defoliators, other biotic, multi-damage. Some events are further attributed to abiotic disturbances such as wind, drought, fire and other abiotic.

Based on the IDS events that overlap with the other datasets, bark beetle and multi-damage disturbances tend to be mostly associated with tree mortality, while disturbances caused by defoliators, wind and other biotic events typically lead to non-mortality damages.

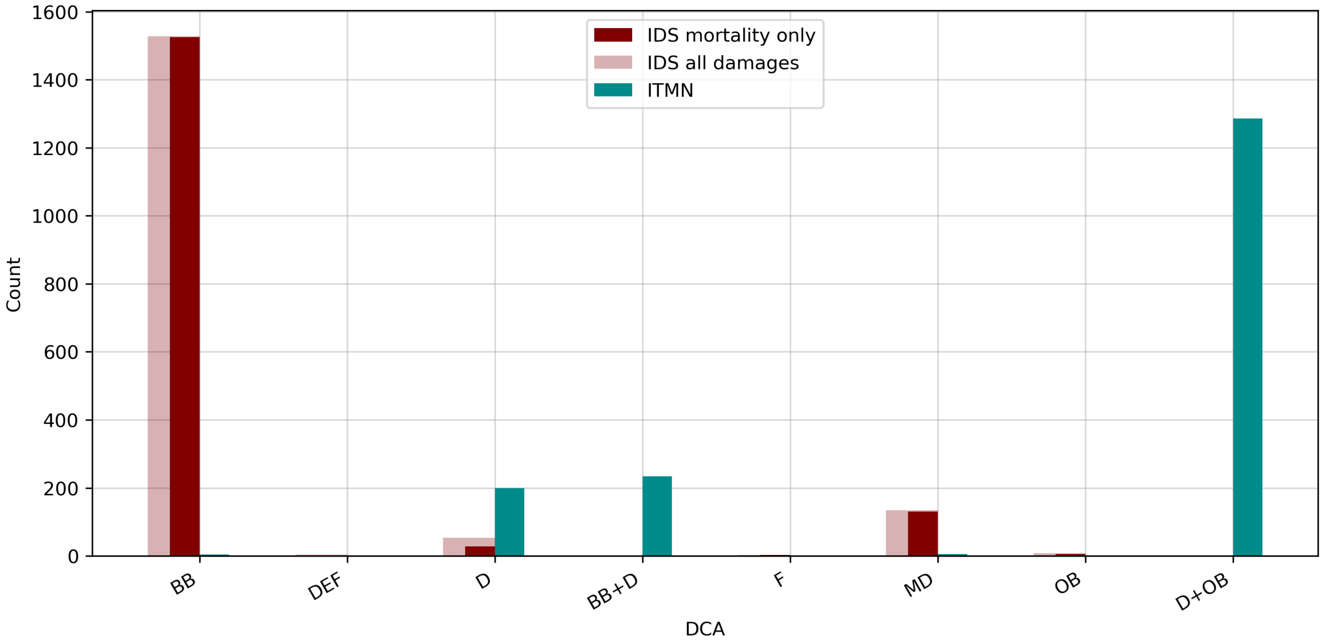

Figure 4. Count of disturbances across different disturbance agents for overlapping events of the comparison of IDS and ITMN and. The IDS data was further separated in *all damages* and *mortality only* events. In 98% of the overlapping disturbances, IDS classified as mortality events.

4.4 Sources of uncertainty

We have found large spread between datasets in terms of reported timing of disturbances and disagreements in reported agents. This indicates that these datasets are still prone to considerable uncertainties, which might be explained by differences in detection and reporting methods, as well as limitations of the accuracy of each approach. We therefore analyse potential sources of uncertainty based on the information provided by the different datasets.

For example, FIA reporting depends on the ownership of the inventory plots (national vs. private forests), since the data is more fuzzed and swapped according to the guidelines of the FIA. The fuzzing approach to protect data of private owners is very

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

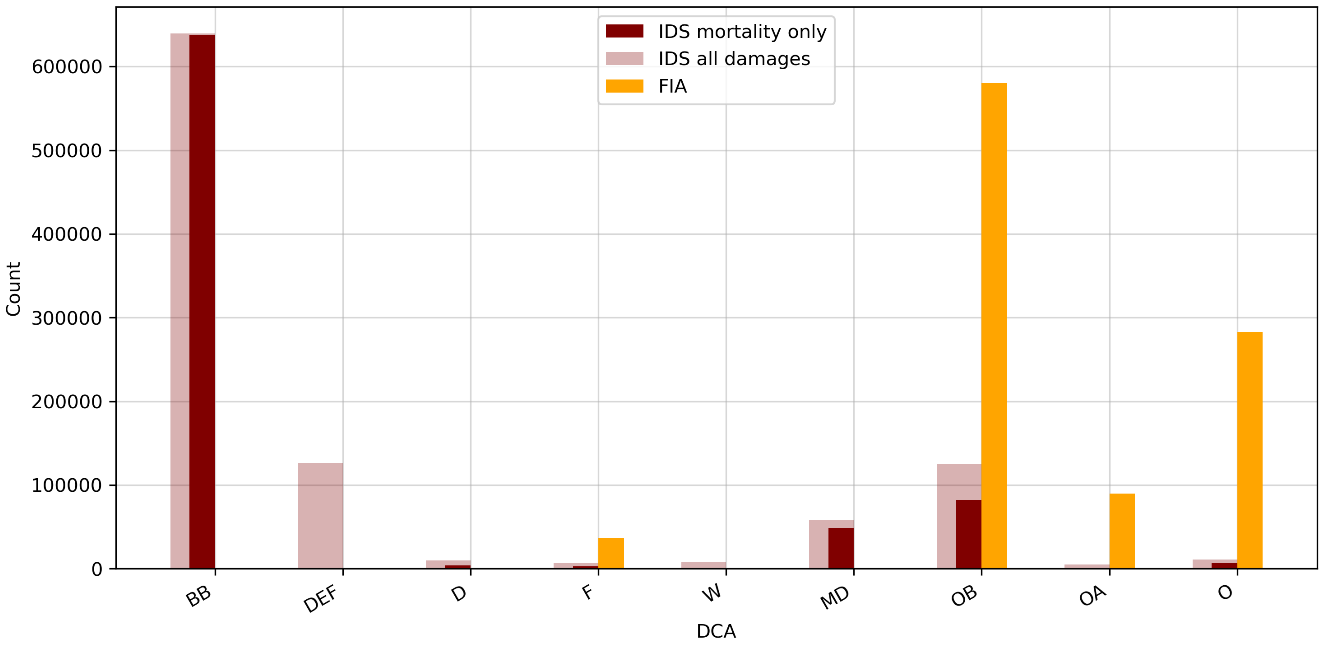

Figure 5. Count of disturbances across different disturbance agents for overlapping events of the comparison of IDS and FIA. The IDS data was further separated in *all damages* and *mortality only* events. In 79% events led to mortality.

likely to contribute to some of the uncertainties we found. Furthermore, the timing of mortality and/or disturbance is likely to be highly uncertain in all datasets because they either depend on visual inspection (ITMN, FIA, IDS) that is not continuous in time, or on time-series analysis that require specific thresholds to report forest loss (GFC).

The FIA data provides the information of plot ownership and of mortality, measure and inventory year, so that the contribution of these uncertainties to the overall discrepancies between datasets can be analysed.

4.4.1 Ownership

First, we analyse the effect of data collection in privately-owned vs. public forests on the agreement between FIA and IDS and GFC (Figure 6 and Table 5).

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

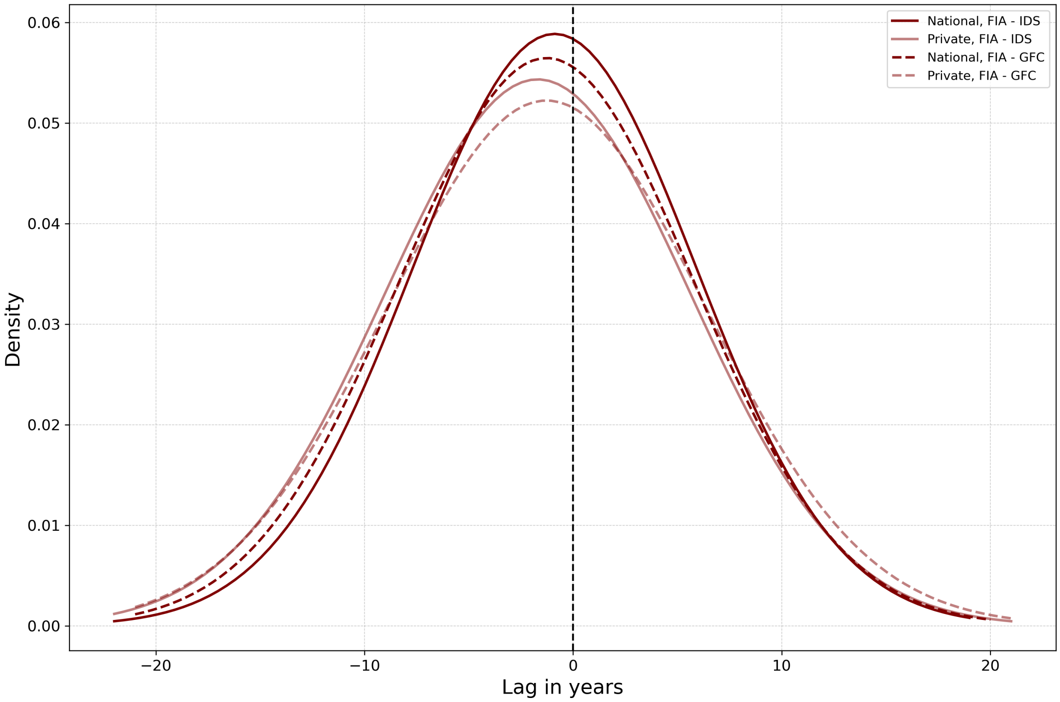

Figure 6. Probability density function of differences in mortality years separated in national (dark) and private (light) forest plots from FIA, compared with IDS (solid) and GFC (dashed). The comparison was made using the 800m buffer around FIA plots.

In the FIA events overlapping with IDS, 153987 events (16%) occurred in privately owned plots, while 835179 events (84%) occurred in public plots. We find that national forests tend to show smaller differences in the reported timing of disturbance, than privately owned forests, with mean differences of -0.9 years and -1.7 years respectively. However, we find a better agreement overall when all data is included. Furthermore, we find a larger spread in the differences of event timing in private forest plots than over national ones, with standard deviation of the differences of 6.8 years and 7.3 years respectively. However, these values are also higher than for all events combined. The data is unevenly distributed in national and private plots, leading to sub-group specific characteristics and larger differences in timing compared to the results of all data.

In the FIA events overlapping with tree loss by GFC, 50358 (59%) occurred in privately owned plots, while 34747 (41%) occurred in public plots. In the case of GFC, the mean differences in the reported timing are very similar for both privately owned (1.3) and public forest plots (1.3 years), being between the values found for the FIA-IDS comparison. The spread in the temporal differences is slightly smaller for public plots (7.1 years) than for private ones (7.6 years), and always larger for the FIA-GFC comparison than the FIA-IDS comparison for each ownership group.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Ownership	Datasets	Mean [years]	Standard deviation [years]
National	FIA - IDS	-0.9	6.8
Private	FIA - IDS	-1.7	7.3
National	FIA - GFC	-1.3	7.1
Private	FIA - GFC	-1.3	7.6

Table 5. Mean and standard deviation values of the differences between FIA and IDS or GFC, when considering the two groups with different ownership status.

4.4.2 Reported versus measurement timing

In addition to the mortality year (MORTYR - the estimated year a remeasured tree died or was cut), FIA reports the inventory (INVYR) and measurement (MEASYEAR) year for each event. The inventory year corresponds to the best representation of the year where the inventory data was collected, while the measurement year corresponds to the year where the plot was completed. We first group FIA events into those in which the inventory (INVYR) and measurement (MEASYEAR) year coincide with the mortality year, and events where these do not coincide. For these groups, we then compare the differences in disturbance timing between FIA and IDS and GFC (Figure 7). In the FIA dataset, 1-3% of the disturbance events analysed have coinciding disturbance year and inventory or measurement year, while for 97–99% of the events analysed, the disturbance year reported is different from the inventory or measurement year. For both groups, the disturbance year tends to be reported 1 to 20 years before the inventory/measurement year. Already 98% of disturbance events are within a range of 1 to 10 years that mortality is earlier than the inventory or measurement took place. On the other hand, there are a few cases where the mortality was described after the inventory or measurement year (1–5 years/ 1–7 years for inventory/ measurement year).

Disturbances recorded by FIA in the same year of the disturbance occurrence show a shift towards positive differences, with the mean values ranging between 1.1 and 1.8 years (IDS/GFC report earlier than FIA), while those in which the inventory or reporting year are different from disturbance year tend to show negative differences, with mean values of -1.1 years, when compared with IDS, and -1.4 years, when compared with GFC. While the differences in reported vs. inventory/measurement year influence the predominant sign of the differences, the two groups show similar values of standard deviation of differences of 6.8–7.5 years, similar to the previous analyses.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

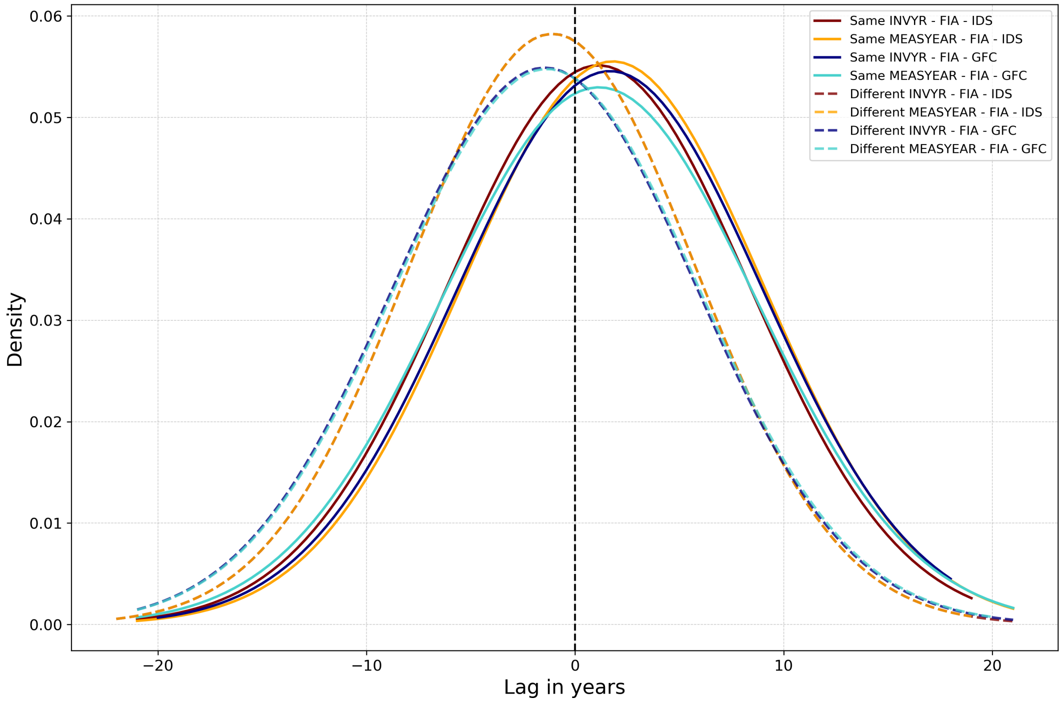

Figure 7. Probability density function of mortality lag years between the comparisons of FIA with IDS (solid) and GFC (dashed). They are further subdivided based on the correspondence between the inventory and measurement year with the mortality year. Same year records are depicted in darker colors, different year records in lighter colors.

Type of year record	Datasets	Mean [years]	Standard deviation [years]
Same INVYR	FIA - IDS	1.1	7.2
Same MEASYEAR	FIA - IDS	1.8	7.1
Same INVYR	FIA - GFC	1.7	7.3
Same MEASYEAR	FIA - GFC	1.1	7.5
Different INVYR	FIA - IDS	-1.1	6.8
Different MEASYEAR	FIA - IDS	-1.1	6.8
Different INVYR	FIA - GFC	-1.4	7.3
Different MEASYEAR	FIA - GFC	-1.4	7.3

Table 6. Mean and standard deviation values of the differences between FIA and IDS or GFC, when considering the groups of events with different reported vs. inventory or measurement years.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

5 Discussion

Forest disturbances have multiple causes and affect forests in varied ways. Wind and bark beetles primarily impact the canopies, whereas fire and drought directly affect tree regeneration, reducing their capacity for recovery and reorganization. Additionally, environmental conditions such as hotter and drier climates can weaken trees' defenses against pathogens and insects, exacerbating the damage caused by disturbances (Anderegg et al., 2015). The type of disturbance and the nature of its impact on trees thus influences our ability to detect and characterize its dynamics.

Here, we compared four different datasets that report timing, location/extent and cause of forest disturbances. Overall, we found variable agreement in terms of disturbance timing and agent across the different datasets. These variations can be explained by inherent uncertainties in the detection methods, spatial and temporal scales, and the level of detail in disturbance records of each dataset.

5.1 Methodological uncertainty

Both FIA and IDS are used for systematic monitoring forest disturbances, but they differ in scale and methodology. The FIA is distributed across the whole CONUS+AK region, but is based on small (<10m) plots which are revisited regularly, typically every 5 or 10 years. Additionally, the swapping and fuzzing approach for data protection of privately owned forests is expected to introduce uncertainty about the exact location and timing of disturbances. Aerial surveys in IDS provide large-scale information as disturbance polygons, but typically focus on regions where outbreaks have been known to occur. Although effective in capturing major disturbance and mortality events, the IDS may miss disturbances in remote, unmanaged forests, and urban areas (Kautz et al., 2017). Furthermore, the IDS data includes hand-drawn disturbance polygons, which introduces subjectivity to their extent. This is particularly evident in USDA Regions 8 and 9, where some polygons exhibit uncharacteristic shapes, spanning vast regions and even covering urban areas (see Figure A1). These inaccuracies affect spatial comparisons and analyses, potentially leading to erroneous results.

The ITMN data is recognized to be inherently biased due to the use of literature-based reports of tree mortality in the field (Hammond et al., 2022), which results in certain regions and forest types being underrepresented in the dataset. Furthermore, the point data format inherently carries uncertainties regarding the exact location and extent of mortality events. For example, the scale of the mortality events reported in ITMN dataset varies considerably, from isolated points to large clusters of points. Moreover, the spatial and temporal accuracy of the coordinates of each event depends on the method used in each original study.

Satellite data allows for spatially and temporally continuous disturbance mapping, but the detection of a given disturbance inherently depends on its size, as satellites operate with varying spatial resolutions. The GFC is based on Landsat with a 30 meter spatial resolution (Hansen et al., 2013), so that smaller scale events might remain undetected (Masek et al., 2013). For example, Baumann et al. (2014) showed that the accuracy of wind disturbance detection algorithms was generally higher for larger disturbance patches compared to smaller ones. Furthermore, for disturbances to be detected, the signal must reach a

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

detectable threshold for satellites to identify and quantify the affected area, meaning the severity of the disturbance plays a critical role (Masek et al., 2013; McDowell et al., 2015) in satellite-based mapping.

Additionally, disturbance events vary widely in both spatial extent - from several meters to hundreds of square kilometers — and temporal duration — from hours to multiple years (Turner, 2010), so that the suitability of each approach is also dependent on the disturbance characteristics. Methodological differences thus result in spatial and temporal uncertainties, as discussed below.

5.2 Spatial uncertainty

Our analysis is primarily based on the spatial overlap between the four datasets considered. Two datasets are point-based (FIA, ITMN), while the other two are spatially explicit (IDS, GFC), but with differences in spatial coverage.

To address this, two buffer sizes—500 (800) meters and 1000 (1600) meters for ITMN (FIA)—were introduced. While this approach increases comparability with other disturbance records, it also introduces additional uncertainty.

The FIA's guidelines suggest that swapping and fuzzing introduces only minimal differences in the data. Our results indicate that indeed uncertainties in disturbance location due to this approach do not significantly influence the spatial overlap with other datasets (Table 4), but they have a small contribution to temporal uncertainty, with larger spread in reported timing of disturbances in privately owned plots than national plots (Table 6).

The IDS inventory reports the total area affected by a type of disturbance agent, rather than the precise disturbed forest area. Contrary, the FIA only reports point-based disturbances. Therefore, satellite imagery might more accurately detect and delineate the exact disturbed area (Meddens et al., 2012; Masek et al., 2013), providing a detailed spatial perspective.

5.3 Temporal bias

In terms of temporal agreement, we found the best consistency between the FIA and IDS datasets, with the smallest mean difference of about -0.6 years, the negative value indicating that FIA tends to record disturbances earlier than IDS and GFC. We found moderately good temporal agreement between ITMN with IDS and between GFC with FIA and IDS, with mean differences in disturbance timing of -1.1 – -1.6 years and -1.3 – -1.5 years, respectively.

While ITMN tends to report mortality earlier than both IDS and GFC, it showed larger temporal biases than FIA. While this can be partly explained by spatial uncertainties, as discussed above, it is also likely to be due to the fact that is it based on events reported in the literature, thus *ad-hoc*.

IDS tends to report this difference later than FIA and ITMN, which can be explained by the fact that aerial surveys are conducted only a few times per year, and often fly regions only once a disturbance has been known. The Landsat-based tree loss dataset from GFC tends to detect disturbances after all the other datasets, which is likely due to the requirement of a given threshold in the signal for change detection algorithm underlying the dataset (Hansen et al., 2013). This difference likely explains the larger mean differences between ITMN with GFC, ITMN being including small events likely to go undetected in the satellite signal.

We find that differences in ownership status in the FIA dataset result in a larger mean differences with IDS datasets, but have a minimal effect on the bias with GFC. This difference might be due to easier access to public lands for aerial disturbance mapping, which can result in delayed mapping of disturbances over private forests.

5.4 Temporal spread

In addition to temporal biases, we find a very large spread in the differences in the timing of disturbances reported by the different datasets, consistently around 6–7 years for most pairs of datasets.

This spread can be explained by the uncertainty introduced by the revisiting frequency of the surveyed plots or areas, which varies across datasets. While the GFC provides temporally continuous data, inventory data often have survey gaps for certain areas and years (Burrill et al., 2021). These gaps can contribute to discrepancies in the timing of disturbance detection. If an area or plot is revisited only every five years, any disturbances occurring within that period might be recorded retrospectively, introducing uncertainty regarding the timing of the events. This is evident in the discrepancies between mortality year and measurement/inventory years in some of the events reported in the FIA dataset, which vary between 0 and 20 years.

Furthermore, disturbed areas in IDS can overlap in both space and time, for example if different agents are recorded across different years in the same region or plot. This results in a higher number of overlapping events but can contribute to a broader distribution of differences in the reported timing of disturbance compared to binary data such as GFC (loss/no-loss) or ITMN (mortality only). Such overlap also occur in the FIA, as plots can be re-visited.

Adding to this uncertainty is the uncertainty in the detection of disturbance in the remote-sensing signal. For example, Hansen et al. (2013) highlight uncertainties associated with the GFC dataset and advise using a 3-year moving window to detect trends, cautioning for area estimation using pixel counts of forest loss. Therefore, uncertainty from the change detection algorithm and can contribute to the larger spread found in the timing differences between GFC and other datasets, compared to IDS, and can partly explain differences in the spatial overlap with other datasets.

5.5 Disturbance agents

The results reveal a low consistency in the determination of disturbance agents, which can be attributed to several factors. One significant reason is the cascading effects of disturbances. For instance, an initial disturbance such as a drought can lead to secondary disturbances like bark beetle infestations, with the combined impact ultimately causing tree mortality (Bentz et al., 2010; Seidl et al., 2017; Burton et al., 2020). Bark beetles require several years of suitable weather conditions to survive and the abundance of the preferred host tress to cause an extensive outbreak (Fettig et al., 2022; Bentz et al., 2010). This highlights the complexity of accurately identifying and recording agents, as multiple interacting factors are often involved.

Information on the disturbance agents is, therefore, strongly dependent on the granularity of the dataset in terms of agent types and temporal dimension covered. For example, the FIA often use broad terms of agent classification, such as "insect" are provided, without specifying whether the agent is a bark beetle, defoliator, or another biotic agent. This broad categorization contrasts with the more detailed IDS data, potentially explaining the low agreement between these datasets. This is evident in our comparison (Figure 5), where FIA reports a large number of events as "other biotic", "other abiotic" and "other" (see

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

classification in Table B6), while IDS resolves these classes more finely into bark-beetle, defoliators and other biotic, and into fire, wind and drought. Note that some uncertainty has been introduced also by our own decision in aggregating data, when broad classifications were provided.

Ad-hoc literature-based synthesis, as is the case of ITMN (Hammond et al., 2022), are prone to biases in terms of the keywords used to select relevant studies, as well as inconsistencies in the interpretation and granularity of driving agents reported in each individual study. Indeed, here we found a tendency towards higher attribution to drought-related disturbances in ITMN, in isolation or in combination with biotic agents, while IDS tended to associate mortality events mostly with bark-beetle disturbance for the same events.

Differences in the reported agents could also arise from the temporal lags in the aerial surveys. For example if these are conducted only once the drought has passed, the disturbance is more likely be attributed only to the subsequent biotic agents. Such systematic bias in the IDS data is, however, unlikely, since we do not find the same differences in the attribution to drought vs. biotic agents in the comparison of IDS with FIA. Nevertheless, Coleman et al. (2018) showed that the detection of damage types and agents in IDS dataset could be improved by aligning flight conditions and timing with the biological windows of disturbance agents. However, they also noted that limited ground-truthing and the subjective nature of aerial surveys—affected by different surveyors, viewing conditions, and flying conditions—introduce variability and uncertainty into the data (Coleman et al., 2018).

Remote sensing technology is widely used in forest disturbance mapping, but attribution to specific agents based on remote-sensing data only is still limited. While now being a standard approach to identifying certain types of disturbances like those caused by fire (Chuvieco et al., 2018, 2022; Otón et al., 2019), distinguishing other disturbances such as different biotic agents, and especially non-stand replacing ones is more challenging (Senf et al., 2020; Senf and Seidl, 2021b; Senf et al., 2017; McDowell et al., 2015). It is worth noting that remote-sensing based detection approaches for disturbance mapping tend to rely on inventory data analysed here, so that they will inevitable inherit their uncertainty. Furthermore, the requirement of a significant threshold for tree loss detection, as applied in the GFC and other studies, might make the detection dependent on human effects, such as selective logging that often follow disturbance events. This can contribute to the lag between the GFC tree loss detection and the timing of disturbance and/or mortality reported by the other datasets.

These factors collectively contribute to the observed low consistency in disturbance agent determination and underscore the challenges involved in accurately identifying and classifying forest disturbances. The discrepancies between datasets, the complexities of cascading disturbances, and the limitations of current detection methods all highlight the need for improved approaches in forest disturbance monitoring.

6 Conclusions

In our study, we demonstrate the importance of consistent and reliable data for understanding forest disturbances. By comparing four datasets on forest disturbances across the continental United States, we identified varying levels of consistency, with inventory-based datasets showing the highest reliability.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

In general, inventory datasets tend to identify lower severity and smaller-scale disturbances earlier than remote sensing methods. The discrepancies found with remote sensing data, especially the delayed detection of disturbances, highlight the challenges of relying solely on satellite observations for accurate and timely monitoring.

Our analysis further revealed inconsistencies in identifying disturbance agents of overlapping disturbances. These discrepancies are due to varying levels of detail in the datasets and the subjective determination of disturbance agents. This finding emphasizes the need for more standardized and detailed attribution of disturbance agents across datasets. Moreover, we show the importance of accounting for inherent data uncertainties, particularly those related to forest ownership and discrepancies between revisiting and reporting times.

Our findings emphasize the need for careful consideration of the multiple sources of uncertainty and strict pre-processing of the data for use in other applications such remote-sensing disturbance classification models. Our analysis of spatial and temporal differences between datasets allowed us to provide quantitative estimates of the accuracy and precision of the different datasets, contributing to more robust and informed decision-making when using such data.

As forests continue to face increasing pressures from climate-induced disturbances, higher integration and increased accuracy in disturbance datasets is crucially needed. Future research could focus on enhancing the consistency between different methodologies and terminology in disturbance detection and classification. By addressing these challenges, we can advance our understanding of forest disturbances and their drivers.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Appendix A: Figures

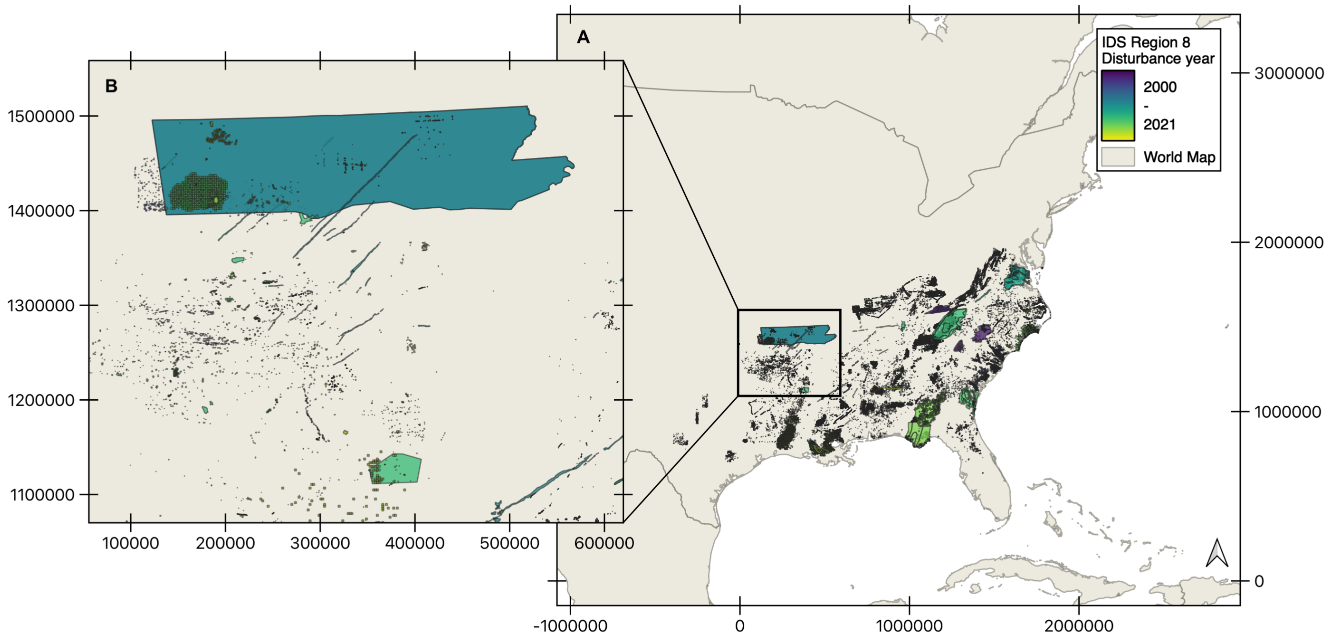

Figure A1. Map of IDS Region 8 with disturbance polygons with the color scheme of earlier disturbance events (2000s) in purple and more recent events in green and yellow (2020s). Panel A shows an overview of the IDS disturbances in Region 8, panel B shows the data with increasing detail zoomed inn to an irregular disturbance shape in blue.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Appendix B: Tables

Column name	Meaning
DCA_ID	Name of disturbance agents based on new coding system from this study
SURVEY_YEAR	Year of survey/ detection of disturbance
geometry	polygon extent in lat long
REGION_ID	number of CONUS/ AK region
DAMAGE_TYPE	type of damage (mortality, crown discoloration, defoliation, dieback, ...)
DAMAGE_TYPE_CODE	Code according to damage type
DCA_CODE	Detailed code of disturbance agent
DA_Code_USDA	Code according to new DCA name (DCA_ID)
PERCENT_AFFECTED	Percentage of affected area within polygon
HOST	Name of host trees
HOST_CODE	Code according to host tree

Table B1. Columns of interest of IDS data.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

DCA_CODE	DCA_NAME
10000	Miscellaneous insects and allies
11000	Bark beetles
12000	Defoliator
14000	Sucking insects
15000	Boring insects
16000	Seed/cone/flower/fruit insects
17000	Gallmaker insects
18000	Predatory insects
19000	General diseases
20000	Biotic damage
21000	Root/butt disease
22000	Stem decays/cankers
23000	Parasitic/epiphytic plants
24000	Decline complexes/dieback
25000	Foliage and shoot diseases
26000	Stem rust
27000	Broom rust
28000	Boring insects - shoot and twig
29000	Root feeding insects
30000	Fire
41000	Wild animals
42000	Domestic animals
50000	Abiotic damage
60000	Competition
70000	Human activies
80000	Multi-damage (insect/disease)
90000	Unknown

Table B2. DCA codes in IDS of different agent categories. The first two digits of the code describe the category code (10, 11, 12), the next three digits (here all 0) describe the agent code, which is normally more detailed (e.g. 001, 002), which was not needed for this study.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Column name	Meaning
PLT_CN	Plot sequence number
COUNTYCD	County Code
FORTYPCD	Forest type code
COND_STATUS _CD	Condition status code
DSTRBCD1	Disturbance code 1
DSTRBYR1	Disturbance year 1
DSTRBCD2	Disturbance code 2
DSTRBYR2	Disturbance year 2
DSTRBCD3	Disturbance code 3
DSTRBYR3	Disturbance year 3
OWNCD	Owner class code
OWNGRPCD	Owner group code
FORINDCD	Private owner industrial status code

Table B3. Columns of interest of FIA condition data.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Column name	Meaning
CN	Sequence number
INVYR	Inventory year
STATECD	State code
PLOT	Plot number
MEASYEAR	Measurement year
MEASMON	Measurement month
MEASDAY	Measurement day
LAT	Latitude
LON	Longitude

Table B4. Columns of interest of FIA plot data.

Column name	Meaning
PLT_CN	Plot sequence number
MORTYR	Mortality year
AGENTCD	Cause of death (agent) code
DAMAGE_AGENT _CD1	Damage agent code 1
DAMAGE_AGENT _CD2	Damage agent code 2
DAMTYP1	Damage type 1
DAMSEV1	Damage severity 1
DAMLOC2	Damage location 2
DAMTYP2	Damage type 2
DAMSEV2	Damage severity 2

Table B5. Columns of interest of FIA tree data.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Agent FIA Name	Agent FIA Code	Agent IDS Name	Agent IDS Code
No visible disturbance	0	No data	0
Insect	10	Other biotic	64
Disease	20	Other biotic	64
Fire	30	Fire	8
Animal	40	Other biotic	64
Weather	50	Other abiotic	128
Vegetation (suppression, competition)	60	Other	256
Unknown	70	Other	256
Silvicultural or landcleaning activity	80	Other	256
Other	90	Other abiotic	128

Table B6. Agent coding mapping for FIA broad agent categories based on IDS new coding system.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Agent FIA Name	Agent FIA Code (Range)	Agent IDS Name	Agent IDS Code
No visible disturbance	0	No data	0
Insect	10 to 19	Other biotic	64
Disease	20 to 29	Other biotic	64
Fire	30 to 39	Fire	8
Animal	40 to 49	Other biotic	64
Weather	50 to 59, except 52 and 54	Other abiotic	128
Wind	52	Wind	16
Drought	54	Drought	4
Vegetation (suppression, competition)	60 to 69	Other	256
Unknown (other)	70 to 79	Other	256
Silvicultural or landcleaning activity	80 to 89	Other	256
Other	≥ 90	Other abiotic	128

Table B7. Agent coding mapping for ranges FIA agent categories based on IDS new coding system. The ranges of FIA codes are based on the information of agent code (*AGENTCD*) and disturbance code (*DSTRBCD1*).

Code availability. The Code will be made publicly available upon publication.

Data availability. Insect and Disease Survey data by U.S. Department of Agriculture can be downloaded here: https://www.fs.usda.gov/science-technology/data-tools-products/fhp-mapping-reporting/detection-surveys; Forest Inventory and Analysis Data by U.S. Department of Agriculture can be downloaded here: https://research.fs.usda.gov/products/dataandtools/tools/fia-datamart; Global Tree Mortality Database by the International Tree Mortality Network can be found here: https://www.tree-mortality.net/globaltreemortalitydatabase/; Global forest change data by Global Forest Watch and Global Land Analysis and Discovery can be downloaded here: https://storage.googleapis.com/earthenginepartners-hansen/GFC-2021-v1.9/download.html

Author contributions. AB and LE designed the study. LE conducted the analysis with scientific and technical input by AB and FM. LE wrote the first draft of the paper. All authors contributed to the writing of the MS submitted.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. FM and AB acknowledge funding by the European Union (ERC StG, ForExD, grant agreement No. 101039567). AB and LE acknowledge support from the CALIPSO project (Carbon Loss in Plant Soils and Oceans) project, funded through the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program.

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

References

Allen, C. D., Breshears, D. D., and McDowell, N. G.: On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene, Ecosphere, 6, art129, https://doi.org/10.1890/ES15-00203.1, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1890/ES15-00203.1, 2015.

Anderegg, W. R. L., Hicke, J. A., Fisher, R. A., Allen, C. D., Aukema, J., Bentz, B., Hood, S., Lichstein, J. W., Macalady, A. K., Mcdowell, N., Pan, Y., Raffa, K., Sala, A., Shaw, J. D., Stephenson, N. L., Tague, C., and Zeppel, M.: Research review Tree mortality from drought, insects, and their interactions in a changing climate, New Phytologist, 208, 674–683, https://doi.org/10.1111/nph.13477, 2015.

Bastos, A., Orth, R., Reichstein, M., Ciais, P., Viovy, N., Zaehle, S., Anthoni, P., Arneth, A., Gentine, P., Joetzjer, E., Lienert, S., Loughran, T., McGuire, P. C., O, S., Pongratz, J., and Sitch, S.: Vulnerability of European ecosystems to two compound dry and hot summers in 2018 and 2019, Earth System Dynamics, 12, 1015–1035, https://doi.org/10.5194/esd-12-1015-2021, publisher: Copernicus GmbH, 2021.

Bastos, A., Sippel, S., Frank, D., Mahecha, M. D., Zaehle, S., Zscheischler, J., and Reichstein, M.: A joint framework for studying compound ecoclimatic events, Nature Reviews Earth & Environment, pp. 1–18, https://doi.org/10.1038/s43017-023-00410-3, publisher: Nature Publishing Group, 2023.

Baumann, M., Ozdogan, M., Wolter, P. T., Krylov, A., Vladimirova, N., and Radeloff, V. C.: Landsat remote sensing of forest windfall disturbance, Remote Sensing of Environment, 143, 171–179, https://doi.org/10.1016/j.rse.2013.12.020, 2014.

Bentz, B. J., Régnière, J., Fettig, C. J., Hansen, E. M., Hayes, J. L., Hicke, J. A., Kelsey, R. G., Negrón, J. F., and Seybold, S. J.: Climate Change and Bark Beetles of the Western United States and Canada: Direct and Indirect Effects, BioScience, 60, 602–613, https://doi.org/10.1525/bio.2010.60.8.6, 2010.

Bonan, G. B.: Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests, Science, 320, 1444–1449, https://doi.org/10.1126/science.1155121, 2008.

Bowman, D. M. J. S., Balch, J. K., Artaxo, P., Bond, W. J., Carlson, J. M., Cochrane, M. A., D'Antonio, C. M., DeFries, R. S., Doyle, J. C., Harrison, S. P., Johnston, F. H., Keeley, J. E., Krawchuk, M. A., Kull, C. A., Marston, J. B., Moritz, M. A., Prentice, I. C., Roos, C. I., Scott, A. C., Swetnam, T. W., van der Werf, G. R., and Pyne, S. J.: Fire in the Earth System, Science, 324, 481–484, https://doi.org/10.1126/science.1163886, 2009.

Burrill, E. A., DiTommaso, A. M., Turner, J. A., Pugh, S. A., Menlove James, C. G., Perry, C. J., and Conkling, B. L.: The Forest Inventory and Analysis Database: Database Description and User Guide for Phase 2 (Version 9.0.1), U.S. Department of Agriculture, Forest Service, 29, 1026, http://www.fia.fs.fed.us/library/database-documentation/, 2021.

Burton, P. J., Jentsch, A., and Walker, L. R.: The Ecology of Disturbance Interactions, BioScience, 70, 854–870, https://doi.org/10.1093/biosci/biaa088, 2020.

Chuvieco, E., Lizundia-Loiola, J., Pettinari, M. L., Ramo, R., Padilla, M., Tansey, K., Mouillot, F., Laurent, P., Storm, T., Heil, A., and others: Generation and analysis of a new global burned area product based on MODIS 250 m reflectance bands and thermal anomalies, Earth System Science Data, 10, 2015–2031, publisher: Copernicus GmbH, 2018.

Chuvieco, E., Roteta, E., Sali, M., Stroppiana, D., Boettcher, M., Kirches, G., Storm, T., Khairoun, A., Pettinari, M. L., Franquesa, M., and Albergel, C.: Building a small fire database for Sub-Saharan Africa from Sentinel-2 high-resolution images, Science of The Total Environment, 845, 157 139, https://doi.org/10.1016/j.scitotenv.2022.157139, 2022.

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

FAO: Global Forest Resources Assessment 2010 (FRA2010), Tech. rep., Food and Agriculture Organization of the United Nations, Rome, 2010.

FAO: Global Forest Resources Assessment 2015 (FRA2015), Tech. rep., Food and Agriculture Organization of the United Nations, Rome, 2015.

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

Fettig, C. J., Egan, J. M., Delb, H., Hilszczan´ski, J., Kautz, M., Munson, A. S., Nowak, J. T., and Negró, J. F.: Management tactics to reduce bark beetle impacts in North America and Europe under altered forest and climatic conditions, in: Bark Beetle Management, Ecology, and Climate Change, edited by Gandhi, K. J. K. and Hofstetter, R. W., pp. 345–394, Academic Press, https://doi.org/10.1016/B978-0-12-822145-7.00006-4, 2022.

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

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

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

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

Hammond, W. M., H., Williams, A. P., Abatzoglou, J. T., Adams, H. D., Klein, T., López Rodríguez, R., Sáenz-Romero, C., Hartmann, H., Breshears, D. D., and Allen, C. D.: Global field observations of tree die-off reveal hotter-drought fingerprint for Earth's forests, Nature Communications, data downloaded: 28.04.2022, 2022.

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

Hartmann, H., Bastos, A., Das, A. J., Esquivel-Muelbert, A., Hammond, W. M., Martínez-Vilalta, J., McDowell, N. G., Powers, J. S., Pugh, T. A., and Ruthrof, K. X.: Climate change risks to global forest health: emergence of unexpected events of elevated tree mortality worldwide, Annual Review of Plant Biology, 73, 673–702, publisher: Annual Reviews, 2022.

Hicke, J. A., Allen, C. D., Desai, A. R., Dietze, M. C., Hall, R. J., Kashian, D. M., Moore, D., Raffa, K. F., Sturrock, R. N., Vogelmann, J., and others: Effects of biotic disturbances on forest carbon cycling in the United States and Canada, Global Change Biology, 18, 7–34, 2012.

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

International tree mortality network: Global Tree Mortality Database, https://www.tree-mortality.net/globaltreemortalitydatabase/, data downloaded: 28.04.2022, 2022.

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

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

Masek, J. G., Goward, S. N., Kennedy, R. E., Cohen, W. B., Moisen, G. G., Schleeweis, K., and Huang, C.: United States Forest Disturbance Trends Observed Using Landsat Time Series, Ecosystems, 16, 1087–1104, https://doi.org/10.1007/s10021-013-9669-9, 2013.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., and Gomis, M.: Climate change 2021: the physical science basis, Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change, 2, publisher: Cambridge University Press Cambridge, UK, 2021.

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

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

Meddens, A. J. H., Hicke, J. A., and Ferguson, C. A.: Spatiotemporal patterns of observed bark beetle-caused tree mortality in British Columbia and the western United States, Ecological Applications, 22, 1876–1891, 2012.

Meigs, G. W., Morrissey, R. C., Bače, R., Chaskovskyy, O., Čada, V., Després, T., Donato, D. C., Janda, P., Lábusová, J., Seedre, M., Mikoláš, M., Nagel, T. A., Schurman, J. S., Synek, M., Teodosiu, M., Trotsiuk, V., Vítková, L., and Svoboda, M.: More ways than one: Mixed-severity disturbance regimes foster structural complexity via multiple developmental pathways, Forest Ecology and Management, 406, 410–426, https://doi.org/10.1016/j.foreco.2017.07.051, publisher: Elsevier B.V., 2017.

Otón, G., Ramo, R., Lizundia-Loiola, J., and Chuvieco, E.: Global Detection of Long-Term (1982–2017) Burned Area with AVHRR-LTDR Data, Remote Sensing, 11, 2079, https://doi.org/10.3390/rs11182079, number: 18 Publisher: Multidisciplinary Digital Publishing Institute, 2019.

Patacca, M., Schelhaas, M.-J., Zudin, S., and Lindner, M.: Database on Forest Disturbances in Europe (DFDE)-Technical report History, State of the Art, and Future Perspectives, 2021.

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

https://doi.org/10.5194/egusphere-2024-3534
Preprint. Discussion started: 19 November 2024

Seidl, R. and Turner, M. G.: Post-disturbance reorganization of forest ecosystems in a changing world, Proceedings of the National Academy of Sciences, 119, e2202190 119, https://doi.org/10.1073/pnas.2202190119, publisher: Proceedings of the National Academy of Sciences, 2022.

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

Senf, C. and Seidl, R.: Persistent impacts of the 2018 drought on forest disturbance regimes in Europe, Biogeosciences, 18, 5223–5230, https://doi.org/10.5194/bg-18-5223-2021, 2021a.

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

Senf, C., Seidl, R., and Hostert, P.: Remote sensing of forest insect disturbances: Current state and future directions, International Journal of Applied Earth Observation and Geoinformation, 60, 49–60, 2017.

Senf, C., Buras, A., Zang, C. S., Rammig, A., and Seidl, R.: Excess forest mortality is consistently linked to drought across Europe, Nature Communications, 11, 6200, https://doi.org/10.1038/s41467-020-19924-1, 2020.

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

Tinkham, W. T., Mahoney, P. R., Hudak, A. T., Domke, G. M., Falkowski, M. J., Woodall, C. W., and Smith, A. M.: Applications of the United States forest inventory and analysis dataset: A review and future directions, Canadian Journal of Forest Research, 48, 1251–1268, https://doi.org/10.1139/cjfr-2018-0196, publisher: Canadian Science Publishing, 2018.

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