

## Referee Comment 1

The paper combines three different datasets on forest disturbances for a large region in the Southern and Eastern US: IDS, S1DM, and Planetscope. The first one is a routine product freely accessible with yearly updates, the second one is generated by the authors based on Sentinel-1 time series, and the third one is merely a case-based evaluation tool including manual delineation of polygons with damaged forest.

**RC1:** Out of the many disturbances, the authors select windthrow, bark beetle attacks, and defoliators, representing important categories, but leaving out fire.

We thank the reviewer for highlighting this point. We agree that fire is indeed a critically important disturbance category. For example, Global Forest Watch reports that between 2001 and 2024, approximately 151 Mha of tree cover loss worldwide was attributed to fire, roughly the size of Mongolia.

At the same time, we chose not to include fire in our study because there are already excellent, globally available fire-monitoring systems, such as NASA's FIRMS (Fire Information for Resource Management), which provide near-real-time detection of active fires from MODIS and VIIRS, often within 3 hours of satellite observation. These resources offer a level of observation and detection that is already outstanding and widely used.

Our study, therefore, focuses on disturbances that are less well monitored by remote sensing, such as windthrow, bark beetle outbreaks, and defoliators, for which we believe our approach can provide novel insights and added value.

**RC2:** Which areas are considered and which are excluded is largely determined by the flight trajectories leading to the IDS dataset - this is a bit disappointing since S1 tiles have global coverage, and the advantage of satellite remote sensing is the ability to conclude on patterns outside ground-based (or aerial for that matter) observations, provided sufficient training data. It would be interesting to see an area with a damage according to the Planetscope manual delineation and the S1 trend detection there, but not covered by the IDS. Of course, the TCC might still be used to exclude non-forested areas. The exercise would demonstrate the power of the S1 disturbance detection completely independent from the IDS data; in general, the two approaches are just compared, they do not depend on each other.

We thank the reviewer for this insightful comment. Indeed, Sentinel-1 has global coverage, and in principle, S1-based disturbance detection could be applied independently of the IDS dataset; combined with TCC, this could provide valuable insights into forest disturbances.

However, the reviewer's suggestion would require identifying disturbed patches based on Planet data that are not detected by IDS. This would imply that other high-quality, spatially explicit, disturbance classification datasets independent of IDS would exist. This is precisely the

knowledge gap we hope to address in the future through satellite-based disturbance detection and classification.

Detailed information on disturbance agents, especially insects, is available in the IDS dataset and, to some extent, in the Forest Inventory Assessment, but it is based on plot-level sampling with irregular revisit times (see Eifler et al., BG, in press).

This is why our analysis is limited to IDS-covered areas: without IDS, we would be unable to associate observed disturbances with their causal agents.

We agree that such an exercise could be valuable and highlight it as an interesting avenue for future research in our manuscript in the new Section 5.3. Methodological Limitations:

*“While Sentinel-1 provides global coverage and enables disturbance detection beyond areas surveyed by aerial or ground-based inventories, the attribution of detected disturbances to specific causal agents remains challenging in the absence of independent, spatially explicit reference data. In this study, the analysis was therefore restricted to areas covered by the IDS dataset, as IDS provides detailed information on disturbance agents that is currently not available from satellite-based products alone. Future work could extend the presented framework by combining Sentinel-1-based disturbance detection with independent high-resolution reference data, such as manual delineations from Planet imagery in regions not covered by IDS, while using tree canopy cover products to exclude non-forested areas. Such an analysis would allow for the full evaluation of Sentinel-1 disturbance detection independent of IDS and further demonstrate its potential for large-scale forest disturbance monitoring.”*

**RC3:** Concerning the exclusion of non-forest areas, the authors set an unnecessary strict threshold for the presence of a forest, i.e. 30% canopy cover. This is not aligned with the FAO definition of a forest as any area of minimum size 0.5 ha with a canopy cover (for trees which can grow to more than 5 meters) of only 10% (<https://fra-data.fao.org/definitions/fra/2020/en/tad>). As the minimum area required according to FAO is only 12.5 pixels for S-1, the 30% seems to be overly restrictive.

We thank the reviewer for this thoughtful comment.

Given the FAO definition, a minimum canopy cover of 10% for a minimum size of 0.5ha corresponds to ca. 500m<sup>2</sup>, which is larger than the 400m<sup>2</sup> (20x20m) of the S-1 pixels considered here. The main reason to exclude pixels with low Tree Canopy Cover (TCC) is, however, the nature of the S-1 signal. At the scale of an S-1 pixel, a large fraction of non-forest vegetation types (e.g., croplands with strong seasonal phenology) could influence the change-detection algorithm, for reasons not necessarily associated with forest disturbances. Therefore, our strict filtering aims to ensure that the patches identified as “disturbed” are associated with forest disturbances rather than other types of interventions, such as agricultural practices.

Nevertheless, we reran the whole analysis using a less restrictive TCC threshold of 10% as suggested. The results remain qualitatively very similar to those presented in the manuscript. The corresponding results for the 10% TCC threshold are shown in Figures R1.1–R1.4.

With the 10% TCC threshold, the total number of detected IDS events increases slightly (Bark Beetle: from 1,177 to 1,186; Wind: from 478 to 508; Defoliators: from 213 to 307). A similar increase is observed for the corresponding S1CD events (Bark Beetle: from 899 to 925; Wind: from 469 to 501; Defoliators: from 144 to 224). Importantly, despite these changes in absolute numbers, the relative loss percentages remain largely consistent: Bark Beetle changes from -23.62% to -22.01%, Wind from -1.88% to -1.38%, and Defoliators from -32.39% to -27.04%.

Overall, this analysis confirms that our results are robust to a less restrictive TCC threshold. However, for the main analysis, we retain the 30% TCC threshold, as it provides a more conservative and reliable estimate of forested areas.

We have added a clarification to the manuscript in Section 3.2 (Tree Canopy Cover):

*“We applied a 30% TCC threshold, which is stricter than the FAO forest definition ( $\geq 10\%$  canopy cover; FAO, 2020), to ensure that detected disturbed patches correspond to forest disturbances and not changes in vegetation structure driven by other processes, such as agricultural practices. Using a 10% threshold yields very similar results, with fewer than 80 additional events per disturbance type.”*

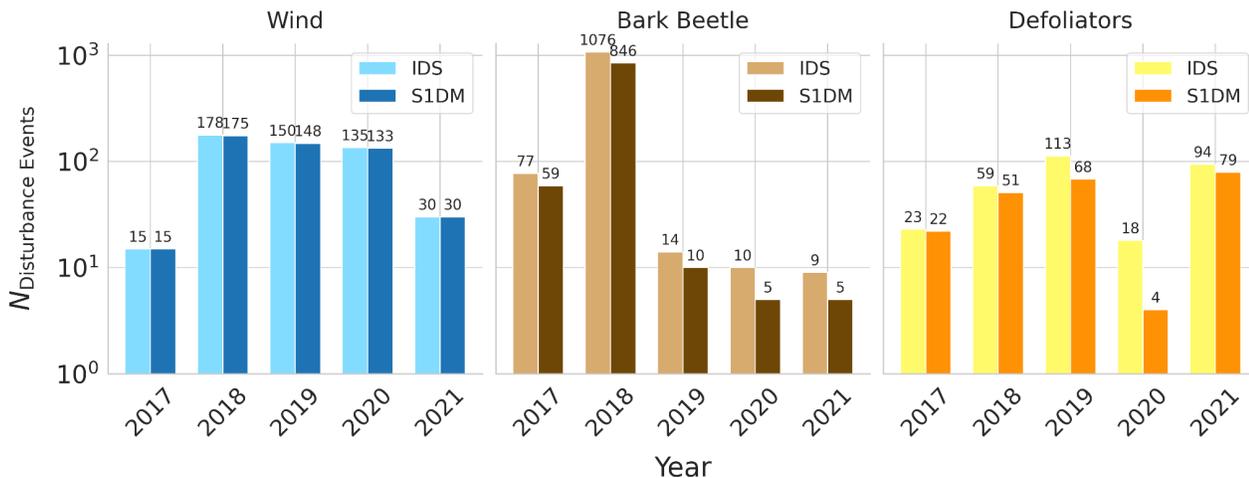


Figure R1.1: Detection efficiency of various disturbance types using Radar Change Detection. The bar plot shows the number of events on the y-axis (log scale) for each disturbance type on the x-axis, with IDS shown in light color and S1DM in darker color. The total number of IDS disturbance events that had a corresponding S1DM signal within the 500 m buffer was 501 for wind (S1DM: 501; IDS: 508), 925 for bark beetle (S1DM: 925; IDS: 1186), and 224 for defoliators (S1DM: 224; IDS: 307).

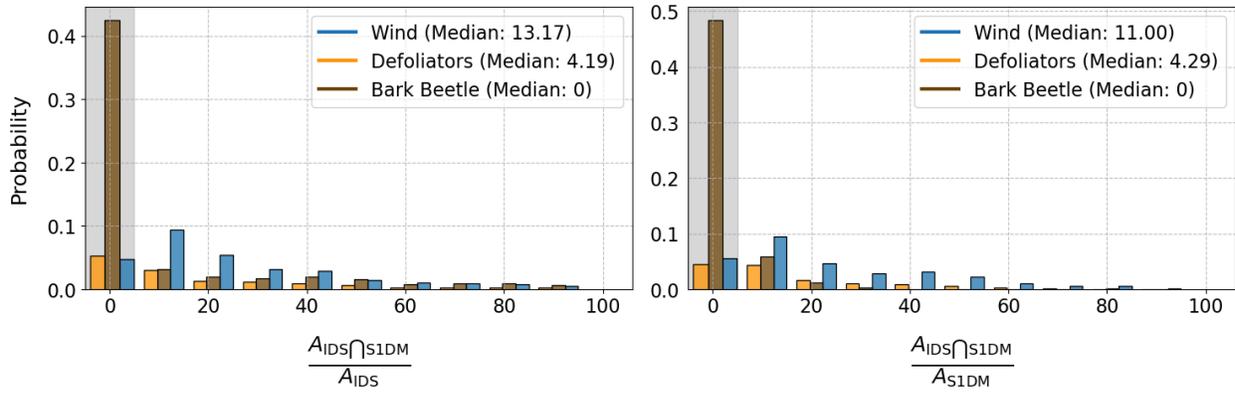


Figure R1.2: Probability of (a) IDS polygons being covered by S1DM, and (b) S1DM polygons overlapping with IDS, for each disturbance type. Disturbance types are color-coded as follows: Wind (blue), Bark Beetle (brown), and Defoliators (yellow). Median probabilities for each disturbance type are indicated in parentheses in the legend.

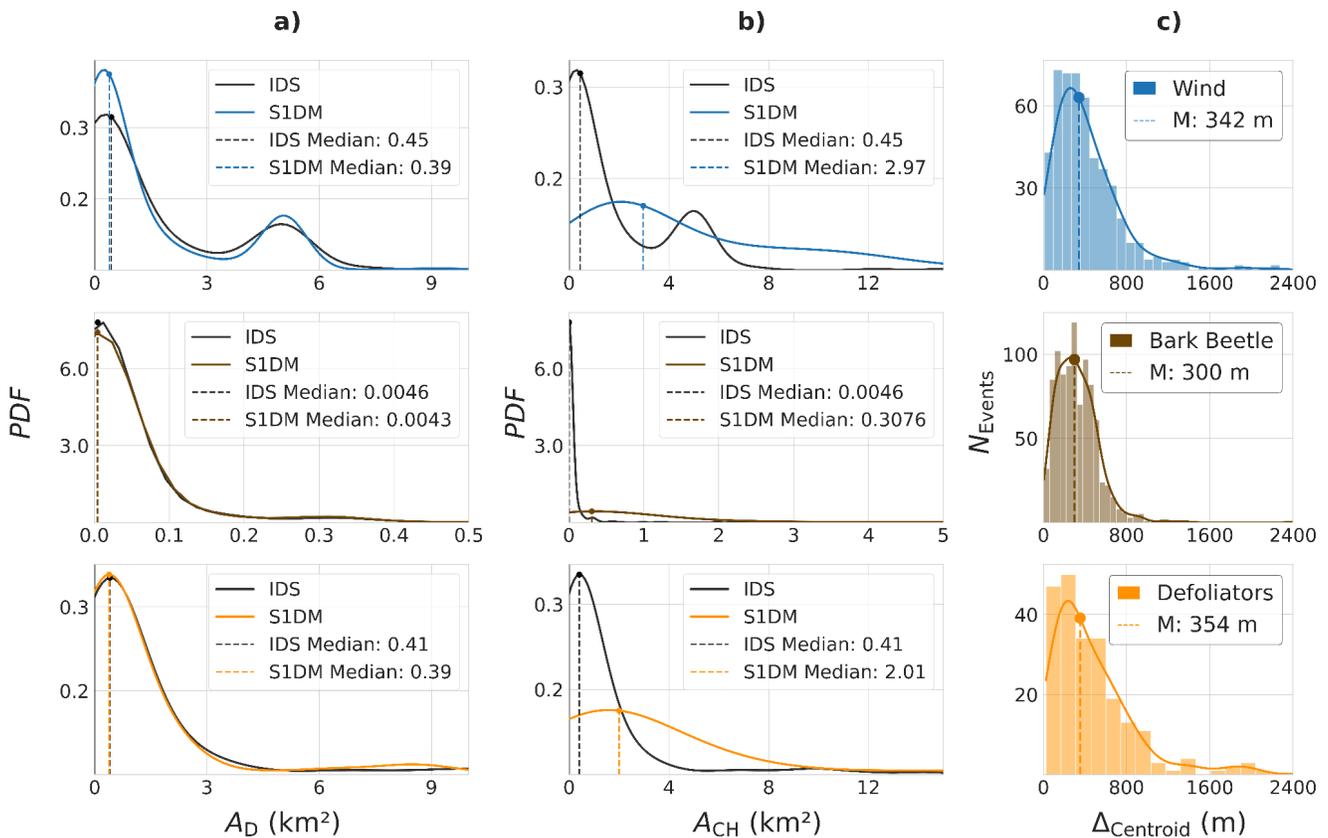


Figure R1.3: Comparison of disturbance size (in  $\text{km}^2$ ) and spatial bias (in m) between the IDS and S1DM datasets. The figure consists of three vertical panels, each displaying the results for three disturbance types: Wind (top, blue), Bark Beetle (second, brown) and Defoliators (third, green). a) shows the probability density function (PDF) of disturbance patch areas for IDS (black lines) and S1DM (colored lines) across the four disturbance types. b) presents the PDF of the convex-hull (CH) areas, representing the spatial spread of the disturbance polygons, using the same color scheme for S1DM and IDS as in the

left panel. c) displays histograms and corresponding PDFs of the spatial distance between IDS and S1DM centroids for each disturbance type.

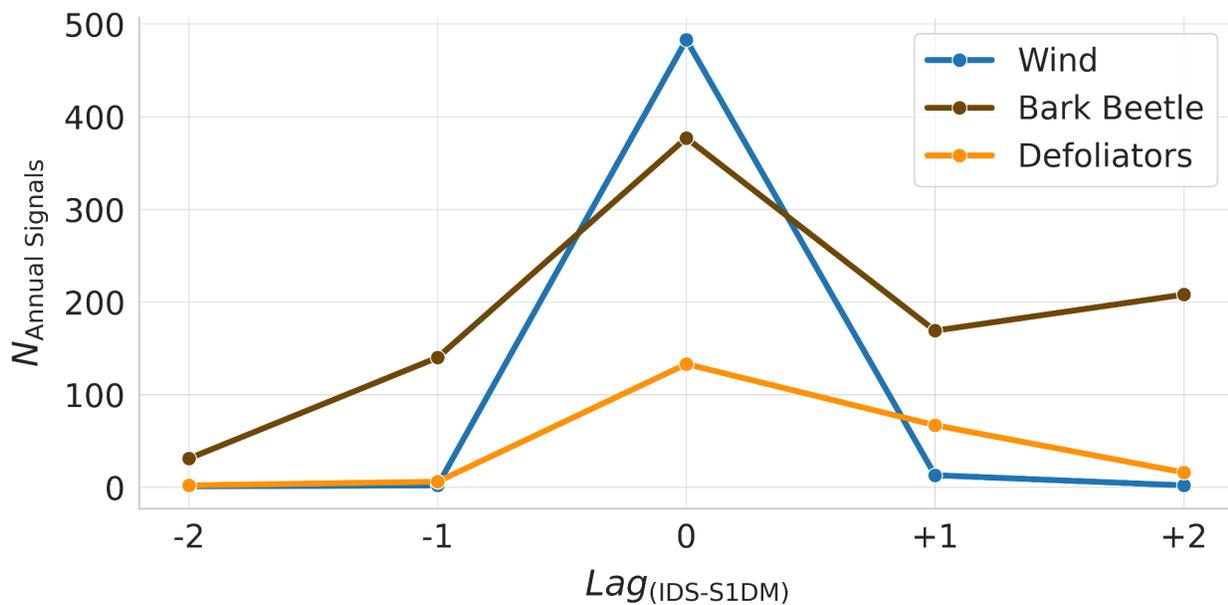


Figure R1.4: Temporal relationship between disturbance detection by S1CD and IDS. The x-axis represents the temporal yearly agreement with IDS - S1CD. Negative values indicate that IDS recorded the disturbance before S1CD detected it. The y-axis shows the count of events corresponding to each year's lag, with different disturbance types represented by distinct colored lines. We focus on the single S1DM timestamp closest to the IDS year for each event. We calculate the temporal lag by subtracting the S1DM detection year from the IDS year. Negative values indicate that S1DM detection occurred after IDS, while positive values indicate earlier radar detection.

**RC4:** On page 3, l. 84-92, ML and DL are mentioned. While it is largely correct what is written here, it seems that not a single ML or DL method is applied in the rest of the paper. What is the purpose of that paragraph? Is it a leftover from earlier versions of the manuscript? It might as well be deleted.

We agree with the reviewer that no machine learning or deep learning model is applied in this manuscript. However, the point we make in the introduction (which is then discussed later in the manuscript) is that these methods have been increasingly used to advance forest disturbance mapping, but they are data-intensive and require reliable labels, which in turn are labor-intensive and costly. Here we show that while IDS provides unique information about disturbance agents, it also suffers from limitations regarding the spatial delineation of the disturbed patches, as exemplified in Figures R1.5 R1.7. Also, see a more in-depth analysis of the spatio-temporal uncertainty of IDS data in Eifler et al. (in press). Our goal is thus two-fold: improve the delineation of disturbed patches identified by IDS using radar information from S-1 and, while doing so, produce a standardized benchmark for training and validating ML and DL models.

Since the detailed discussion of DL/ML takes up considerable space in the introduction but is not central to the manuscript's topic, we have shortened the paragraph as follows:

*“Increasingly, remote sensing studies use data-driven approaches (Artificial Intelligence) to detect, map, and classify forest disturbances, demonstrating the ability to capture complex spatiotemporal patterns (Andresini et al., 2024; Bárta et al., 2021; Hawryło et al., 2018; Gibson et al., 2020). The effectiveness of these models depends on high-quality reference data, as accurate ground-truth labels are essential for training and validating disturbance detection algorithms. This data-hungry approach creates a bottleneck, as labeling forest disturbances is labor-intensive and costly, requiring expert knowledge to distinguish agents such as insects, drought, or biotic stress.”*

**RC5:** Regarding the size of disturbed areas, a maximum size of 15 km<sup>2</sup> is used. This seems to be an arbitrary threshold and a huge difference to the maximum size of 2000 km<sup>2</sup> used by Eifler et al. (2024). The only justification is (l. 278) “we applied a stricter filter”. Certainly you did, but why? Many beetle attacks happen on or spread to larger areas, similar with defoliators.

We thank the reviewer for this important comment. We agree that the 15 km<sup>2</sup> maximum disturbance threshold may currently appear arbitrary, especially compared to the 2000 km<sup>2</sup> threshold used by Eifler et al. (in press).

To provide context, we analyzed the distribution of disturbance sizes in our dataset, focusing on bark beetles, defoliators, and wind. Out of 27,776 disturbances (before excluding compound events), only 257 events (~0.93%, i.e., the top 1%) exceeded 15 km<sup>2</sup>, while the vast majority, 27,519 events (~99.07%), were smaller or equal to this threshold. The mean area of the larger disturbances was 120.67 km<sup>2</sup>.

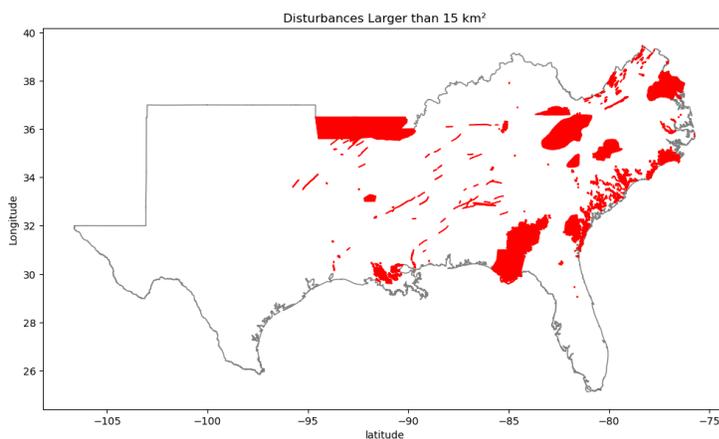


Figure R1.5: Disturbances (bark beetles, defoliators, and wind) before preprocessing with sizes larger than 15 km<sup>2</sup> in the study region.

This size, combined with visual inspection (in the provided Figure R1.5 above), indicates that these unusually large disturbances are likely the result of human error/imprecision rather than true ecological events. Given that these are outliers in terms of extent and likely unreliability, we applied the stricter 15 km<sup>2</sup> filter to focus our analysis on realistically sized disturbance events while avoiding potential bias.

To clarify this point, we have added the following sentence to explain our decision of excluding the above 15 km<sup>2</sup> elements:

*“Less than 1% of the disturbances exceed 15 km<sup>2</sup>, but these rare events cover unrealistically large and unspecific areas (mean size ~121 km<sup>2</sup>) and were therefore removed to avoid introducing bias and uncertainty into our analysis.”*

**RC6:** The polygon sizes mentioned in Table 1 are hard to believe. The smallest one (0.5 m<sup>2</sup>) would cover a single tree at most, and would be undetectable in aerial imagery. The largest one would be a significant fraction of all forest in region 8. Double-check these numbers.

We thank the reviewer for carefully examining these numbers. We agree that the polygon sizes reported in Table 1 may appear extreme at first glance.

Analyzing the raw IDS data for Region 8 before any filtering, we find that of the total 74,281 disturbances recorded, the largest was a snow-ice event covering 62,230.66 km<sup>2</sup>, and the smallest was a southern pine beetle attack covering 0.00000041 km<sup>2</sup> (≈0.41 m<sup>2</sup>).

In the paper, we reported polygon sizes ranging from 0.5 m<sup>2</sup> to 62,231 km<sup>2</sup> (rounded), reflecting the dataset before the filtering applied to Region 8. These extreme values highlight the heterogeneous nature of the dataset: very small entries indeed likely represent single-tree disturbances or isolated points, whereas the largest events correspond to large aggregated polygons.

**RC7:** The RQA TREND method for change detection is perfectly valid; however, the method has three parameters: the embedding dimension  $m$ , the recurrence threshold  $\epsilon$ , and the delay  $\tau$ . The results for the slope away from the main diagonal depends on them. None of the parameters is provided in the text; unfortunately, Cremer et al. (2020) does not mention them either; you also refer to the “European Commission... (2023)” proceedings, what you really mean is the Cremer et al. article on pp. 361-364 in that book (please be more precise in your referencing), but that article does not contain the values for the parameters either. Thus, the threshold for the trend -1.28 (l. 316 – what you probably mean is -1.28 yr<sup>-1</sup>) appears completely arbitrary (again, also this threshold is not mentioned in the Cremer et al. articles), what is its justification? It seems to be THE crucial parameter to distinguish non-disturbance to disturbance – how robust are your results against changing it? The extreme patchiness of the disturbance areas (e.g., as seen in Figure 7) could be a result of that choice. – You are also stating the

opposite of what you intend to say in I. 316f, the correct version would be “Pixels with a RQA-Trend value ABOVE the threshold of -1.28 were considered to show no significant change.” Please be more explicit here, and check the consequences of changing the trend threshold. Did you calculate one RP per time series, or did you move a window (e.g. of one year length) across the time series and calculated a separate RP and thus a TREND each time? If a disturbance sets in, it is to be expected that some of the RQA variables (TREND, but possibly also DET) “react” more or less suddenly (e.g., for wind). That would provide an opportunity to put a more precise date for the onset of the disturbance.

Thank you for the careful consideration of the RQA TREND Methodology. We agree that we should add more detail about these three parameters.

We have corrected the previously cited “European Commission...(2023)” proceedings to [2] and added Cremer et al. (2020) [1]; together, these represent the most recent and appropriate references for this method. The full references are provided at the end of this response.

Regarding the RQA parameters, following the framework implemented in [1, 2], recurrence plots were constructed directly from the univariate Sentinel-1 backscatter time series without phase-space reconstruction. The RecurrenceAnalysis.jl package (the Julia implementation we are using, [RecurrenceAnalysis.jl](#)) supports the use of embedded state vectors if provided, in this study, we did not perform time-delay embedding, and thus no embedding dimension or delay parameters were involved. The recurrence threshold  $\epsilon$  was the sole parameter used to construct the recurrence matrix. As such, embedding dimension and delay parameters are not applied in this formulation; the recurrence condition is based solely on the absolute distance between scalar observations (Eq. 1) in [1]. Consequently, the only tunable RQA parameter in our workflow is the recurrence threshold  $\epsilon$ . The recurrence threshold ( $\epsilon$ ) was set to 3 dB. This choice has now been explicitly stated in the Methods section.

The RQA TREND threshold of  $-1.28$  ( $1/\text{timestep}$ ) with Timestep being 6-Days for Sentinel-1, was used to distinguish between stable and disturbed pixels. We agree with the reviewer that this threshold is not explicitly reported in Cremer et al. (2020). In our case, the threshold was derived empirically from the 5th percentile of the TREND distribution observed in a large reference dataset across Germany, representing extreme negative TREND behavior associated with structural change. While this choice is necessarily heuristic, it is grounded in observed data distributions rather than being arbitrarily selected. We have clarified this in the revised manuscript and corrected the text to explicitly state that pixels with RQA-TREND values above  $-1.28$  (unitless) were considered to show no significant change, as correctly noted by the reviewer.

We acknowledge that the selected TREND threshold can influence the spatial coherence and patchiness of detected disturbance areas. To assess the robustness of our threshold selection, we evaluated the optimal threshold for each disturbance type (DCA\_ID) by maximizing the F1 score, which balances Precision and Recall. The resulting best thresholds per DCA\_ID were: Bark Beetle  $-2.78$ , Defoliators  $-0.78$ , and Wind  $-2.78$ . These results demonstrate that the optimal threshold differs between disturbance types.

To allow a single, medium threshold applicable across all DCA\_IDs, we retain our selected threshold of  $-1.28$ , which falls within the overlapping range of the optimal thresholds across the different disturbance types and evaluation metrics (Precision, Recall, F1). This choice ensures a reasonable compromise between detecting true disturbances and limiting overestimation. Our analysis shows that this medium threshold generally performs well, although it is somewhat restrictive for fully capturing defoliator disturbances, which represents a limitation of the current approach.

Importantly, this analysis highlights that, in principle, there is no universal “change-detection-fits-all” threshold. Different disturbance agents exhibit distinct spectral and spatial characteristics, leading to varying optimal detection thresholds.

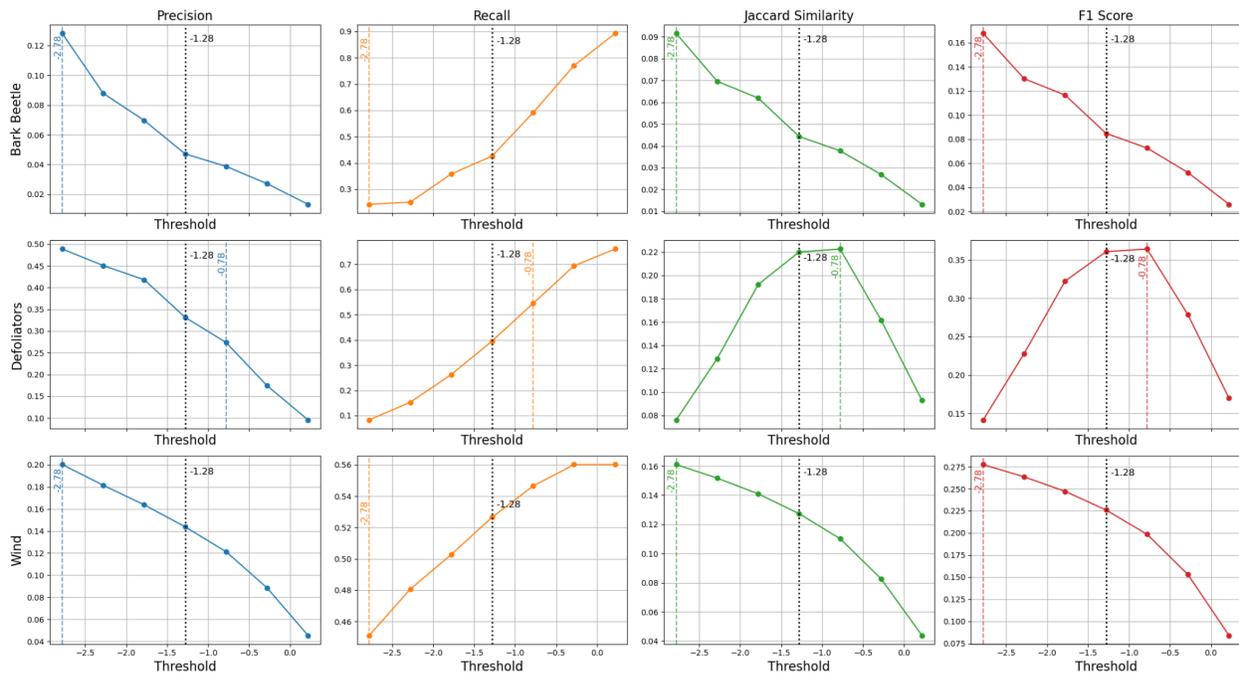


Figure R1.6: Threshold analysis for Bark Beetle, Defoliators, and Wind (top to bottom). Columns show the evaluated metrics. The x-axis represents the tested thresholds. The black dotted line indicates the selected threshold of  $-1.28$ , while the colored dashed lines represent the F1-maximizing threshold for each disturbance type.

We now also explicitly acknowledge this dependency on the threshold and the fact that there is no universal “change-detection-fits-all” threshold as a methodological limitation in the separate Discussion Section 5.3. *Methodological Limitations* and added Figure R1.6 to the Appendix of our manuscript.

For the computation of RQA metrics, we calculated one recurrence plot and one TREND value per pixel using a fixed two-year time window rather than a moving window. We do not agree that applying a moving window would necessarily allow for a more precise estimation of disturbance timing. However, if the editor deems it appropriate, we could explore such a temporally adaptive

framework in a small sample. Overall, we prefer to retain the fixed window size for this study. Investigating moving or adaptive window approaches is beyond the scope of the present work but is an important direction for future research.

The rewritten paragraph in the method section now states (changes in **bold**):

*“**Following Cremer et al. (2020), we applied Recurrence Quantification Analysis (RQA) to each pixel individually to quantify the similarity between time steps in the series. This algorithm is schematically illustrated in Appendix A2. In brief, the method plots the time series against itself to generate a Recurrence Plot (RP), where the main diagonal ( $x = 0$ ) shows a perfect match as each time point correlates with itself. The RP was computed following Cremer et al. (2020) using the RecurrenceAnalysis.jl package (RecurrenceAnalysis.jl). The recurrence condition is based solely on the absolute distance between scalar observations (Eq. 1) in Cremer et al. (2020), and the recurrence threshold ( $\epsilon$ ) was set to 3 dB.***

*Following Cremer et al. (2020), we computed the Recurrence Rate (RR) for each diagonal, defined as the fraction of recurrent points along the diagonal relative to its length. Plotting RR across diagonals produces a recurrence line, and the slope of this line reflects the temporal dynamics of the pixel: a slope of zero indicates no change, whereas negative or positive slopes indicate decreasing or increasing recurrence, respectively. **The RQA TREND threshold of  $-1.28$  (1/timestep) was used to distinguish between stable and disturbed pixels. This threshold was tested with a small subset analysis on the three different disturbance types. Pixels with RQA-TREND values above  $-1.28$  were considered to show no significant change. Values below this threshold were classified as indicating a structural change in that specific year. This process produced a binary raster, where each pixel indicated whether a structural change in forest cover occurred during the analyzed year.***

Additionally, the new Discussion section **5.3. Methodological Limitations** now addresses these Limitations explicitly:

*“Another methodological consideration relates to the selection of the RQA TREND threshold used to classify structural changes. In this study, we applied a threshold of  $-1.28$  (1/timestep) to distinguish between stable and disturbed pixels. While this value was empirically derived from reference datasets, it remains a heuristic choice and can influence the spatial coherence and apparent patchiness of detected disturbances. To further evaluate the robustness of this choice, we conducted a small disturbance-type-specific threshold analysis on each disturbance type. The results indicate that the optimal thresholds differ between disturbance agents, with best-performing thresholds varying for bark beetle, defoliators, and wind disturbances. This demonstrates that there is no single universally optimal threshold and that disturbance-specific spectral and structural characteristics lead to different detection sensitivities.*

*The selected threshold of -1.28 falls within the broader range of optimal values across disturbance types and metrics, representing a compromise solution that ensures consistent applicability across all disturbance categories. Visual inspection and comparative analyses indicate that this threshold captures meaningful disturbance structures while minimizing noise; however, results may vary if alternative threshold values are applied.*

*We explicitly acknowledge this dependency as a limitation of our workflow. The analysis highlights that, in principle, a universal “one-size-fits-all” change-detection threshold does not exist. We therefore recommend that future studies further explore the sensitivity of disturbance detection outcomes to the TREND threshold, potentially adopting disturbance-specific calibration strategies to better constrain its impact.”*

**RC8:** Concerning the IDS dataset, you mention “over 1000 selectable agents” (l. 162). This is surely a source of uncertainty; how can any image interpreter ever pick the right one out of so many choices under time pressure etc.? How many of these 1000 did you have to aggregate to get to the broad categories “wind”, “bark beetle”, “defoliator”? Please discuss. What is the connection between the > 1000 choices and Table A.3 (the transformation of the choices into DCA\_ID)?

Thank you for this question. We agree that the description of disturbance agents in the manuscript could lead to confusion and therefore warrants clarification.

In the IDS database, disturbance agents are organized into 28 broad groups (e.g., general insects, bark beetles, defoliators, chewing insects, general disease, biotic damage, domestic animals, and abiotic damage). Each group can be further subdivided into more specific subgroups or species-level categories, resulting in more than 1,000 distinct agent codes.

Disturbance agents are encoded using a five-digit numerical code, in which the first two digits identify the major disturbance group and the remaining digits specify subgroups or individual species. For example, bark beetles belong to the major group coded as 11\*\*, where the generic bark beetle category is represented by code 11000 (bark beetles, *Scolytinae*). If surveyors identify a specific species, more detailed codes are used, such as 11007, which represents the Douglas-fir beetle (*Dendroctonus pseudotsugae*). Despite this increased taxonomic specificity, these agents remain within the same major group, bark beetles. Comparable hierarchical subdivisions exist for other disturbance categories, including abiotic subgroups such as drought or wind.

Before each new flight season, Quick Key lists are prepared that summarize the disturbance agents, host species, and damage types most commonly expected within a given state, region, or survey area, and these are provided to the surveyors. In addition, ground checks are conducted to support and validate aerial observations.

In the GIS Handbook of USDA [3], they write: “Regardless of survey method, ground checking should always complement data collection with real-time feedback to keep observers tuned and adjust future observations during operations. Ground checks should be done where core attributes are unknown, following any survey method. Ground checks should also be done where new damage types or agents are suspected, and always when regulated pests are suspected. The extent of ground checking is subjective but should be thorough enough that the observer is confident the ground checks are representative of the feature of interest in terms of spatial position and core attributes; adjustments to any features should be made prior to data submission.”

In our analysis, disturbance agents were aggregated to their highest hierarchical level (major group), except for wind, which (although formally a subgroup of abiotic damage) was treated as a separate category due to its distinct disturbance characteristics. Accordingly, the DCA\_IDs in Table A.3 represent these broad categories (Bark Beetle, Defoliator, and Wind) rather than the detailed IDS agent codes.

We agree with the commenter that surveyors' ability to distinguish disturbance agents at the species level represents an impressive level of expertise and observational detail. Therefore, the aggregation to the broader categories of bark-beetle, defoliator, and wind allows to reduce the uncertainty regarding the particular agent of each disturbed patch. To emphasize these rich details in the dataset, we revised the manuscript and added a table to the Appendix.

*“Disturbances are mapped as geo-referenced polygons that represent areas where damage was observed, though not all trees within a polygon are necessarily affected. For each polygon, surveyors record a range of attributes; those used in this study include the affected tree species, the causal agent, the survey year, the damage type (e.g., mortality, defoliation percentage, crown dieback, branch breakage, or other non-mortality damage), and the severity class, expressed as the percentage of live and damaged trees within the polygon (Very Light: 1–3% through Very Severe: >50%).*

*The IDS database allows interpreters to select from more than 1,000 fine-grained disturbance agent codes, which are hierarchically organized into 28 broader categories. To reduce interpreter uncertainty and ensure analytical consistency, these 28 broader categories from the IDS disturbance classification hierarchy were used; for example, 63 bark beetle–related subgroups are recorded in the category Bark Beetle (see Table A.4). Disturbance agents were aggregated to their highest hierarchical level (major group), with the exception of wind, which—although formally a subgroup of abiotic damage—was treated as a separate category due to its distinct disturbance characteristics. The resulting aggregated disturbance categories are reported as **DCA\_ID** and are listed in Table A.3, which documents the final set of 28 broader disturbance agents used in this study.”*

Table A.4: Disturbance Agents Categories and Subgroups

No.	Disturbance Agent	Number of subgroup
1	General Insects	16
2	Bark Beetles	63
3	Defoliators	210
4	Chewing Insects	32
5	Sucking Insects	80
6	Boring Insects	97
7	Seed/Cone/Flower/Fruit Insects	56
8	Gallmaker Insects	23
9	Insect Predators	5
10	General Diseases	0
11	Biotic Damage	4
12	Root/Butt Diseases	35
13	Stem Decays/Cankers	90
14	Parasitic/Epiphytic Plants	23
15	Decline Complexes/Dieback/Wilts	34
16	Foliage Diseases	79
17	Stem Rusts	15
18	Broom Rusts	6
19	Terminal, Shoot, and Twig Insects	2
20	Root Insects	0
21	Fire	4
22	Wild Animals	19
23	Domestic Animals	6

24	Abiotic Damage	22
25	Competition	0
26	Human Activities	15
27	Multi-Damage (Insect/Disease)	4
28	Unknown	5

**RC9:** The annotated pdf attached to this review contains a further **31** comments, mostly rather specific ones. Please consider them as well.

We have carefully reviewed the annotated PDF supplement and identified **31** specific comments, which we have combined into a single answer when necessary. We have addressed each of these in line comments at the end of this response, providing detailed explanations and clarifications for each concern.

**RC10:** The paper has some strong points on being self-critical, indicating the limitations of the study, the problems with spatial inaccuracy and thus the necessity to introduce a buffer zone around the polygons, etc. It becomes obvious that the three disturbance categories are very different in their spatial structure. Rendering IDS and S1DM truly comparable for bark beetle and defoliator attacks is a long way to go, as Figure 4 demonstrates.

We thank the reviewer for the kind comment. We have made every effort to be transparent about the limitations of our study, including the spatial inaccuracy of the data and the resulting need for buffer zones around polygons. We agree that the three disturbance categories differ substantially in their spatial structure, and as Figure 4 shows, fully rendering IDS and S1DM comparable for bark beetle and defoliator events remains a significant challenge. Our results demonstrate that Sentinel-1 offers a promising avenue for improvement, but this work represents only the first step toward fully integrating these datasets.

**RC11:** A rather tricky issue is the timing of the onset of a disturbance; here, the authors go to a very coarse resolution of even more than one year, indicating that “online detection” of new damages is impossible. This is really a pity, since the strength of S-1 (and also S-2 for that matter) is short revisit cycles, with the potential to detect emerging attacks early and potentially act accordingly, very relevant for ecosystem managers! (See Jamali et al. 2023 for an approach using S-2). As the setup is now, the S1DM is for documentation of past events only.

We agree with the reviewer that this is a limitation of our current approach. While Sentinel-1 and Sentinel-2 offer short revisit cycles that could, in principle, enable early detection of emerging disturbances, our reference dataset (IDS) is only available at an annual resolution, and with an uncertainty of ca.±0.7 years (rounded to ~1 year) with a standard deviation of ±3.7 years (~4

years) compared to FIA, as discussed in Eifler et al. (in press). Furthermore, since the goal of our study is to develop a more refined reference dataset for training and benchmarking satellite-based disturbance models, it is focused on documenting past events. We note, nevertheless, that even with its annual resolution, S1DM tends to detect some disturbances earlier than IDS, particularly bark-beetle. Therefore, we hope our dataset will support the development of real-time or near-real-time detection algorithms in the future.

**RC12:** The last sentence (l. 649f.) is talking about an “alternative to manual labeling”; ironically, you are judging the quality of the S1DM as compared to IDS based on a third dataset which was obtained by manual labeling. The “fully automatic forest disturbance classification methods” (l. 645) are still to be developed.

We acknowledge the reviewer’s critical perspective on these sentences. We would like to clarify that PlanetScope manual labels are not used as training data for the S1DM workflow. Instead, they serve as an independent quality-control dataset to assess whether our fully automatic method correctly captures disturbance outlines. Such quality control is necessary because the original IDS dataset contains spatial and temporal uncertainties (e.g., Eifler et al. (in press); Coleman et al., 2018), and although the Sentinel-1 method has been developed, published, and validated (Cremer et al., 2020), combining these two sources does not eliminate the need for independent verification. Therefore, manual labeling was used for validation, even though the S1DM pipeline is fully automated.

The changes suggested still add up to “minor revisions” only, this is regarding the text. The sensitivity analysis for the TREND threshold and the selection of PlanetScope/S1DM but NOT IDS damaged areas require additional work however.

## Reference

Jamali, S., P.-O. Olsson, A. Ghorbanian and M. Müller (2023). "Examining the potential for early detection of spruce bark beetle attacks using multi-temporal Sentinel-2 and harvester data." *ISPRS Journal of Photogrammetry and Remote Sensing* 205: 352-366.  
<https://doi.org/10.1016/j.isprsjprs.2023.10.013>

## Supplement:

**L12: acronym not introduced before**

Thank you for noticing. We have now introduced the acronym appropriately at first use.

**L42: i.e. being reported?**

Thank you, we have updated this sentence. It now reads:

*“Coverage of reports on severe weather disturbances included 50% of forest areas in North America, 86% in Europe, and just 8% in Asia.”*

**L99: Remove "drought"**

Thank you for noticing, this is a fragment of previous drafts and was removed.

**Fig1: what about disturbances > 15 km<sup>2</sup> in size? and where do the 15 km<sup>2</sup> come from?**

We addressed this concern in **RC5**, where we explained our intentions for the 15 km<sup>2</sup> threshold. In this case, we keep the original Figure Caption but added the following clarification in Section 3.1 IDS Preprocessing:

*“Less than 1% of the disturbances exceed 15 km<sup>2</sup>, but these rare events cover unrealistically large and unspecific areas (mean size ~121 km<sup>2</sup>) and were therefore removed to avoid introducing bias and uncertainty into our analysis.”*

This addition makes our handling of extremely large disturbances explicit and justifies their exclusion from the analysis.

**L162: how many of them are attributed correctly?**

According to Coleman et al. (2018), the overall accuracy for identifying damage type was 97%, with a kappa statistic of 0.95. Their analysis was based on a subset of 567 aerial survey polygons from 2012 to 2014 that were ground-checked.

As explained in **RC8**, we revised the sentence to clarify the mapping approach and simplified the description of disturbance agents, replacing “over 1,000 selectable agents” with “28 broader disturbance agents.” The new sentence reads:

*“ Disturbances are mapped as geo-referenced polygons (or occasionally points) that represent areas where damage was observed, though not all trees within a polygon are necessarily affected. For each polygon, surveyors record the tree species affected, the causal agent (with 28 broader disturbance agents), the damage type (e.g., mortality, defoliation percentage, crown dieback, branch breakage, or other non-mortality damage), and the severity class, expressed as the percentage of live and damaged trees within the polygon (Very Light: 1–3% through Very Severe: >50%).”*

**L166/Tab1: make sure you really mean km<sup>2</sup> here, not m<sup>2</sup>. This largest polygon would cover almost 9% of the whole forested area of region 8.**

The reviewer’s comment overlaps with **RC6**, which we have already addressed.

As also discussed in our response to **RC5**, forest disturbances reported by IDS can occasionally span extremely large areas, in some cases covering several percent of the total forested region. As these very large polygons are likely attributable to human error or overly generalized mapping, we excluded them by restricting the analysis to disturbances smaller than 15 km<sup>2</sup>.

L191: in time probably, not in space?

Exactly — in this context, *activity windows* refer to a temporal window, not a spatial one. It indicates the specific time frame during the year during which a biotic disturbance (e.g., insect defoliation or bark beetle attack) occurs. The revised sentence now reads clearly:

“Biotic disturbances, such as defoliators and bark beetles, occur within specific **temporal** activity windows that must be captured accurately through timely ADS.”

L263: this is repetitive

Agreed. We have removed the sentence “Furthermore, our analysis focused on three specific types of disturbances: wind, bark beetles and defoliators.”, as this point is explained later in the manuscript.

L269: deleted ‘type’

Thank you. We agree and have deleted the word “type”.

L278: that's drastic difference, what is the rational behind it?

We addressed this concern in **RC5**, where we explained our decision on the 15 km<sup>2</sup> threshold, and updated the manuscript in Section 3.1, IDS Preprocessing. Please refer to **RC5** for the revised text.

L295-296: since the pixels are small (20 m x 20 m), the 30% threshold might exclude quite a bit of low-density forest. According to FAO, the threshold is 10% (but over an area of 0.5 ha, or 12.5 S1 pixels) <https://fra-data.fao.org/definitions/fra/2020/en/tad>

We already addressed this concern in **RC3**, where we reran the analysis using a less restrictive 10% Tree Canopy Cover (TCC) threshold as suggested by the reviewer. The results were very similar to the original analysis using 30% TCC, with only minor changes in the number of detected disturbance events and no change in the relative loss percentages. Given this, we decided to retain the 30% threshold for the main analysis to provide a more conservative and reliable estimate of forested areas.

Please refer to **RC3** for the revised text.

L316. no, this has to be the magnitude (absolute value) of the RQA trend; i.e. all trend values in the intercal [-1.28; 1.28] are considered as no change. Where does the 1.28 come from? What was the embedding dimension used? What was the recurrence threshold  $\epsilon$ ? Need more details here

This comment has been addressed in **RC7**.

L320-329: Why do you focus on these only?

The comment refers to the sentence in the manuscript: “*To focus on enhancing disturbance information rather than detecting new events, we only considered S1CD elements located within a buffer around IDS polygons, excluding all areas outside both the buffer and the flown survey paths.*”

We focused on these areas because the goal of this study is to refine the spatio-temporal attributes of existing disturbance datasets, rather than to detect new disturbances. Detecting new disturbances does not yet allow for reliable attribution of disturbance agents, which is ongoing work. Attributing disturbance types across all forest patches at a large scale is currently not feasible. Previous studies (Senf & Seidl, 2021; Viana-Soto & Senf, 2024) have focused on limited subsets of disturbance types, such as combined wind–bark beetle events, fire, and harvest. However, comprehensive disturbance attribution remains challenging—particularly when distinguishing among insect disturbance types (e.g., bark beetles and defoliators) and between insect- and wind-driven disturbances.

In our work, we focus on these disturbance types to enhance their spatial and temporal accuracy in legacy datasets, creating a high-quality reference that other researchers can later use to train classification algorithms and include for new events detected by remote sensing. By focusing on IDS-buffered areas, we ensure this enhanced dataset is reliable and suitable for future modeling beyond IDS-covered regions.

We revised the sentence in the manuscript.

*“To focus on **refining the spatio-temporal attributes of known disturbances**, rather than detecting new events, we **considered only** S1CD elements located within a buffer around IDS polygons, excluding all areas outside both the buffer and the flown survey paths.”*

## 2 Comments: L387 & L391: Removed Jaccard similarity

We are unsure why the Jaccard similarity was suggested for removal, as it is an important metric for understanding our method. To improve clarity, we have revised the first mention to “Jaccard similarity index” and retained it throughout the manuscript.

## Figure5: a)1: any explanation for this peak in the pdf for both IDS and S1DM at around 5 km<sup>2</sup> ?

We thank the reviewer for this important observation.

An analysis of IDS data without removing compound events shows that the peak in the probability density function around 4-6 km<sup>2</sup> for wind disturbances in 2020 is associated with unusually large events, primarily concentrated in a specific region of Louisiana, between Houston and New Orleans. Many of the polygons in this region are square-shaped, suggesting that part of the observed pattern may reflect mapping artifacts rather than the exact spatial extent of disturbances.

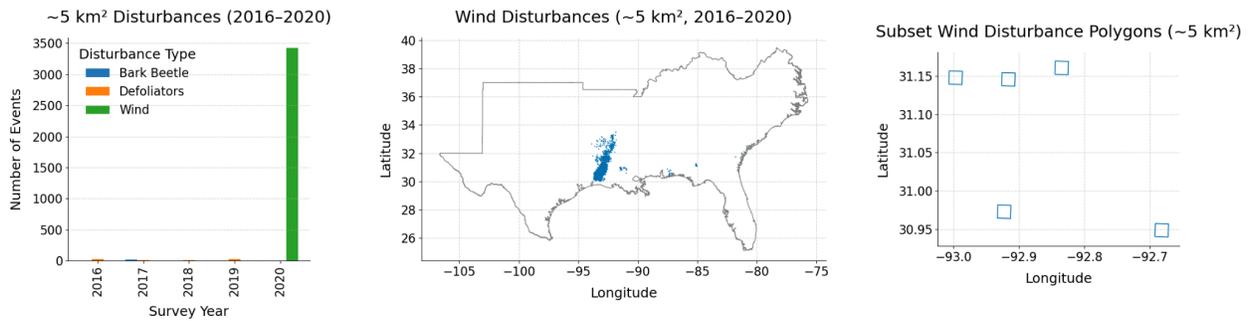


Figure R1.7: Left: Bar plot showing the number of disturbance patches with areas between 4–6 km<sup>2</sup>, grouped by disturbance type (bark beetles, defoliators, wind) and survey year. Bar colors indicate disturbance type, and the y-axis represents the number of disturbance polygons. Center: Spatial distribution of all wind disturbance polygons with areas between 4–6 km<sup>2</sup> detected between 2016 and 2020 across the study region. Right: Detailed view of a subset of wind disturbance polygons (4–6 km<sup>2</sup>) detected in 2020, with individual polygon boundaries (blue) shown to illustrate disturbance shape.

There are two main factors contributing to this pattern:

1. Actual large disturbances: In 2020, the region experienced major wind damage from several tropical cyclones, including Tropical Storm Cristobal, Hurricane Marco, Hurricane Laura, Hurricane Delta, and Hurricane Zeta. In particular, Hurricane Delta caused extensive forest damage, producing naturally large disturbance areas that contributed to the large wind-driven forest destruction in the data IDS and S1DM.
2. Mapping artifacts due to COVID-19 restrictions: The USDA modified its aerial survey protocols in 2020 because of pandemic-related restrictions. Standard aerial flights were reduced or not conducted, and alternative methods, including ground checks and imagery interpretation, were used instead. As a result, many IDS polygons in this region are square-shaped and approximately 5 km<sup>2</sup> in area, reflecting changes in survey methodology rather than actual disturbance extents.

Together, these factors explain the peak at ~5 km<sup>2</sup> in Figure 5: it arises from both true large-scale wind disturbances and methodological artifacts arising from COVID-affected survey procedures, producing unusually uniform polygon sizes while still capturing the impact of these destructive storms.

We have added the following information in the results and discussion sections of the manuscript.

**Results:** “Figure 5a shows a pronounced peak in the probability density function at approximately 4–6 km<sup>2</sup> for wind-related disturbances, which is consistently observed in both the IDS and S1DM datasets.”

**Discussion:** *“The peak at approximately 5 km<sup>2</sup> for wind disturbances reflects a combination of genuine large-scale wind damage and methodological artifacts in the IDS data. In 2020, several major tropical cyclones (including Cristobal, Marco, Laura, Delta, and Zeta) affected the region between Houston and New Orleans, producing extensive areas of wind-disturbed forests. In addition, aerial survey procedures were modified due to COVID-19 restrictions, leading to the use of alternative mapping approaches. This resulted in more uniform, often square-shaped disturbance polygons of similar size, particularly in this region, which contributes to the observed peak. Despite this artifact, the mapped polygons still capture the spatial footprint of severe wind-induced forest disturbances.”*

12 Comments in the following paragraph (L431–436):

Line 434: Removed: 50% of S1DM Wind events being 0.1 km<sup>2</sup> and 90% being 0.03 km<sup>2</sup> smaller than IDS.

434-435: not “unlike” if numbers in Table 2 are correct

L435-436: Deleted the red underlaid parts in the following sentence: For Defoliators, **more than 50%** of S1DM events **are** 0.01 km<sup>2</sup> larger than **their** IDS counterparts, and 90% of events **are** nearly twice as large in S1DM as in IDS.

Thank you very much for your careful review and for catching this inconsistency. You are absolutely correct: the values reported in Table 2 are correct, but the accompanying text had not been updated from a previous version. We have therefore completely reworded the paragraph (L431–436) to ensure full consistency with the table. The revised text now reads:

*“Overall, disturbance areas derived from S1DM closely match those from IDS, with only minor differences across disturbance types. Bark beetle disturbances show the strongest agreement, with nearly identical median areas (IDS: 0.0046 km<sup>2</sup>; S1DM: 0.0043 km<sup>2</sup>) and 90% of events in both datasets smaller than 0.08 km<sup>2</sup>.*

*For wind disturbances, median S1DM events are slightly larger than IDS (0.34 vs. 0.24 km<sup>2</sup>), while the 90th percentile areas are nearly identical (5.03 vs. 5.06 km<sup>2</sup>).*

*Defoliator disturbances tend to be larger in S1DM than in IDS, with similar median values (0.50 vs. 0.49 km<sup>2</sup>) but substantially larger upper-tail events, where the 90th percentile in S1DM is nearly twice that of IDS (6.33 vs. 3.68 km<sup>2</sup>).”*

L466: Deleted majority in sentence: “While a **majority** of events — 359 out of 899 (39.9 %) — are detected by S1DM in the same year as IDS (Lag=0), a substantial number fall within the ± 1 ±1- year window around the IDS detection year, with 140 events (15.6 %) detected one year earlier and 157 events (17.5 %) one year later.”

Absolutely correct! 39% is not a majority; rather, it is the largest proportion of events. We have revised the sentence to now read:

*“While the largest proportion of events — 359 out of 899 (39.9%) — are detected by S1DM in the same year as IDS (Lag = 0), a substantial number fall within the ±1-year window around the IDS detection year, with 140 events (15.6%) detected one year earlier and 157 events (17.5%) one year later.”*

**L518: That number is not provided in the reference**

The original sentence read: “In this case, the low sensitivity of S1CD to subtle and slow-onset disturbances might stem from the change detection threshold, set at **-1.28 (1/yr)** by the European Commission. Joint Research Centre. (2023).”

This is correct — the threshold was not reported in the reference. We updated the citation to Cremer et al. (2020, <https://ieeexplore.ieee.org/document/9186165>). The threshold itself was decided by us based on the method described in this paper and in correspondence with the authors. Accordingly, the sentence has been revised to:

*“In this case, the low sensitivity of S1CD to subtle and slow-onset disturbances may stem from the change detection threshold, which we set at  $-1.28$  (1/timestep) based on the method described in Cremer et al. (2020) and correspondence with the authors.”*

**L529: Three?**

Yes, thank you for noticing that. There are three disturbances, and the manuscript has been corrected accordingly.

References:

- [1] Cremer, F., Urbazaev, M., Cortes, J., Truckenbrodt, J., Schmullius, C., & Thiel, C. (2020). Potential of recurrence metrics from Sentinel-1 time series for deforestation mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5233–5246. <https://doi.org/10.1109/JSTARS.2020.3019333>
- [2] Cremer, F., Gans, F., Cortes, J., & Thiel, C. (2023). Mapping forest loss in Europe with Sentinel-1. In P. Soille, S. Lumnitz, & S. Albani (Eds.), *Proceedings of the 2023 Conference on Big Data from Space (BiDS'23) – From Foresight to Impact* (pp. 361–364). Publications Office of the European Union. <https://doi.org/10.2760/46796>
- [3] U.S. Forest Service. (2022). *Forest Health Detection Survey: GIS handbook and data conformity standards* (Version 1.0). U.S. Department of Agriculture, Forest Service.

- [4] Senf, C., & Seidl, R. (2021). Mapping the forest disturbance regimes of Europe. *Nature Sustainability*, 4, 63–70. <https://doi.org/10.1038/s41893-020-00609-y>
- [5] Viana-Soto, A., & Senf, C. (2025). The European Forest Disturbance Atlas: A forest disturbance monitoring system using the Landsat archive. *Earth System Science Data*, 17, 2373–2404. <https://doi.org/10.5194/essd-17-2373-2025>
- [6] Coleman, T. W., Graves, A. D., Heath, Z., Flowers, R. W., Hanavan, R. P., Cluck, D. R., & Ryerson, D. (2018). Accuracy of aerial detection surveys for mapping insect and disease disturbances in the United States. *Forest Ecology and Management*, 429, 301–311. <https://doi.org/10.1016/j.foreco.2018.08.020>
- [7] Eifler, L., Müller, F., & Bastos, A. (2024). Evaluating the consistency of forest disturbance datasets in continental USA. *EGUsphere* [preprint]. <https://doi.org/10.5194/egusphere-2024-3534>
- [8] FAO. (2020). *Global Forest Resources Assessment 2020: Main report*. Rome. <https://doi.org/10.4060/ca9825en>
- [9] Andresini, G., Appice, A., & Malerba, D. (2024). A deep semantic segmentation approach to map forest tree dieback in Sentinel-2 data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 17075–17086. <https://doi.org/10.1109/JSTARS.2024.3460981>
- [10] Bárta, V., Lukeš, P., & Homolová, L. (2021). Early detection of bark beetle infestation in Norway spruce forests of Central Europe using Sentinel-2. *International Journal of Applied Earth Observation and Geoinformation*, 100, 102335. <https://doi.org/10.1016/j.jag.2021.102335>
- [11] Hawryło, P., Bednarz, B., Wężyk, P., & Szostak, M. (2018). Estimating defoliation of Scots pine stands using machine learning methods and vegetation indices of Sentinel-2. *European Journal of Remote Sensing*, 51(1), 194–204. <https://doi.org/10.1080/22797254.2017.1417745>
- [12] Gibson, R., Danaher, T., Hehir, W., & Collins, L. (2020). A remote sensing approach to mapping fire severity in south-eastern Australia using Sentinel-2 and random forest. *Remote Sensing of Environment*, 240, 111702. <https://doi.org/10.1016/j.rse.2020.111702>