Response to Reviewer #2

The authors wish to thank the anonymous referee #2 for his/her constructive comments and suggestions, which have substantially improved our manuscript. We have considered the recommendations carefully and revised the manuscript accordingly. We made significant changes to the manuscript to further clarify the methods and validation process.

Please, find a detailed response to the comments below. Original comments are in black and our response in blue. Line numbers correspond to the track-changes file.

***

General comments

It is certainly a great paper, describing a huge processing effort to identify long-term changes in forested areas in Europe. The generated dataset will certainly be very useful to monitor changes throughout the European territory and identify the main drivers which should lead to better conservation of European forest patrimony.

I have a few general remarks and a longer list of specific comments that may help the authors to improve the current version of the manuscript, but my general impression is very positive.

First, it is not clear why the authors selected pixel instead of plots as reference data for calibrating and validating the outputs. Single pixels are difficult to characterize and may be affected by geolocation errors, particularly when detecting changes. In this regard it is not clear why the authors start commenting at length about this pixel reference database, but then make several analyses at patch level when they do agent attribution. It seems inconsistency here, which would be convenient to clarify. It would be convenient to clarify the text now included in sections 2.2 and 2.6 and explain why two different approaches were necessary to classify and to assign agents.

We thank the reviewer for their comment, which pointed out a sub-optimal description of our methods. We used two distinct reference datasets for 1) annual disturbance mapping, and 2) disturbance agent attribution. To date, there is no reference dataset that fulfils all the requirements to calibrate and validate disturbance models per agents in a single step (i.e. a systematically collected, spatially explicit and covering all of Europe and all years since 1984). We therefore relied on two separate existing datasets to first calibrate and validate a pixel-based disturbance model and following calibrate a patch-based attribution model. Separating disturbance detection and attribution into two steps is quite common and has been done also in other studies (Kennedy et al., 2015; Sebald et al., 2021; Senf et al., 2017; Shimizu et al., 2019; Stahl et al., 2023).

The first step focuses on annually mapping disturbed and undisturbed pixels within forested areas relying on spectral changes from a given year (t_0) and the previous year (t_{-1}) at the pixel level. We did this mapping-step at the pixel level, because the reference database used for model training exists at the pixel level and it would have been impossible to translate this pixel-based reference database into a patch-based database (we would need high-resolution imagery back to 1984 for doing this). While we acknowledge that single-pixel reference data may introduce potential geolocation uncertainties, we minimized these issues by employing robust preprocessing methods (i.e. spatial co-registration) and given the large number of samples (n = 20,084), it can safely be assumed that uncertainties in matching pixels to reference points will not affect the overall training success and validation precision.

The second step (agent attribution) was done at the patch level, because several patch-based predictors were derived from the maps (i.e. patch size and form). For doing this, patches need to be identified, for which a pixel-based classification is needed in the first place. Thus, doing both classification steps in one model would be impossible. Using patch-based predictors for agent attribution has been done in many previous studies (Oeser et al., 2017; Sebald et al., 2021; Senf and Seidl, 2021b; Stahl et al., 2023) and builds upon ecological understanding on the distinct spatiotemporal dynamics of natural disturbances. To better explain the two processing steps (disturbance mapping at the pixel level and agent attribution at the patch level), we expanded the information on how each reference dataset was created and more prominently highlight the difference between the two classification steps:

> Lines 84-89: "*The overall workflow behind the EFDA contains six processing steps summarized in Figure 1 and described in full detail in the following sub-sections: First, building a consistent Landsat data cube for Europe. Second, compiling a reference dataset on forest land use and forest disturbances. Third, creating a consistent forest land use mask. Fourth, developing a classification-based approach to detect disturbances annually at the pixel level. Following this, validating the forest land use mask and disturbances maps based on an independent reference sample. Fifth, identifying disturbance patches and the likeliest agent of disturbance at the patch level. Lastly, creating a set of summary layers on forest disturbances and forest disturbance agents*."

> Lines 240-249: "*To identify the most likely agent of disturbances, we adapted an existing attribution algorithm (Sebald et al., 2021; Seidl and Senf, 2024; Senf and Seidl, 2021b) for the annual disturbance maps. The algorithm first detects disturbance patches by grouping pixels disturbed in the same year that are connected by an edge or corner using queen-contiguity. That is, the analysis is performed at the patch-level and not at the pixel level anymore, allowing us to derive patch-level predictors important for disturbance agent attribution (Oeser et al., 2017; Sebald et al., 2021; Stahl et al., 2023). To correct for timing errors in the disturbance map (e.g., a fire mapped over two years might appear as two separate fires), patches from consecutive years that share an edge are merged, with the disturbance year assigned based on a majority vote (see Senf and Seidl 2021b for details). For each patch, we generated a set of 18 predictors, including its size, shape, spectral characteristics, and surrounding landscape (see Table 2 for details and Figure A6 for further details on variable importance, as well as Seidl and Senf, 2024; Senf and Seidl, 2021b).*"

References:

Kennedy, R. E., Yang, Z., Braaten, J., Copass, C., Antonova, N., Jordan, C., and Nelson, P.: Attribution of disturbance change agent from Landsat time-series in support of habitat monitoring in the Puget Sound region, USA, Remote Sensing of Environment, 166, 271–285, https://doi.org/10.1016/j.rse.2015.05.005, 2015.

Oeser, J., Pflugmacher, D., Senf, C., Heurich, M., and Hostert, P.: Using Intra-Annual Landsat Time Series for Attributing Forest Disturbance Agents in Central Europe, Forests, 8, 251, https://doi.org/10.3390/f8070251, 2017.

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

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

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, https://doi.org/10.1016/j.jag.2017.04.004, 2017.

Shimizu, K., Ota, T., Mizoue, N., and Yoshida, S.: A comprehensive evaluation of disturbance agent classification approaches: Strengths of ensemble classification, multiple indices, spatio-temporal variables, and direct prediction, ISPRS Journal of Photogrammetry and Remote Sensing, 158, 99–112, https://doi.org/10.1016/j.isprsjprs.2019.10.004, 2019.

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

Second, it is not clear whether the authors used multitemporal changes in the detection of disturbances or classified single years without any consideration of previous years. There is a comment in lines 214-215 that indicates that double disturbances were removed in post-processing, but it would have been simpler to classify with both current year and previous year indices, so the change would have been included explicitly in the classification, thus removing double detections.

To identify disturbed and undisturbed pixels annually, we indeed use the spectral information from the previous year (t-1) and the target year (t0) (see section 2.4. lines 200-203). We thus do multi-temporal classification. The case presented in Figure 4 is an example of the illogical sequences that may appear when relying on year-to-year changes. In the example shown (sequence of Disturbed-Undisturbed-Disturbed), the second falsely classified disturbance is difficult to prevent even when considering additional previous years, because there is a negative change not only with respect to the previous year, but also with the year t-3 and backwards. Further, including years back in time would have restricted the time series starting point to early 1990s.

My third comment refers to the confusion that appears in some paragraphs of the paper between validation and uncertainty characterization. They are not the same, the former meaning the level of agreement with the reference data, and the latter the probability of being sure that the classification was properly done. This is clear when authors use the probability of the RF classifier, but they do not mention uncertainties derived from the pre-processing or the compositing periods or even the lack of observed areas. In addition, in lines 404-05 it is indicated that the authors aimed to provide a full characterization of uncertainty, but they talk about validation results.

We thank the reviewer for this comment, and we agree that we wrongly wrote that we provide a full quantification of uncertainty, which – in fact – we do not. We provided a detailed assessment of classification errors and a comparison of our data products to existing products. We further provide classification probabilities derived from the Random Forest model, but we do not use those probabilities in later analyses. They are just provided as a decision support for user and can be used to apply more/less strict thresholds for disturbance classification (e.g. to reduce commission errors) (see Figure B1 appendix). Therefore, we revised the manuscript to strictly refer to validation and not uncertainty quantification throughout the manuscript.

Finally, the spatial validation is well explained but nothing is very little is said about the temporal reporting accuracy, that is how longer after the actual disturbance your product detects it. Include this on your validation approach with a proper consideration to the impacts of your annual compositing method.

We assessed temporal accuracy in lines 358 to 367 showing annual commission/omission errors (see Figure 8) and detailed confusion matrices of disturbance detection assessment per period can be found in Table B2 of the Appendix. We further showed in Figure 9 the validation of the disturbance year by comparing the estimated disturbance year versus manually interpreted year, with a mean absolute error of 1.91 years, meaning that we detect disturbances within the next 2 years after and many fall into the same year. This is already an improvement compared to previous assessments which reported more than 3 years (Senf and Seidl, 2021a). We are unsure what additional analyses is proposed by the reviewer.

Beyond the results section, we further discuss sources of errors in the discussion section, and specifically how our compositing approach contributes to the delayed detection of disturbances.

> Lines 461-462: *"Further, image availability and observation gaps (e.g. ETM+ SCL-off problems) can also contribute to missing disturbed areas in large disturbance patches"*.
>
> Lines 482-485: *"We further found that omission errors were caused by a delayed detection of disturbances (i.e. disturbances detected one year after the reference). The detection of disturbances with a delay of one year is a common problem when working with annual summer composites, where disturbances can occur in the same year but after the compositing date (i.e. a pixel is selected from August, but the disturbance occurs in November and will thus only be detected the next summer)."*

Specific comments

Lines 114-15. Please further clarify the methods used for the annual compositing. The reader should not need to read another paper to understand yours.

Thanks for the suggestion. We reviewed the description and added more information on the compositing process.

Lines 115-121: "*For creating the best available pixel composites, we selected for each pixel the best observations based on a parametric weighting scheme established in previous research (Griffiths et al. 2013). Observations were ranked per pixel according to distance to clouds and cloud shadows, haze opacity and proximity to a predefined target date (1st of August; Figure A1- Appendix). Each of these factors is assigned a specific weight in the scoring algorithm and only high-quality pixels are considered, i.e. observations with very low* *cloud or haze score are discarded. The observation with the highest cumulative score is then selected as the best pixel for each location, ensuring both spectral and temporal consistency in the composites."*

Lines 118-19. Gap-filling with a previous year is quite controversial. In my opinion, it would have been more convenient to fill the date with a pre-summer image or just keep the unobserved data as unobserved and more the analysis to the following year.

Thanks for this comment. Indeed, gap-filling can introduce problems. That said, in most cases the trajectory is stable over time and filling the missing observation with the previous year will not change the outcome (i.e. a stable forest pixel). The worst case is that a disturbance happens exactly in the year of a missing observation. In this case, the disturbance will be detected one year later (as the stable forested conditions from the year before will be superimposed). This drawback is, however, minor compared to having many missing observations in the final map, because the bitemporal classification does not work with missing observations. Using a pre-summer observation is also challenging, as in many regions there will still be snow before summer, introducing large spectral changes and thus a high risk of false positives. Hence, our filling only "delays" the analysis into the next year, as proposed by the reviewer. Consequently, we did not change our approach.

Line 189: Not clear to me why you used a single forest mask for the whole period. What happens with areas that converted from crop to forest or from forest to urban during your time period? You indicate in line 191 that "All non-forest land use pixels were excluded from following analyses", non-forested when, anytime in between 1983-2023. How about areas of new urbanization in that period?

Thank you for this comment. First and foremost, it is important to understand that we aimed at creating a land use mask instead of a land cover mask. For forest areas, land cover can be dynamic, while land use remains static. A harvested area is still considered a forest, even though temporary unstocked. Using a forest land cover map (i.e. existing maps) would have masked out those temporally unstocked areas, leading to the exclusion of many disturbed areas. Further, creating a reliable land use mask allowed us to mask out all unrelated land uses (e.g. permanent agricultural areas, urban areas, waters, etc.), even though they might have trees as land cover (e.g. in urban areas).

We approach the problem of land use mapping by applying a multitemporal classification to map forest and non-forest land uses, leveraging the full time series of reference data (1985–2018) as input to the Random Forest model (see lines 190 onwards: "*we trained a random forest model with a stack of all annual best available pixel composites as input (i.e. 34 years with 6 spectral layers and 6 spectral indices, resulting in 408 features, see Table A1 and Figure A4*)". This method allowed us to classify each pixel into its predominant land use over the entire time frame, rather than being limited to a specific subset of years or a single year. This approach was first and foremost used to exclude areas that have been non-forest land use over the full temporal period. This approach also ensures that young forests of lower tree cover, areas of reforestation and temporally unstocked areas (e.g. forestry roads) will be included. We consider this multitemporal classification an effective solution compared to the use of a static forest mask, which would mask out recently disturbed areas that are temporarily unstocked.

We acknowledge, however, that this approach does not account for all types of land use changes, such as urbanization or afforestation events. Yet, forest disturbances related to land use conversions in Europe account for only 2.6% of all disturbances (Senf and Seidl, 2021b), and those will not cause any problems as the majority are related to infrastructure development (e.g. highways, parking lots, industrial areas) and show little spectral changes after disturbance (i.e. forest to non-forest land cover change). Problems would arise if a forest turned into cropland with year-to-year spectral changes after disturbance that would be falsely detected as disturbance. Forest to agricultural conversion is – however – rare in Europe. Afforestation areas will also be included in our mask, but those afforestation patches will likely not cause any false positives, because spectral properties are very different to disturbances (in fact, the opposite behaviour). Thus, while there will be uncertainty from the multi-temporal land use mask, it ensures the inclusion of all forests and does not mask out temporally unstocked and recently disturbed areas.

A year-specific land cover and thus dynamic land use map would be an important next step in understanding land cover dynamics across Europe. However, creating such a dataset is beyond the scope of this study.

Line 198, you include here the use of NDVI and Disturbance index but none of the two are quoted in table 2. Please clarify why?

We used a different set of predictors for the annual disturbance mapping and for the agent attribution. To provide clarity, we included a figure in the appendix (Table A1) detailing the indices used as inputs for the forest land use and forest disturbance classifications (see also Lines 190–194 and Lines 202–205 and Figures A4 - A5 for feature-space information on the predictors). Table 2 summarises the variables used only in the disturbance agent attribution model. For the agent attribution, we used the NBR and Tasseled Cap components because they showed higher predictive importance for the forest disturbance detection (see Figure A6).

Line 239-240: Not sure why do you consider that a burned area cannot occur in two consecutive years in a neighbour region. In Europe is less frequent but in other ecosystems is quite common.

We agree that fires can occur in two consecutive years but is very rare Europe – as pointed out by the reviewer. As we focus on Europe in our study, we consider our wording appropriate. The reason we merged patches from consecutive years is that temporal errors in our disturbance map can cut large disturbance events into two pieces. This happened most often with large fires that burnt over the summer and thus over the temporal cut-off for the compositing algorithm. The problem became most obvious with SLC-off strips, showing gaps in the fire pattern that are filled in the following year (see Figure 11a). We thus consider the potential error introduced by this processing step as minor to the benefits. We note that a similar technique is also used by Acil et al. (2024) and Hermosilla et al. (2015).

Acil, N., Sadler, J. P., Senf, C., Suvanto, S., and Pugh, T. A. M.: Landscape patterns in stand-replacing disturbances across the world's forests, Nat Sustain, 8, 86–98, https://doi.org/10.1038/s41893-024-01450-3, 2024.

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., and Hobart, G. W.: Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using Landsat-derived time-series metrics, Remote Sensing of Environment, 170, 121–132, https://doi.org/10.1016/j.rse.2015.09.004, 2015.

Line 241: How patches were created? What is the impact of your compositing period to create those patches?

We created individual disturbance patches using queen-contiguity (i.e. considering the 8-connected neighbours), combining all pixels disturbed in the same year sharing either an edge or node. We clarified that now in Lines 244-249: *"The algorithm first detects disturbance patches by grouping pixels disturbed in the same year that are connected by an edge or corner using queen-contiguity (…) To correct for timing errors in the disturbance map (e.g., a fire mapped over two years might appear as two separate fires), patches from consecutive years that share an edge are merged, with the disturbance year assigned based on a majority vote".*

The compositing period can impact the formation of patches (see comment before), and we discuss this in detail in the discussion (see Figure 11). Lines 479-482: *"We further found that omission errors were caused by a delayed detection of disturbances (i.e. disturbances detected one year after the reference). The detection of disturbances with a delay of one year is a common problem when working with annual summer composites, where disturbances can occur in the same year but after the compositing date (i.e. a pixel is selected from August, but the disturbance occurs in November and will thus only be detected the next summer)."*

Additionally, we show here an example of a fire in Portugal that was mapped over two years due to the composite dates (DOY before and after the fire event occurred):

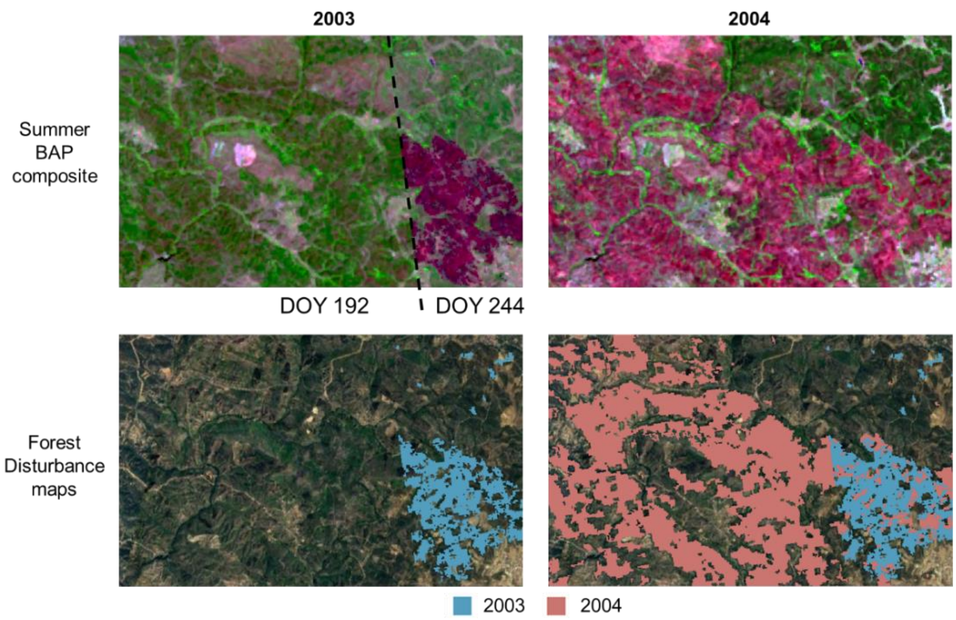

Figure 5. It is not clear what the colours mean, as three different colour legends are included above.

We edited the figure for better clarity and placed the legends below each corresponding layer for the 3 examples.

Figure 7. Include country abbreviations (perhaps in Table 1)

We included country abbreviations in Table 1.

Line 350. Indicate why accuracy is reduced before 2000. I assume it will be a question of image availability, but this should have been considered when computing the uncertainty, and therefore the estimations done with broader confidence intervals.

Accuracies are lower before 2000 due to higher commission errors related to image noise (i.e. remaining clouds, difficult illumination conditions) and omission errors are also slightly higher for that period, likely due to lower image availability (though this is hard to test). We included the following details in the results:

Lines 358-362: *"The high commission error prior to 2000 was driven by the very early years of our time series (<1990), where commission error was on average 22.5% (compared to 16.4% between 1990 and 1999), likely due to higher noise in the underlying data. Omission errors showed a less clear pattern, with also higher omission rates in the 80s (23.3%), likely due to lower image availability, but less pronounced differences before/after the year 2000 (22.5% and 20.5%, respectively)."*

How the image availability should be reflected in the confidence intervals is unclear to us. It is important to note that our study does not involve any kind of area estimation (the numbers shown in Figure 6 are raw pixel counts). There are thus no confidence intervals given for the disturbance area estimates.

Line 394. Include a comment of the disturbance return times in the paper, as you indicate is a novelty of your product.

We included a comment on the disturbance return times in lines 406-409. "*We found, for example, return times of 15-20 years in recurrent fires in the Mediterranean or in short rotation plantations across Europe (e.g. Hungary, Sweden, Netherlands), and further temporal interactions between harvest and wind disturbances in the Gascony area in France (i.e. wind disturbance following a harvest event).*"

Line 426. You are right that is difficult to compare you map with static land cover. Maybe you can make a comparison with the LC CCI product, which provides annual changes, even though at coarser resolution product than yours.

Although a comparison of our forest land use map would be interesting (we compared already with country forest cover form Forest Europe 2020), we decided to focus on comparing our product to maps that specifically map forest disturbances/changes, and that provide information at the same spatial resolution as our product.

Further, several studies have highlighted limitations in using the CCI-LC product for annual land cover changes analysis. For instance, Li et al (2018) used the annual ESA CCI land cover maps (1992–2015) for gross and net land cover changes in the main plant functional types and noted discrepancies when compared with other datasets, suggesting caution in their application for detailed land cover change analyses, particularly due to its spatial resolution limitations (1km).

Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., ... & Peng, S. (2018). Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015). Earth System Science Data, 10(1), 219-234.

Figure 11 shows most likely ETM derived strips in the fire area of central Portugal. This was not commented in the paper and it should, as it creates an evident noise in the output. Please include the map location of the three sites.

Figure 11(a) shows fire events in Portugal that were recorded across two different years in EFDA. The pattern observed in the year 2003 due to ETM+ data corresponds to true disturbed pixels, with the disturbance patch being fully completed in the subsequent year 2004. We acknowledge this as a valid observation; however, it is important to note that this is distinct from ETM+ artifacts, which can introduce false disturbances. This example nicely shows the problem of using annual composites for disturbance mapping (see also discussion above) and is a great example of why we group consecutive and neighbouring disturbance events together before doing the patch-based agent attribution.

We edited Figure 11 to include the map location of the three sites for better clarity and commented on the ETM derived strips in Lines 461-462: *"Further, image availability and observation gaps (e.g. ETM+ SCL-off problems) can also contribute to missing disturbed areas in large disturbance patches."*

Line 461. You indicate the potential problems to detect grading changes. Have you explore the use of spectral unmixing approaches?

Spectral unmixing is indeed a valuable technique used to disentangle changes in vegetation cover types within a pixel. It could certainly be a promising approach for detecting more subtle disturbances, particularly those that do not produce a clear opening of the canopy (e.g. gradual defoliation). Testing spectral unmixing is, however, beyond the scope of this study. We like to note, though, that we aim at exploring spectral unmixing in a follow-up project.

Line 485. You indicate potential problems with several years with poor image availability. Have you tried to use MODIS 250m resolution reflectances for those years? A few methods that compare Landsat-MODIS to merge reflectances are available in the dedicated literature.

Incorporating MODIS reflectance for years with limited image availability could be an interesting avenue to explore. However, most of the gaps in our time series occur in the years before the 2000s, for which MODIS is not available (see Figure A2 showing the annual percentage of no-data pixels for best available pixel composites). Further, this approach would introduce additional limitations. For instance, MODIS lacks SWIR bands at a 250m resolution (only available at 500m - 1km), which are crucial for detecting forest disturbances due to its high sensitivity to forest structure and vegetation water content (see importance of

NBR and TCW in Figure A6). Moreover, merging datasets with differing spatial resolutions would introduce further uncertainties to the disturbance detection, especially in Europe where disturbances are small, and forests are highly fragmented.

Line 505. I agree on the need of having publicly access validation databases. May be you can refer to the BARD, which is a global (not just European, but with European sites) validation database of burned perimeters (Franquesa et al. 2020). References:
Franquesa, M., Vanderhoof, M.K., Stavrakoudis, D., Gitas, I.Z., Roteta, E., Padilla, M., & Chuvieco, E. (2020). Development of a standard database of reference sites for validating global burned area products. *Earth Syst. Sci. Data 12*, 3229-3246.

Indeed. We incorporated examples of public databases in Lines 520-522: "*Further research should thus increasingly focus on generating reference data that can be used for remote sensing applications (e.g. Franquesa et al., 2020; Senf, 2019), and we urge authors to make their reference data openly available, ultimately leading to a better understanding of forest change across Europe.*"