Reply to Reviewer #1.

Legend:

-Reviewer's comments
-authors reply

We would like to thank reviewer #1 for their time. We are very grateful for their extensive and valuable input that contributed to the improvement of the manuscript.

This manuscript aims to find the most relevant predictors for forest damage in Europe at monthly to annual timescales, using a random forest approach applied to NDVI from the Advanced Very High Resolution Radiometers (AVHRR) as the predictand, and a range of variables from the ERA5 and ERA5-Land reanalysis as hydrometeorological predictors. I have several criticisms related with the concept of forest damage, methodology (e.g., the quality of NDVI dataset, the multicolinearity of predictors, the) and lack of clarity of results (low quality of the figures and no insufficient discussion) listed below that should be addressed before considering the paper for publication in Natural Hazards and Earth System Sciences journal.

Major suggestions/comments:

Introduction:
The title mentions forest low greenness events in Europe, but within the document the authors speak about forest damage. The concept must be introduced, namely in terms of biophysical and phenologic characteristics of the forests. For different readers and purposes, forest losses in terms of density, forest mortality or simples forest cleaning are included in the classification of 'forest damage' used in this paper. For instance, the measures for fire risk mitigation that include forest cleaning or measures to control invasive species are in the context of this work classified as 'forest damage' and they are in fact measures that aim to protect forest and biodiversity. I strong suggest to change the wording in the manuscript to other options, e.g., forest low greenness, losses of forest greenness or forest browning.
Thank you very much. We removed all the occurrences of "forest damage" and replaced them with "low greenness events" and "forest browning".

Another topic is the use of NDVI to capture forest damage. The use of satellite data to study vegetation health should also be included in the introduction, mentioned the studies that have used NDVI to evaluate the relationship between climate variability and vegetation dynamics. In particular, the choice of using NDVI and advantages and disadvantages for such choice should be provided.

Thank you for this valuable suggestion. In response, we have expanded the introduction to better contextualize the use of satellite data, particularly NDVI, for assessing vegetation health and forest damage. We now include references to key studies that have used NDVI to evaluate the relationship between climate variability and vegetation dynamics. Additionally, we clarified our rationale for selecting NDVI over other indices and datasets (especially kNDVI, SAR and NDWI), and addressed their limitations in the discussion section.
Introduction section:
"Forest canopy greenness, widely employed as a proxy for vegetation conditions, exhibits strong sensitivity to hydro-climatic variability across sub-seasonal to interannual timescales (Obuchowicz et al., 2023). Satellite-based remote sensing, particularly vegetation indices such as the Normalized Difference Vegetation Index (NDVI), enables a systematic and spatially consistent monitoring of canopy greenness across large regions, offering the capability for the detection and attribution of forest stress (Liu et al., 2013; Sun et al., 2021). For example, satellite-derived low NDVI values have been associated with severe canopy impacts during the 2013 and 2018 droughts in Germany (Brun et al., 2020; Buras et al., 2020; West et al., 2022). Hermann et al. (2023) introduced a pragmatic and systematic approach that advances this capability by explicitly linking NASA Terra MODIS observations to multi-annual meteorological records. Their methodology identifies low-NDVI events at a 50 km spatial scale across Europe, showing that reduced summer forest greenness is consistently preceded by anomalous temperature and precipitation patterns during the preceding months and seasons. Satellite imagery is a powerful tool for monitoring drought impacts, yet the wide range of available sensors, each with distinct characteristics, requires careful consideration when selecting the most suitable ones (Verbesselt et al., 2010; West et al., 2019). Among these, the Advanced Very High Resolution Radiometer (AVHRR) is unique in providing a continuous 40-year record (Weber et al., 2021; Dupuis et al., 2024; Barben et al., 2024). AVHRR is onboard the National Oceanic and Atmospheric Administration (NOAA) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). AVHRR provides global coverage in the Global Area Coverage format with a spatial resolution of approximately 0.083°, and higher-resolution data in the Local Area Coverage (LAC) format at about 0.01°."

Discussion section:
"The NDVI tends to saturate at high leaf area index (LAI) values, particularly in dense broad-leaved forests (Aklilu Tesfaye and Gessesse Awoke, 2021). Our compositing strategy, using the median value of the NDVI for 10 days, mitigates saturation and potential cloud contamination (Asam et al., 2023). An alternative is the kernel-based NDVI (kNDVI), which can address the saturation problem of the NDVI (Wang et al., 2023). However, kNDVI requires parameter tuning and kernel selection tailored to each use case, with these settings potentially varying across GPs. In addition, Wang et al. (2022) found no significant improvement between using the NDVI versus the kNDVI, likely due to the default kernel values not being universally applicable. The

use of NDVI ensures consistency across locations, facilitating comparability across Europe while mitigating biases through median compositing. While other studies have explored vegetation dynamics using total reflectances, Synthetic Aperture Radar technologies (SAR, e.g., Flores-Anderson et al., 2025; Thonfeld et al., 2022), or indices like the Normalized Difference Water Index (NDWI, e.g., Sturm et al., 2022), we selected NDVI due to the long temporal coverage and high revisit frequency of the AVHRR sensor, which is essential for a robust statistical analysis across Europe."

To improve clarity, we added the following sentence in section 2.1: "Selecting the maximum among the MED diminishes the impact of potential disturbances such as clouds, snow, and aerosols, which tend to lower NDVI values (Holben, 1986; Cihlar et al., 1994)."

Obuchowicz, C., Poussin, C., and Giuliani, G.: Change in observed long-term greening across Switzerland – evidence from a three decades NDVI time-series and its relationship with climate and land cover factors, Big Earth Data, 8, 1–32, https://doi.org/10.1080/20964471.2023.2268322, 2023.
Liu, G., Liu, H., and Yin, Y.: Global patterns of NDVI-indicated vegetation extremes and their sensitivity to climate extremes, Environmental Research Letters, 8, 025 009, https://doi.org/10.1088/1748-9326/8/2/025009, 2013
Sun, H., Wang, J., Xiong, J., Bian, J., Jin, H., Cheng, W., and Li, A.: Vegetation Change and Its Response to Climate Change in Yunnan Province, China, Advances in Meteorology, 2021, 8857 589, https://doi.org/10.1155/2021/8857589, 2021.
West, E., Morley, P., Jump, A., and Donoghue, D.: Satellite data track spatial and temporal declines in European beech forest canopy characteristics associated with intense drought events in the Rhön Biosphere Reserve, central Germany, Plant Biology Journal, 24, 1120–1131, https://doi.org/10.1111/plb.13391, 2022.
Hermann, M., Röthlisberger, M., Gessler, A., Rigling, A., Senf, C., Wohlgemuth, T., and Wernli, H.: Meteorological history of low-forest-greenness events in Europe in 2002–2022, Biogeosciences, 20, 1155–1180, ttps://doi.org/10.5194/bg-20-1155-2023, 2023.
Verbesselt, J., Hyndman, R., Newnham, G., and Culvenor, D.: Detecting trend and seasonal changes in satellite image time series, Remote Sensing of Environment, 114, 106–115, https://doi.org/10.1016/j.rse.2009.08.014, 2010.
West, E. et al.: Remote sensing for drought monitoring & impact assessment: Progress, past challenges and future opportunities, Remote Sensing of Environment, https://doi.org/10.1016/j.rse.2019.111291, 2019.
Brun, P., Psomas, A., Ginzler, C., Thuiller, W., Zappa, M., and Zimmermann, N. E.: Large-scale early-wilting response of Central European forests to the 2018 extreme drought, Global Change Biology, 26, 7021–7035, https://doi.org/10.1111/gcb.15360, 2020.
Buras, A., Rammig, A., and Zang, C. S.: Quantifying impacts of the 2018 drought on European ecosystems in comparison to 2003,550 Biogeosciences, 17, 1655–1672, https://doi.org/10.5194/bg-17-1655-2020, 2020
Asam, S.; Eisfelder, C.; Hirner, A.; Reiners, P.; Holzwarth, S.; Bachmann, M. AVHRR NDVI Compositing Method Comparison and Generation of Multi-Decadal Time Series—A TIMELINE Thematic Processor. *Remote Sens.* **2023**, *15*, 1631. https://doi.org/10.3390/rs15061631

The last topic that should be also included in the introduction is the use of Random Forest and Lasso Methods and advantages and disadvantages of each method, referring studies that used those methods for similar studies over Europe and/or other regions.

Thank you, we added the following sentences at the end of the second last paragraph of the introduction:
“For instance, Lasso regression can identify key drivers through feature selection and performs well in the presence of correlated variables (Tibshirani, 1996; Vogel et al., 2021). Random Forest, a more flexible model based on decision trees, is an interpretable method that ranks predictor importance and captures non-linear relationships (Breiman, 2001; Oliveira et al., 2012).”

NDVI data:
Lines 67-69: provide references.
Line 78-80: The reference provided is about the correction of AVHRR database for Fennoscandia. Provide results of correction for the other biomes in Europe, namely the Mediterranean region that presents very different vegetation types and biomes. Alternatively compare with GIMMS dataset (at 8 km spatial resolution) after resampling to the same resolution. This exercise make sense as the authors did not take advantage of the spatial resolution of 1 km.
Thank you for your remark. Barben et al. (2024) indeed focused on Fennoscandia by utilizing a pan-European AVHRR LAC dataset, extracting only the relevant regional data. The correction is not specifically tailored for Fennoscandia but rather designed for broader European applicability. Following your remark, we added more references and further described the spectral response function method in the following paragraph in the Appendix A:

“To account for differences in the spectral response function (SRF) among the AVHRR sensors (AVHRR/1, AVHRR/2, and AVHRR/3), the data were normalized to NOAA-9 AVHRR/2 following the method described by Fontana et al. (2008) and Barben et al. (2024), using the polynomial coefficients proposed by Trishchenko et al. (2002). This SRF correction was applied to the entire LAC dataset, covering all of Europe. As demonstrated in both Fontana et al. (2008) and Trishchenko et al. (2002), the correction performs well across various biomes. ”

Fabio M.A. Fontana, Alexander P. Trishchenko, Konstantin V. Khlopenkov, Yi Luo, Stefan Wunderle,Impact of orthorectification and spatial sampling on maximum NDVI composite data in mountain regions,Remote Sensing of Environment,
https://doi.org/10.1016/j.rse.2009.08.008
Trishchenko, A. P., Cihlar, J., & Li, Z. (2002). Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors. Remote Sensing of Environment, 81(1), 1–18. https://doi.org/10.1016/S0034-4257(01)00328-5

We also modified the section 2.1 NDVI data accordingly:
“Prior to that, the dataset was orthorectified, radiometrically calibrated, and filtered for clouds, and the NDVI values were spectrally corrected for the different versions of the AVHRR instrument (see section A in the appendix for further details).”

The GIMMS dataset is created from AVHRR Global Area Coverage (GAC) data using maximum NDVI values reported twice monthly (Pinzon and Tucker, 2014). Our NDVI dataset is obtained using 10-days Median value composites from AVHRR Local Area Coverage (LAC) and using a NWPSAF-PPS cloud mask. Furthermore AVHRR GAC are obtained from a unique sampling method. Regarding these differences between both datasets, while we carefully considered your suggestion to compare with the GIMMS dataset, we ultimately decided against it, as the justification provided above supports the broader applicability of our correction method, and such a comparison would fall outside the scope of this study.

Pinzon, J. E., & Tucker, C. J. (2014). *A non-stationary 1981–2012 AVHRR NDVI3g time series*. Remote Sensing, 6(8), 6929–6960. https://doi.org/10.3390/rs6086929

Lines 85-86: explain why the maximum NDVI value is better than the median. provide references.

The MED (median value) compositing strategy has been applied to generate 10 days composites as suggested and shown in Asam et al. (2023), cited in the manuscript. As several AVHRR sensors collect data concurrently, the maximum value (from each 'MED' time series) has been chosen to diminish the impact of potential disturbances such as clouds, snow, and aerosol which tend to lower NDVI values (Holben et al., 1986; Cihlar et al., 1994).

The compositing strategy has been done with the MED and not instead of the MED. Following your remark, we clarified this point in the manuscript:

“As several platforms (NOAA-6 to -18 and the MetOp-series) may be available for a given 10-day composite, we selected the maximum NDVI value among the MED (10-day median value composite) NDVI values to obtain a single time series. Selecting the maximum among the MED diminishes the impact of potential disturbances such as clouds, snow, and aerosols, which tend to lower NDVI values (Holben, 1986; Cihlar et al., 1994).”

HOLBEN, B. N. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing*, *7*(11), 1417–1434. https://doi.org/10.1080/01431168608948945,

J. Cihlar, D. Manak and M. D'Iorio, "Evaluation of compositing algorithms for AVHRR data over land," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 2, pp. 427-437, March 1994, doi: 10.1109/36.295057.

From NDVI to binary forest damage

Lines 94-107: The ‘binarization’ of NDVI is a tricky issue. Several assumptions must be justified in the context of entire Europe:

-Why considering only summer? spring is the most active season in terms of vegetative cycle for several species.

- Why only July and August and not June July and August or July, August and September?

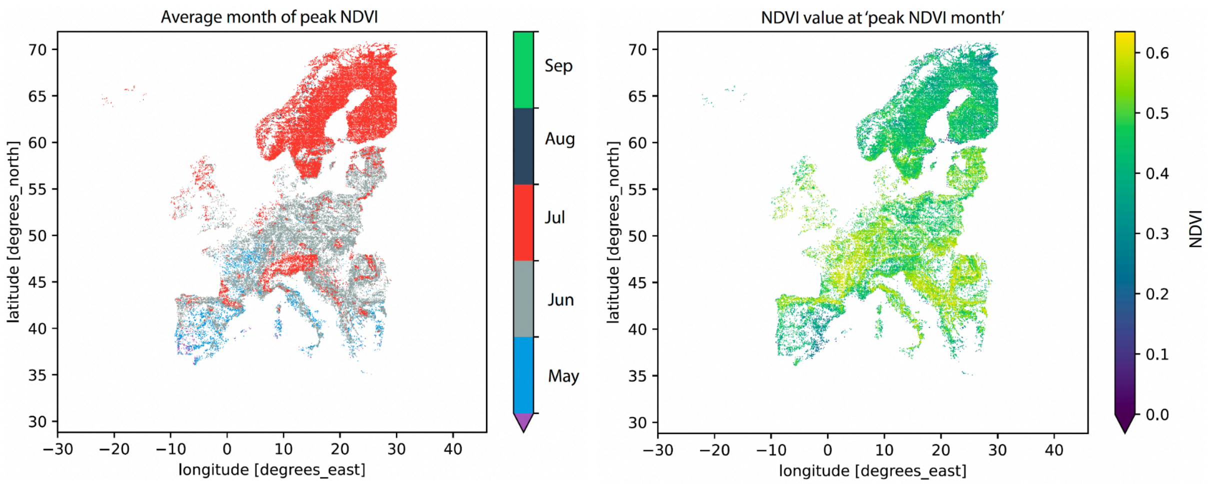
To justify the above mentioned points provide the annual vegetative cycle for the

different types of forest. Provide maps of the months of NDVI peak for the forest GP and maps of NDVI mean value for the Peak Month for each GP across the europe.
Thank you for these two relevant points regarding the months employed to compute low greenness events. Following your recommendation, we have added a new figure to the appendix (now Fig. B1), which shows the average month of yearly NDVI maximum for each 0.1° forest grid point across Europe, and the value of NDVI during this peak NDVI month. This visualization helps illustrate the spatial variability in forest phenology.

We chose to focus on July and August to ensure that all forested areas across Europe, including those at higher latitudes and altitudes, had reached their peak greenness. This approach minimizes the risk of misclassifying low NDVI values due to delayed phenological development rather than actual summer browning.

Using a consistent time window (July-August) across all grid points also allows for harmonized analysis of potential drivers of low greenness events (e.g., in Figure 4), facilitating comparability across regions. While this choice emphasizes late-summer damage, the method can be extended to other months or seasons in future work.

This rationale is now clarified in Section 2.1 of the manuscript: “We selected July and August to ensure that all forested regions across Europe had reached peak greenness, allowing consistent comparison of low NDVI events and their drivers across locations (see Fig. B1 in the appendix). To ensure consistency and comparability across Europe, we selected the same time window (July-August) for all grid points in the analysis.”



New Figure B1. Spatial patterns of the average month of the NDVI peaks across European forest grid points (left panel) and mean value of the NDVI peak (right panel). The average month of yearly NDVI maximum is computed for each 0.1◦ forest grid point. Colors on the left panel indicate the month of peak NDVI: purple for April, blue for May, grey for June, red for July, navy blue for August, and green for September.

- Why the threshold of 80%?

We adopted the 80% threshold following the methodology of Hermann et al. (2023), where it represents a balance between identifying sufficiently extreme events and retaining enough data points for robust analysis. This threshold ensures that the classified events are spatially extensive while avoiding overly restrictive criteria that would result in too few events. This is now better explained in the method section (together with the modification following your next point on NDVI anomaly<0): “The 80% spatial threshold and the requirement of negative NDVI anomalies for at least 5 out of 6 summer datapoints ensure that browning events are both widespread and temporally persistent, reducing noise from short-term or localized fluctuations (Hermann et al., 2023)”.

- Why NDVI Anomaly <0? NDVI anomaly equal to zero is the common state. To assume a decrease that could be related with a real loss (or ‘damage’) the use of a standard deviation value would be more realistic.

Thank you for your remark. The use of NDVI anomaly < 0 is also based on Hermann et al. (2023). A year is classified as an extreme event only if the anomaly is negative for at least 5 out of the 6 summer datapoints. This criterion ensures temporal persistence of the anomaly, distinguishing genuine vegetation stress or damage from short-term fluctuations. Additionally, the 80% spatial threshold further filters out localized or random variations.

Although NDVI anomaly = 0 represents the average state, using a standard deviation-based threshold would introduce complexity in defining consistent thresholds across diverse forest types and regions. Our approach, combining spatial extent and temporal persistence, results in a binary classification that captures relatively rare but significant browning events.

The choices regarding the thresholds are now better explained in the method section of the manuscript: “The 80% spatial threshold and the requirement of negative NDVI anomalies for at least 5 out of 6 summer datapoints ensure that browning events are both widespread and temporally persistent, reducing noise from short-term or localized fluctuations (Hermann et al., 2023)”, and supported by Fig. B4, which shows that most grid points have fewer than 15% of years classified as browning events.

Lines 108-110: Another very important point is the characterization of a GP as forest. Taking in account the original spatial resolution of CORINE (100m) and the *resolution used for this study (around 10km) the resampling assume a crucial role and should be explained.*

Thank you for your remark. The classification of a GP as a forest is performed at the native spatial resolution of the AVHRR NDVI dataset (0.01°), in order to exclude the NDVI of gridpoints that are not forests (l.110). The classification of a gridpoint as forest is therefore not done at the final resolution for the study (0.1).

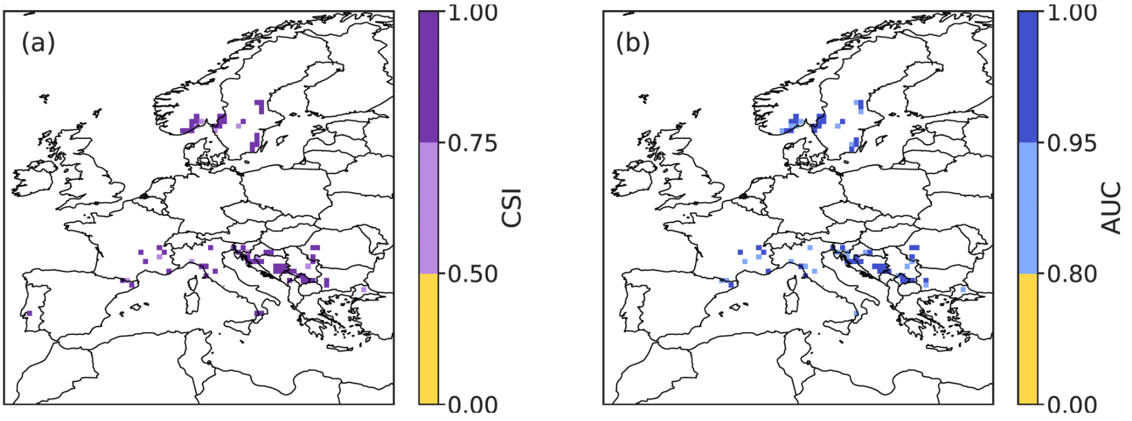
We interpolated the CORINE land cover classification to 0.01°, considering the 'most frequent' occurrences of classes in 0.01° pixels. This is now better explained in the manuscript, section 2.2: “1. Retain only the 0.01° pixels classified as forests in the CORINE Land Cover dataset (EEA, 2020a), considering the most frequent occurrences of classes in 0.01°pixel”

Besides this issue, the GP seems to be very coarse and therefore the threshold of 10% of forest seems very low. This will allow to have GPs that represent better cities or other types of vegetation like shrub land, than forests. I strongly suggest to use a threshold of 80% or reduce the size of the GP. A map with the % of forest for the selected GPs should be presented and discussed in discussion section.

Thank you for your comment. We would like to highlight that, within a 0.1° gridpoint, before the aggregation step, all the 0.01° pixels that are not classified as forest are discarded (l.110). To further clarify this point, we added a step in the revised method section, section 2.2:

“1. Retain only the 0.01° pixels classified as forests in the CORINE Land Cover dataset (EEA, 2020a), considering the most frequent occurrences of classes in 0.01°pixel”.

Following your suggestion, we re-ran the model for gridpoints with a forest cover of at least 50% per 0.1° gridbox. This adjustment left only 87 gridpoints on which the RF procedure could be run (see figure below). We show below the performance metrics for these gridpoints, to illustrate the sparsity of the obtained model output.

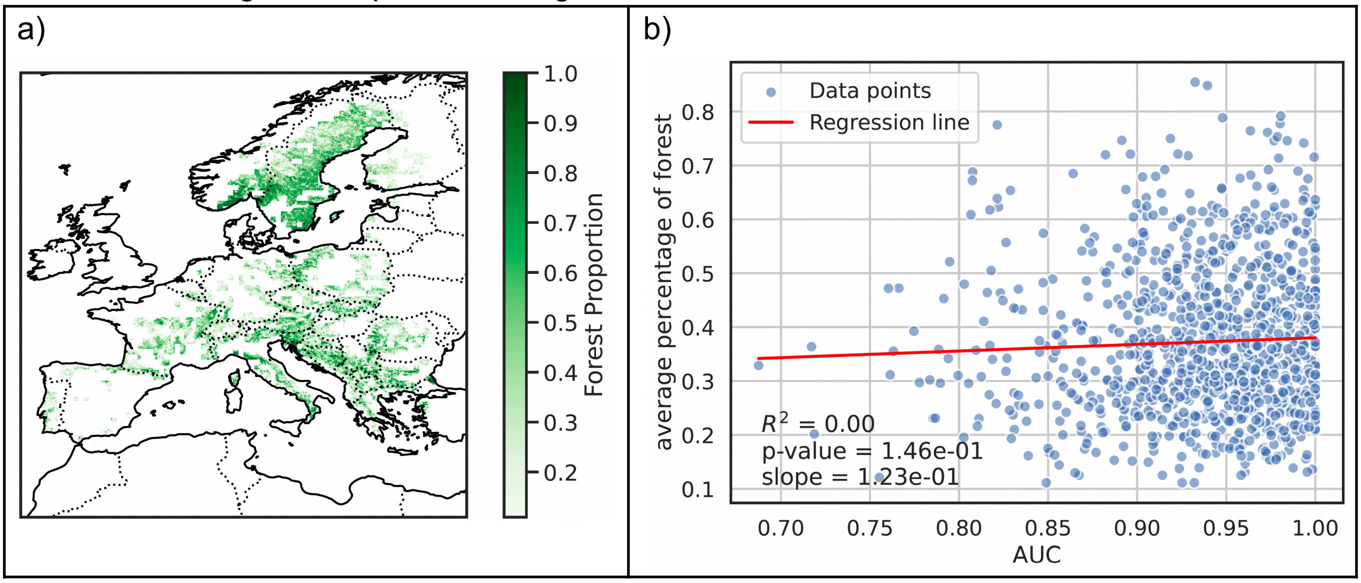


We intentionally set the forest cover threshold at a relatively low level (10%) to ensure sufficient spatial coverage at the 0.1° resolution. It’s important to note that within each 0.1° grid cell, only the 0.01° pixels classified as forest (based on CORINE land cover) are used in the analysis. For instance, if a grid cell contains

11% forest pixels, the binary browning signal is computed exclusively from the NDVI values of those forest pixels. Thus, the 0.1° browning data reflects the condition of forested areas only, and non-forest pixels are entirely excluded from the computation.

Following your remark, we quantified the impact of the forest threshold on the performance. The following plots represent a) the spatial distribution of the forest proportion per 0.1° gridpoint, and b) the percentage of forest within the 0.1° gridpoints, as a function of the CSI. We do not observe a significant relationship between CSI and forest cover percentage, as indicated by a near-zero slope and an explained variance of $R^2=0$. As the forest percentage does not impact the performance of the model, we kept a forest cover threshold of 10% to ensure sufficient spatial coverage at the 0.1° resolution.

This is now shown in the manuscript: we added the forest percentage map in the manuscript (fig. B2, introduced in the data section “Figure B2 in the appendix illustrates the proportion of forested area within each 0.1°× 0.1°GP, showing only GP with more than 10% forest coverage.”) and the scatterplot of forest percentage vs CSI (that is now called figure E1, independent of the former figure E1).
We added the following sentence in the result section:
“We do not observe a dependence between model performance and the percentage of forest cover within the GP area (fig E1). This supports the use of a 10% threshold for forest pixel inclusion, which ensures sufficient representation of forested pixels while maintaining broad spatial coverage.”



Lines 110-112: Why discarding the pixels that experienced a forest change between 2006, 2012 and 2018? The change may also be related with forest browning and mortality.

We agree that with this procedure, we are removing cases with high mortality. However, when running the random forest to identify drivers, we assume that the link between hydro-meteorological anomalies and forest browning does not change through time. If there is a change of forest type, there might be a change of drivers for browning. Our study aims to stay general at the scale of Europe. Such grid points with discontinuities can be analyzed with case studies, which is beyond the scope of our manuscript. We have now clarified this point in the method section:
“Removing pixels that experienced forest change is necessary, as the random forest model we employ (see Section 3.1) assumes that the relationship between hydro-meteorological anomalies and forest browning remains stable over time. A change in forest type could alter the sensitivity to hydro-meteorological conditions and introduce different response mechanisms, thereby confounding the identification of consistent drivers”

Lines 115-119: Instead of concatenating, use the NDVI correction method proposed by Stockli and Vidale, 2004.
Stockli R, Vidale PL. 2004. European plant phenology and climate as seen in a 20-year AVHRR land-surface parameter dataset. International Journal of Remote Sensing 25: 3303–3330.
Thank you for the suggestion.
The methodology proposed by Stöckli and Vidale applies a second-order Fourier adjustment algorithm to smooth NDVI curves and reduce noise, enabling the efficient detection of key phenological events such as the start of vegetation growth, snow melt-out, or leaf-out. This approach is commonly used for deriving long-term phenologies.
In our study, however, the focus lies on detecting anomalous NDVI events. Moreover, the concatenation (l.116-119) does not aim for smoothing the NDVI but rather obtaining time series long enough for applying a random forest model. Unfortunately, the method developed by Stöckli and Vidale would provide a solution to having time series long enough to run a statistical model.

Hydro-meteorological predictors
Lines 122-124: The selected predictors show potential high multicolinearity, as they are correlated parameters. For instance, the soil moisture variables are strongly correlated and the surface soil moisture also correlated with precipitation. Provide an assessment of multicolinearity, selected the less affected variables or use statistical methods to reduce the multicolinearity.
Following your remark and remark from reviewer#2 on multi-collinearity, we removed 3 variables: dewpoint temperature, soil moisture (0-7cm) and soil moisture (7-28cm). We still kept soil moisture (28-100cm) in addition to precipitation, as they still bring complementary information (e.g. if in case of precipitation extremes, high run off, implies soil moisture not as extreme as precipitation). We justified our decision in the

manuscript in section 2.3 and added a figure showing the correlations between variables, in the appendix:

"To capture adverse hydro-meteorological conditions for forests, we select four hydro-meteorological variables as potential drivers for low-greenness events from an initial set of seven (see figure C1 in Appendix). The selection was based on two criteria: (i) limiting pairwise correlations between variables to the range of -0.4 to 0.4, a conservative threshold compared to commonly used values (e.g., Dormann et al., 2013), and (ii) prioritizing variables known to influence NDVI based on previous studies and expert knowledge (Hermann et al., 2023; Grossiord et al., 2020; Young et al., 2017; Alavi, 2002). The final set of variables includes: maximum 2-m temperature (Max T2m), soil moisture (soil moist.) at depth 28-100 cm, total precipitation (total precip.), and surface latent heat flux (s.l.heat flux)."

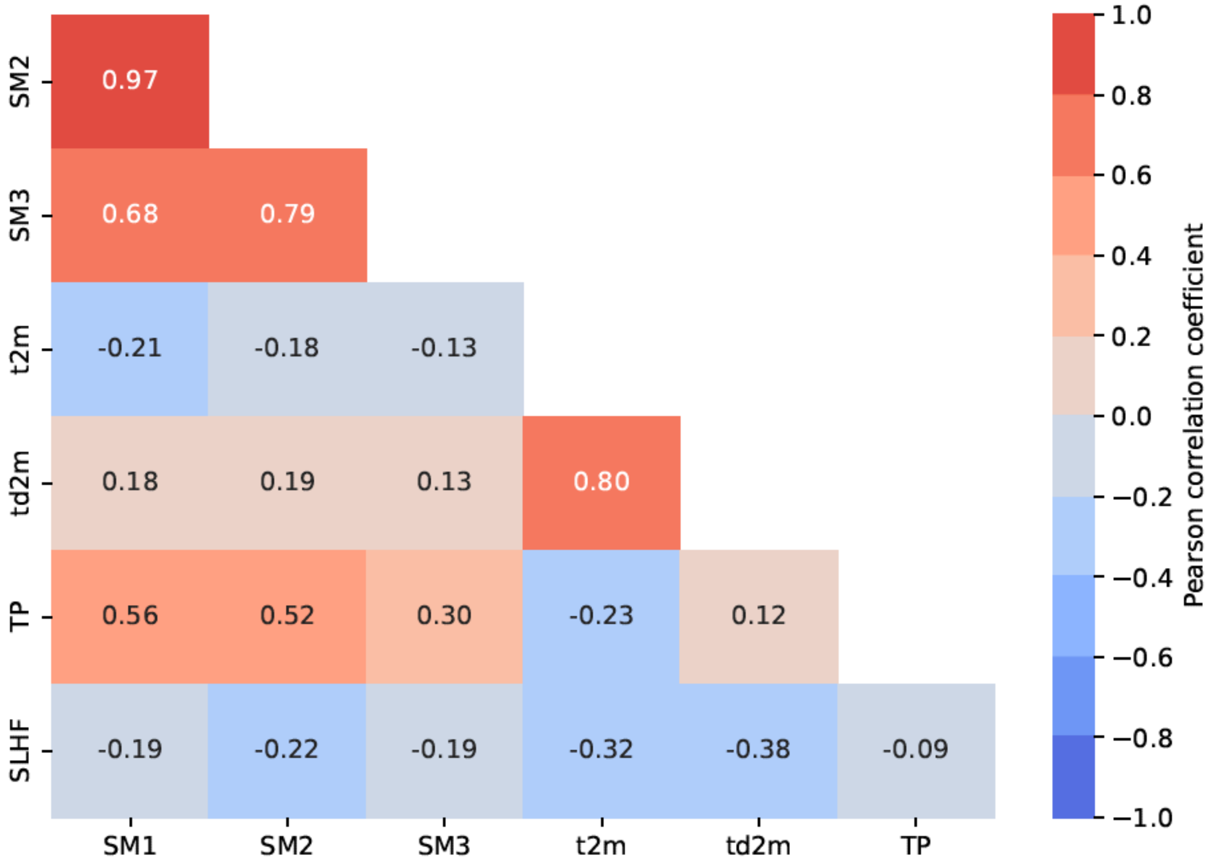


Figure C1. Pairwise correlation coefficients between hydro-meteorological predictor variables, averaged across months and over Europe. Variables include soil moisture at three depths (SM1: 0–7cm, SM2: 7–28cm, SM3: 28–100cm), 2-meter air temperature (t2m), 2-meter dew point temperature (td2m), total precipitation (TP), and surface latent heat flux (SLHF). We excluded td2m, SM1, and SM2 from the analysis to ensure that all remaining predictors have cross-correlations between –0.4 and 0.4, thereby reducing multicollinearity and enhancing model interpretability.

Lines 131-132: ‘monthly (seasonal) standardised anomaly’ instead of monthly (seasonal) anomaly’
Figure 10: correct the figure caption changing to: 7 variables and 10 time-steps
Thank you, we changed the text accordingly.

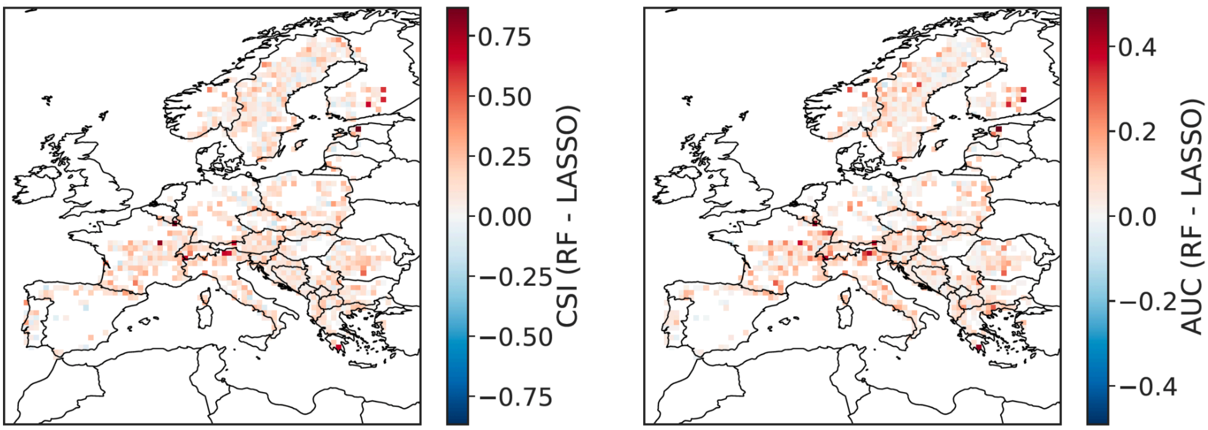
Method
RF Method is a robust approach, but still potentially affected by multicolinearity. Therefore, despite the use of only 3 predictors for each model, the results are potentially affected by overfitting. Consider to evaluate the possible overfitting using a cross validation approach.
Thank you for your suggestion. To address potential multicollinearity, we removed highly correlated variables, namely the two top layers of soil moisture and dewpoint temperature (see our reply to the previous comment). After re-running the model with the reduced set of predictors, the results remained consistent, with the same key variables emerging as most important across Europe, now with even stronger importance scores.
While Random Forest can be influenced by multicollinearity, its ensemble nature and variance reduction through averaging decision trees help mitigate this issue. Additionally, we implemented a train-test split (70% training, 30% testing) to evaluate model performance. The model shows consistently high predictive accuracy across Europe, suggesting that overfitting is not a major concern.

The LASSO method seems to produce less good results (but maybe more robust, taking in account the overfitting that likely affect RF), was designed to deal with correlated variables and therefore seems to be more appropriate, if the authors decide to maintain the same predictor variables. In this case, a detailed description of the method and references should be provided here. If not, LASSO section as does not correspond to a significant result in the paper.
Thank you for your remark. We reduced the number of variables to prevent multi-collinearity, following your remark. As the RF model is able to catch some non-linear dependence between forest browning and meteorological variables, we would like to still keep RF as the main model. While the LASSO regression is indeed not a major result in our paper, we would still like to keep it, as a benchmark for performance. Therefore, we added a figure comparing the performance of RF and LASSO for all the gridpoints (figure E2).

Results
Lines 202_204: Present also the %s for CSI>0.75 and AUC> 0.95 and for CSI<0.5 and AUC<0.80. Justify the low skill for these GPs.
Thank you, we added AUC> 0.95, this is now clarified in the revised manuscript:
“The AUC also indicates a good performance, with 98% of the GPs having an AUC greater than 0.8, and 48% of the GPs with an AUC larger than 0.95.”

CSI>0.75 was already present in the text. We decided to not specify the percentage of CSI<=0.5 (1%) as CSI>0.5 (99%) is already present in the manuscript. The same is true for AUC<0.80.
We chose not to explicitly justify these low skill values in the main text, as they represent only a very small fraction of the total. Given that we independently trained over 1,059 models, one per grid point, some variability in model performance is expected. A small number of low-skill cases is therefore not unexpected and does not compromise the robustness of the overall results.
Following your remark above, we also investigated whether model performance was related to forest cover within each grid point. As shown in Figure E1, there is no significant relationship between CSI and the percentage of forest cover ($R^2 \approx 0$), and low-skill GPs are not concentrated in any particular forest type or region. This suggests that the few low-performing models are not systematically associated with specific land cover characteristics, but may instead reflect local data limitations, weaker signal-to-noise ratios, or more complex disturbance dynamics that are harder to capture with the available predictors. A more detailed investigation of these local limitations is left for future work; here, we focus on analyzing the broader-scale patterns and drivers of forest browning across Europe.

Rewrite the Figure 2 Caption as: Spatial distribution of performance metrics for RF model: (a) CSI (critical success index) and (b) AUC (area under the ROC curve, both defined in section 3.2) of the random forest model, evaluated for the training dataset.
Lines 211-215: Show the results for LASSO or remove this paragraph
Thank you, we changed the caption of figure 2.
We added a new figure in the appendix (figure E2), showing the performance difference between the random forest model and the LASSO regression. We refer to this figure in section 4.1: "In comparison with RF, the LASSO model exhibits a lower CSI for 83% of the GPs, and a lower AUC for 89% of the
GPs (see Fig E2 in appendix)"



New Figure E2: Spatial distribution of difference in performance metrics between the RF and the LASSO model: (a) CSI (critical success index) and (b) AUC (area under the ROC curve, both defined in section 3.2), evaluated with the testing dataset.

Example for two grid points
Figure 3 is very difficult to read (very small font size) and to interpret. Maybe I missed, but the CSI and AUC for this 2 places is not mentioned in the manuscript. Figure should be change to have more readability and be more easy to follow. The percentage of forest for this GP should also be discussed.
Thank you, we improved the figure and the text:
-We added the CSI and percentage of forest for these GP, a justification behind the selection of the 2 GP:
“These GPs were selected to showcase the diversity in forest type and the variability in driver responses captured by the RF model. The first grid point, predominantly broadleaf forest, is located in northeastern France and has 41% forest cover. The RF model achieves a CSI of 0.92 and an AUC of 0.99 at this location. The second grid point, characterized by coniferous forest, is situated in southwestern Sweden, with 66% forest cover, a CSI of 0.97, and an AUC of 1.”
-We guided the description of Figure 3:
“Positive 2-m temperature anomalies in the summer of the preceding year and in July of the current year (left side of the x-axis), as well as negative precipitation anomalies in MAM of the previous year (right side of the x-axis), are all associated with an increased probability of forest browning (Fig. 3, grid point (A), second, third, and last panels).”

More locations among different region of the europe, such as eastern, southeastern, southwestern europe, must be analysed,
Thank you for your valuable comment. We acknowledge the importance of including a broader range of locations across Europe. In our study, we selected two representative sites primarily to help guide the reader through the interpretation of the mean decrease in accuracy and the partial dependence plots. These examples were not intended to be exhaustive but rather illustrative.
Following your remark, to address spatial representativeness more explicitly, we have now integrated Figure 6 in the main text rather than in the appendix, which includes maps and highlights key drivers across Europe. This addition provides a more comprehensive spatial overview of the model outputs.

Results over europe
In my opinion the most interesting part of the paper, but very incomplete.
Figure E1 should be moved to the maintext.
Thanks a lot for your suggestion. We moved the former figure E1 to the main text (Figure 6 now), added some panels and described it more extensively:
“We now examine the spatial distribution of the most influential predictors across Europe, highlighting regional patterns in predictor importance (Fig. 6). Temperature was selected as one of the top 10 most important predictors for the vast majority of

grid points across Europe, with a few exceptions in Scandinavia and Bosnia and Herzegovina (top left panel). Soil moisture (28–100 cm) was not retained as an important variable in Switzerland, Eastern France, and the Northern Balkans, but it was selected in regions such as southern Iberia, the southern Balkans, and parts of Norway (top row, second panel). Interestingly, precipitation was also identified as an important predictor in Scandinavia, alongside soil moisture, and in some grid points in Central Europe where soil moisture was not retained. In contrast, surface latent heat flux did not emerge as a dominant predictor in any specific region; its importance appears relatively evenly distributed across Europe (Fig. 6, left and middle panels). Temporal variables from the previous year were among the top 10 predictors at most grid points, with slightly higher
importance observed at higher latitudes, particularly in Scandinavia (Fig. 6, bottom row, left panel). Similarly, conditions during June–July of the current year were frequently selected, with especially high importance in central France and eastern Germany, where between 6 and 8 of the top 10 predictors correspond to this period (bottom row, right panel). DJF of the current year played a moderate role, being selected in scattered grid points across Europe (bottom row, second panel). Finally, spring conditions (MAM) of the current year were highly influential, with 4 to 6 predictors among the top 10 in several regions, including southern Italy, Portugal, central France, the Balkans, and Scandinavia (bottom row, third panel)."

We also restructured the discussion section to increase the attention given to the comparison with existing literature. The new subsections are: a comparison of findings with existing literature (5.1 Identified adverse conditions), methodological considerations (5.2 Preprocessing and predictor selection) and dataset limitations (5.2).

Maps of Top 3 predictores used for each GP considering the type of forest will help to understand the role played by each variable across the europe.
Thank you for your remark. As there are 40 potential predictors, a map of the top1 variable would need 40 different colors, which would not be readable. The new figure 6 (improved version of former figure E1) aims at summarizing this information, indicating which type of meteorological variable or which time periods were important for each gridpoint.

A discussion of the role played by each predictor in function of the % of forest in the GP should be added to the analysis and discuss accordingly.
Thank you for this suggestion. Upon visual comparison of Figures 6 and B2, we do not observe consistent patterns between them. This is expected, as the low NDVI events under investigation are restricted to GPs predominantly covered by forest and are therefore not influenced by other land cover types. We would like to highlight here too that, within a 0.1° gridpoint, before the aggregation step, all the 0.01° pixels that are not classified as forest are discarded. To further clarify this point, we added a

step in the revised method section, section 2.2: “1. Retain only the 0.01° pixels classified as forests in the CORINE Land Cover dataset (EEA, 2020a), considering the most frequent occurrences of classes in 0.01°pixel”.

Figure 5 is also very difficult to read (very small font size) and to interpret. Use the same y-axis range for all the figures and change colors, as it is not possible to discriminate the different latitudes

Thank you for your remark, we adapted the figure accordingly

Identified adverse conditions

The results do not show a high level of novelty and the discussion in mainly focused on previous works from other authors. The authors must highlight the novelty of the results obtain.

Following your suggestion, we highlighted further the novelty and added value of our method in the abstract, introduction, and conclusion.

The role played by each hydro-meteorological predictors for forest browning across the europe is not presented and discussed in a systematic way, allowing the reader to identify the main drivers for the different regions.

Thank you for this valuable comment. In response, we have added a new Figure 6, which provides a spatially explicit summary of the most important variables and time periods. While we cannot represent the importance of each of the 40 drivers for each grid point, this figure summarizes the importance of variables and time periods for each grid point across Europe.

Conclusions

Line 391: The authors only studied these seasons, so the word ‘primarly’ is not a result from this study.

Thank you for your remark. We added “to predict summer browning” to avoid any misunderstanding in this sentence (we meant primarily, among all the other time periods considered as potential drivers).

Line 397: The crucial conditions and periods are identified, but not presented clearly in the manuscript.

Thank you. We added fig6 and we adapted the conclusion accordingly.

The authors should show the novelty of this work and the added value of the presented

Thank you for your recommendation. We highlighted the novelty and added value of our work in

- The abstract:

“In this study, we present a novel, large-scale, spatially explicit analysis of forest browning drivers across Europe, using a homogeneous and automated random forest modeling framework. By running independent models at each 0.5° grid point, we enable a region-specific comparison of hydro-meteorological drivers, capturing the diversity of forest responses across the continent.”

- The introduction:

“In this work, we introduce a novel, automated framework to identify and quantify adverse sub-seasonal to seasonal (S2S) hydro-meteorological conditions driving forest browning across Europe.”

“We run independent RF models at each 0.5° grid point, enabling a region-specific identification of key predictors and time windows. This allows us to capture the diversity of forest-climate interactions across Europe, rather than relying on a one-size-fits-all model. Our goal is to pinpoint the most critical hydro-meteorological drivers, at monthly to annual timescales, using only variables that are available from operational S2S forecast systems. By linking forest browning to forecastable climate conditions, our method provides actionable insights for forest managers, offering valuable lead time to anticipate and mitigate the impacts of climate extremes through targeted interventions.”

- The conclusion:

“In this study, we presented a large-scale, spatially explicit identification of forest browning drivers across Europe, using a homogeneous and fully automated modeling framework. Our approach achieves high predictive performance while running independent models for each 0.5° grid point, enabling a regionally nuanced comparison of hydro-meteorological drivers. ”

“The identification of critical conditions and time periods at the local scale, combined with the use of hydro-meteorological variables available from sub-seasonal to seasonal forecast products, offers practical opportunities for proactive forest management. Regional forest agencies can leverage our methods and findings to anticipate periods of increased risk and implement targeted preventive measures, such as monitoring key variables in preceding seasons and years (e.g., autumn soil moisture or spring temperature). This approach enables more strategic allocation of resources, including irrigation, thinning, or pest control, tailored to the most influential local drivers. Such preventive measures would mitigate the economic and environmental costs of forest damage.”

Reply to Reviewer #2.

Legend: *-Reviewer's comments* -authors reply

General Comments:
First, I would like to commend the authors for their detailed response to all the reviewers' comments. I recognize that the clarity and flow of the manuscript have improved somewhat in this revised version. However, as outlined in my observations, several key concerns remain unaddressed. Both the writing and the figures still require substantial improvement. For these reasons, I maintain my initial recommendation against publication of the manuscript in its current form in the NHESS journal. I apologize for any inconvenience that my comments may have caused, and I wish the authors the best of luck with their next submission.

We would like to thank you again for your time, very valuable comments, and kind remarks, to help us improve the manuscript.

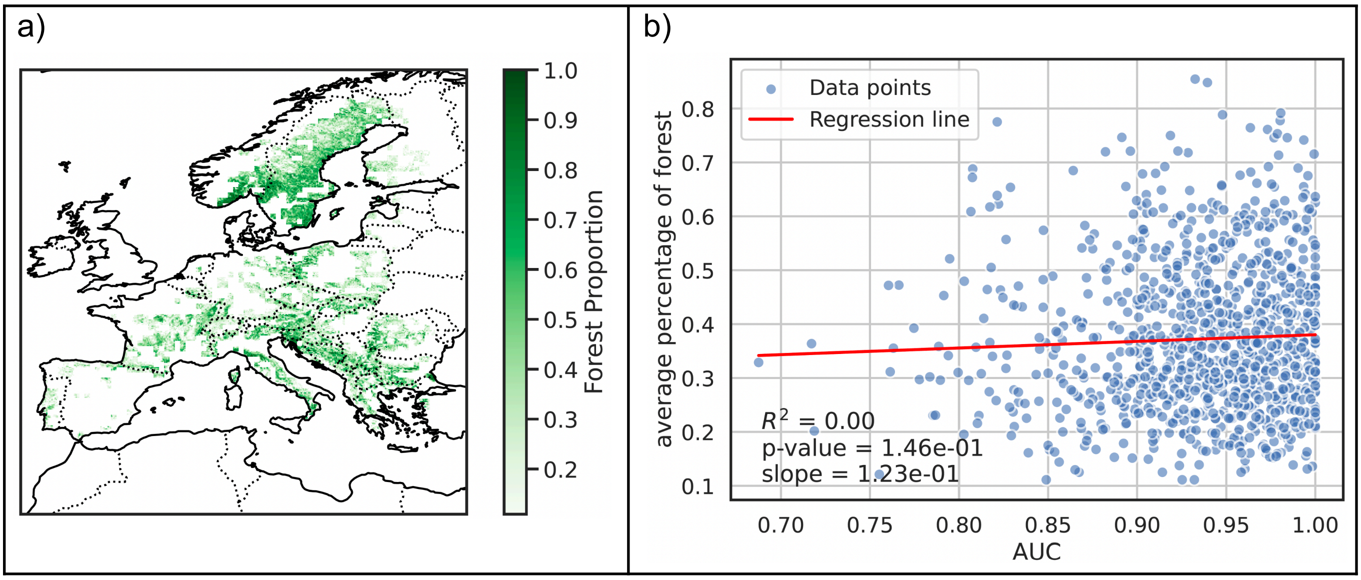
Regarding the term "forest damage": Once again, I advise against using the term "forest damage", as the authors cannot demonstrate—based solely on NDVI data—that the observed low greenness events result in measurable negative biological impacts at the population or community level in broadleaf or coniferous forests (see my previous comments in the first round review). Therefore, I strongly recommend that the authors consistently use terms like "forest browning" or "low greenness events", as they already do in the paper's title.
Thank you, we removed all occurrences of "forest damage" and replaced them with "low greenness events" and "forest browning".

Regarding the land use concerns: I appreciate the authors' response to my concerns regarding potential biases arising from land use change. While I recognize their efforts to address this issue, I remain concerned about the lack of clarity in describing the adopted procedures and the limitations of the underlying data. From their explanation, I understand that only forest pixels (at a 0.01° resolution) were used to calculate the average NDVI values, which were then aggregated to produce the 0.1° GP values, with non-forest pixels excluded from the calculation. I would appreciate confirmation of whether this interpretation is correct. If so, this approach would indeed help mitigate biases introduced by annual crop variation. However, it does not fully account for silvicultural forestry areas, such as those subjected to pruning, which can lead to abrupt changes in NDVI while remaining consistently classified as forest over time. Portugal, for instance, contains extensive regions of such managed forests. In this context, my earlier suggestion to include control *areas—such as unmanaged protected areas—could help reduce these uncertainties. While I acknowledge the author's rationale, I believe it is important that the manuscript includes a more detailed description of these methodological steps, especially regarding the upscaling process from 0.01° to 0.1°, and the justification for including GPs where forest pixels represent as little as 10% of the area—a threshold that seems quite low, in my view.*

Thank you for your follow-up. Your interpretation was indeed correct, only the 0.01° pixels classified as forest (based on the CORINE Land Cover dataset) were retained for analysis. NDVI values were computed exclusively from these forest pixels and then aggregated to the 0.1° grid point (GP) level. We improved the clarity of this step in the revised manuscript by explicitly adding the following initialization step in the methods section: “1. Retain only the 0.01◦ pixels classified as forests in the CORINE Land Cover dataset (EEA, 2020a), considering the most frequent occurrences of classes in 0.01° pixels”

Regarding your concern about silvicultural practices (e.g., pruning) within consistently classified forest areas, we agree that such management activities can introduce abrupt NDVI changes that are not necessarily linked to climate-driven browning. While our current approach helps mitigate biases from land use change (e.g., conversion to agriculture), it does not explicitly distinguish between managed and unmanaged forests. We acknowledge this as a limitation and have added a note in the discussion section suggesting that future work could explore differences in model performance and browning signals between managed and unmanaged forest areas: “Future local studies work could also explore differences in browning dynamics and model performance between managed and unmanaged forest areas, such as protected zones, to better account for the influence of silvicultural practices not captured by land cover classifications alone.”

Regarding the impact of the 10% forest threshold for exclusion of gridpoints, the following figures represent a) the spatial distribution of the forest percentage per 0.1° gridpoint, and b) the percentage of forest within the 0.1° gridpoints, as a function of the CSI. We do not observe a significant relationship between CSI and forest cover percentage, as indicated by a near-zero slope and an explained variance of R^2=0, even for gridpoints with a forest coverage near 0. As the forest percentage does not impact the performance of the model, we intentionally set the forest cover threshold at a relatively low level to ensure sufficient spatial coverage at the 0.1° resolution. Following your remark on forest percentage, we added the forest percentage map in the manuscript (fig. B2), the scatterplot of forest percentage vs CSI (that is now called figure E1, independent of the former figure E1, moved to the main text as figure 6 now). We added the following sentence in the result section: “We do not observe a dependence between model performance and the percentage of forest cover within the GP area (fig E1). This supports the use of a 10% threshold for forest pixel inclusion, which ensures sufficient representation of forested pixels while maintaining broad spatial coverage.”



Regarding the use of hydro-meteorological predictor variables: I respectfully disagree with the author's justification for including predictor variables that are likely to be autocorrelated (note that this is also a concern raised by Reviewer 1). For instance, the authors state that they selected the dew point temperature because it provides "information about water content in the air". However, there are other variables—such as specific humidity or absolute humidity—that also quantify atmospheric moisture and are not directly correlated with air temperature. Combining temperature with a measure of air humidity can reduce the risk of multicollinearity while still providing an effective proxy for vapor pressure deficit (VPD). Similarly, the inclusion of three different soil moisture metrics raises questions. Is there a specific ecological rationale for this choice, particularly given that NDVI measurements in forested areas are likely to reflect the condition of mature trees with deep root systems? In this context, what added value do the authors expect to gain from surface soil moisture (0–7 cm) that is not already captured by deeper measurements (28–100 cm)?

Thank you for your remark about multicollinearity among predictors. In response to your comment, we removed 3 variables with high correlations with temperature or soil moisture (28cm-100cm): dewpoint temperature, soil moisture (0-7cm) and soil moisture (7-28cm). We would like to highlight that the key variables and time periods identified over Europe remained mostly unchanged (apart from the variables that were removed).

To support this decision, we now include a figure showing the correlation coefficients between all hydro-meteorological variables (averaged over months and across Europe). In this figure, SM1, SM2, and SM3 refer to soil moisture at depths of 0-7 cm, 7-28 cm, and 28-100 cm, respectively, t2m to 2-meter temperature, td2m to 2-meter dew point temperature, TP to total precipitation, and SLHF to surface latent heat flux. After removing the three variables mentioned above, all remaining correlations fall between -0.4 and 0.4, an acceptable threshold for multicollinearity in this context.

We appreciate the reviewer's suggestion, which helped us refine our predictor set and strengthen the interpretability of our model. These changes are now reflected in the revised manuscript, Section 2.3 and Figure C1.

"To capture adverse hydro-meteorological conditions for forests, we select four hydro-meteorological variables as potential drivers for low-greenness events from an initial set of seven (see figure C1 in Appendix). The selection was based on two criteria: (i) limiting pairwise correlations between variables to the range of -0.4 to 0.4, a conservative threshold compared to commonly used values (e.g., Dormann et al., 2013), and (ii) prioritizing variables known to influence NDVI based on previous studies and expert knowledge (Hermann et al., 2023; Grossiord et al., 2020; Young et al., 2017; Alavi, 2002). The final set of variables includes: maximum 2-m temperature (Max T2m), soil moisture (soil moist.) at depth 28-100 cm, total precipitation (total precip.), and surface latent heat flux (s.l.heat flux)."

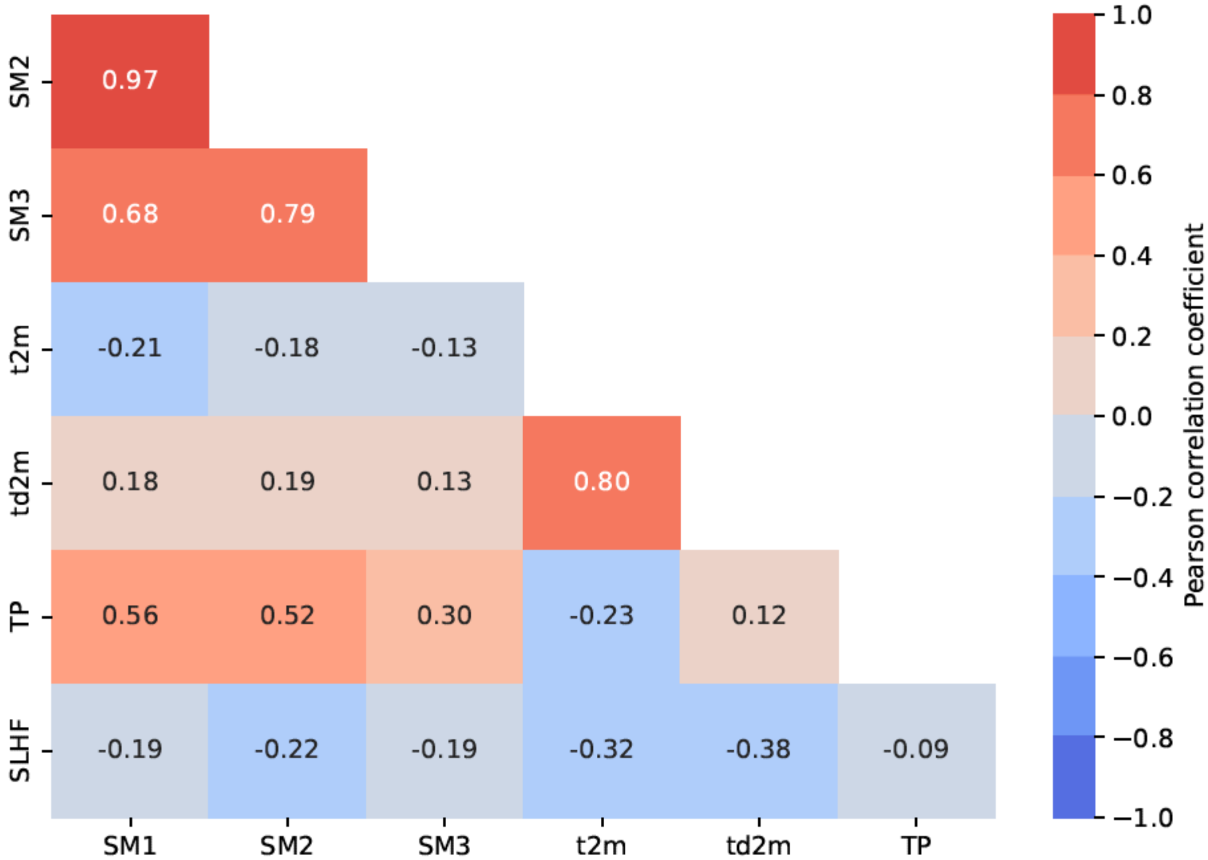


Figure C1. Pairwise correlation coefficients between hydro-meteorological predictor variables, averaged across months and over Europe. Variables include soil moisture at three depths (SM1: 0–7cm, SM2: 7–28cm, SM3: 28–100cm), 2-meter air temperature (t2m), 2-meter dew point temperature (td2m), total precipitation (TP), and surface latent heat flux (SLHF). We excluded td2m, SM1, and SM2 from the analysis to ensure that all remaining predictors have cross-correlations between –0.4 and 0.4, thereby reducing multicollinearity and enhancing model interpretability.

Interestingly, the authors themselves acknowledge the challenges associated with soil moisture data (Section 5.2), yet they opted to include not just one but three soil moisture variables in their models. Moreover, the study they cite in support of the reliability of soil moisture data—Zheng et al. (2024)—explicitly discusses the limitations of these time-series datasets, particularly their tendency to underestimate values in dry regions, such as the Mediterranean, during summer. The reduction of the number of soil moisture variables would significantly reduce the number of predictors—from 70 to approximately 50—potentially improving model interpretability without sacrificing performance.

Thank you for this point. In response, we removed the two shallow soil moisture variables (see also the comment above), reducing the number of predictors and improving model interpretability. Regarding the reliability of soil moisture data, we acknowledge the limitations, particularly in dry regions. However, our analysis focuses on anomalies rather than absolute values, and we interpret the shape of the partial dependence plots rather than exact magnitudes. This approach helps mitigate systematic biases. We have now clarified this discussion point in the revised manuscript:

"Soil moisture from reanalysis datasets can present uncertainties due to sparse in-situ measurements and heterogeneous soil properties that influence local moisture dynamics. While these limitations are well-documented, including potential underestimation in dry regions such as the Mediterranean during summer, Zeng et al. (2022) show that ERA5-Land soil moisture remains among the most reliable reanalysis products, with good agreement with observations. We chose ERA5-Land to maintain consistency across all climate variables used in the model. Moreover, we rely on soil moisture anomalies rather than absolute values, which helps reduce systematic biases and emphasizes relative changes over time. Although the quality of reanalysis data may influence model performance, the high predictive accuracy of the random forest substantially reduces concerns regarding its impact."

In a future submission, the authors should consider adding more detail to Figure 1 or including a new table that provides information on each variable, such as the measurement method, units (which are currently missing), the ecological rationale for its inclusion, and supporting bibliographic references.

Thank you for this helpful suggestion. In response, we have added units to Figure 1 for clarity. Additionally, we clarified in the caption of Figure 2 that all predictors used in the random forest model are unitless, as they have been transformed into standardized monthly and seasonal anomalies (see Section 2.3 for details). This transformation ensures comparability across variables and removes scale-related biases.

All hydro-meteorological data were extracted from the ERA5 and ERA5-Land reanalysis datasets (Hersbach et al., 2019; Muñoz-Sabater et al., 2021), which provide dynamically consistent variables with broad spatio-temporal coverage. To avoid redundancy, we retained the relevant references in the main text: "We extract these variables for the period 1980-2022 from the ERA5 and ERA5-Land reanalysis datasets (Hersbach et al., 2019; Muñoz-Sabater et al., 2021). Reanalysis data offers dynamically consistent variables with a large, uniform spatio-temporal coverage."

Furthermore, we revised the methods section to include more detail on the rationale behind variable selection:
"To capture adverse hydro-meteorological conditions for forests, we select four hydro-meteorological variables as potential drivers for low-greenness events from an initial set of seven (see figure C1 in Appendix). The selection was based on two criteria: (i) limiting pairwise correlations between variables to the range of –0.4 to 0.4, a conservative threshold compared to commonly used values (e.g., Dormann et al., 2013), and (ii) prioritizing variables known to influence NDVI based on previous studies and expert knowledge (Hermann et al., 2023; Grossiord et al., 2020; Young et al., 2017; Alavi, 2002). The final set of variables includes: maximum 2-m temperature (Max T2m), soil moisture (soil moist.) at depth 28-100 cm, total precipitation (total precip.), and surface latent heat flux (s.l.heat flux)."

As I highlighted in a previous review round, Random Forest models are not entirely free from biases introduced by highly correlated variables. Similarly, Lasso regression also has limitations and does not solve all problems involving multicollinearity. See the example of grid point (A) in France (Section 4.2.2), where two different soil moisture layers measured in the same period were identified as the most important predictors for explaining low greenness events. Assuming a strong correlation between these two variables, how will the authors be able to detect or discard the influence of multicollinearity on the final model result, or identify which variable is influencing the low greenness event?
In conclusion, I have a strong suspicion that this probable multicollinearity present in the predictor variables may have biased the Random Forest results. To help avoid this issue in future work, I strongly recommend that the authors conduct exploratory analyses (e.g., VIF- Variance Inflation Factor) to identify and address autocorrelated variables prior to model development. Alternatively, the use of Empirical Orthogonal Functions (EOFs) could be used to remove correlations in the spatial and spatiotemporal data. Another option would be to use the elastic net approach. That said, I recognize that implementing such changes is no longer feasible for the current submission.

Thank you for these insightful comments. We acknowledge the limitations of Random Forest and LASSO models in handling multicollinearity, and we appreciate your suggestions for alternative approaches. In response to your concerns, we removed three highly correlated variables, 2-meter dew point temperature, and the two shallow soil moisture layers (0-7 cm and 7-28 cm), to reduce redundancy and improve interpretability. This adjustment was guided by pairwise correlation thresholds and ecological relevance, as detailed in Section 2.3.

We chose not to apply Empirical Orthogonal Functions (EOFs) in this study, as they transform input variables into abstract components that are difficult to interpret ecologically. Given the large number of grid points analyzed, applying EOFs individually would also hinder comparability across regions. We appreciate the reviewer's suggestion to apply Variance Inflation Factor (VIF) analysis to further address multicollinearity. In our study, we now proactively mitigated this issue by removing variables with pairwise correlation coefficients larger than 0.4 or lower than -0.4. Moreover, our analysis focuses on identifying the top 10 most important predictors rather than interpreting their exact ranking. Given this focus, and the fact that Random Forest models are relatively robust to multicollinearity due to their use of random feature selection and ensemble averaging, we decided not to conduct an additional correlation analysis, VIF, to keep the focus on the interpretation of drivers. Nonetheless, we acknowledge the value of VIF and similar diagnostics for future studies aiming to refine variable selection, and we have added a note in the discussion to reflect this point and the remark on EOF:

"To address potential multicollinearity among predictors, we applied a conservative variable selection approach by removing those with pairwise correlation coefficients greater than 0.4 or lower than -0.4. This step was taken to reduce redundancy while preserving ecological interpretability. Given our focus on identifying the top 10 most important predictors, rather than interpreting their exact ranking, we did not apply additional diagnostics such as the Variance Inflation Factor (VIF, Fox and Monette, 1992). While Empirical Orthogonal Functions can be effective in reducing collinearity (North and Wu, 2001), they transform variables into abstract components that are difficult to interpret ecologically. Additionally, applying EOFs individually across thousands of grid points would hinder comparability across regions. Nonetheless, we acknowledge the value of both VIF and EOF approaches for future studies aiming to refine predictor selection for local studies."

Regarding the elastic net approach, we appreciate the reviewer's suggestion and agree that it offers a useful alternative for handling correlated predictors. The elastic net approach combines the strengths of both LASSO and Ridge regression and may be more robust in the presence of multicollinearity. However, like LASSO, the elastic net is a linear regression method. Importantly, Random Forest allows for nonlinear relationships, which are particularly relevant in ecological contexts where both low and high extremes of a variable can negatively affect forest health. This aspect is now emphasized in the abstract, introduction, and conclusion.

The Random Forest model demonstrated strong performance (for both the CSI and AUC), while offering practical advantages: it is computationally efficient across thousands of grid points, accommodates nonlinear relationships, and provides interpretable outputs that support regional comparisons across Europe.

Regarding the section 4.2.1: I still have some reservations regarding Section 4.2.1. The authors should clearly explain the criteria that guided their selection of these two specific grid points. Moreover, the wording in this section is unclear, and the absence of a theoretical foundation undermines the credibility of the results. Notably, this section lacks any bibliographic references—a shortcoming that also applies to Section 4.2.2. The authors are encouraged to engage more thoroughly with the physiological and ecological literature, which would help readers better understand how variation in the predictors relates to photosynthetic activity. While I acknowledge that some of this supporting literature is addressed in the discussion, I would like to reiterate a point raised in my previous review: the current structure of the article hampers the overall reading flow (see my next comment).

Thank you for your feedback. We have revised Section 4.2.1 to clarify the rationale behind the selection of the two GP:
"As an illustration of the RF model output analyzed in this study, we present the mean decrease accuracy and the partial dependence plots for two GPs in Europe highlighted in red in Fig. D1. These GPs were selected to showcase the diversity in forest type and the variability in driver responses captured by the RF model."
Additionally, we now specifically refer to the findings of section 4.2.1. and 4.2.2. in the discussion section, to anchor the results in existing literature and provide a theoretical foundation:
"Our model identified dry and hot conditions as adverse conditions for European broad-leaved forests, i.e., high temperature and low soil moisture (Fig. 3.1 A. and 5 A.), agreeing with existing literature (Rita et al., 2019; Beloiu et al., 2022; Rubio-Cuadrado et al., 2018; Senf et al., 2020; Knutzen et al., 2025; Schnabel et al., 2023)."
"The observed link between dry conditions (July) and warm conditions (May-July) and the increased coniferous forest browning probability (Fig. 3.1 B. and 5 B.) may similarly be linked to drought stress, "
The new structure of the discussion allows for an analysis of the results and comparison with existing literature, directly following the result section:
"5.1 Identified adverse conditions
In this section, we examine the key hydro-meteorological drivers associated with low greenness events identified by our models and compare these findings with the existing literature to contextualize and validate our results."

Lines 250-255: This paragraph would be more appropriately placed in the study area section of the methodology, as it primarily serves as a methodological justification. Its current placement disrupts the flow of the results section. I would also like to note that these concerns about section structure were already raised in the first round of

review.
Thank you, we moved this paragraph to the methods section.

Regarding the quality of the figures: Figure 5 remains difficult to interpret and requires improvement. I recommend increasing the size of each plotting panel to prevent the legend boxes from overlapping with the lines representing different latitudes. Additionally, the y-axis scaling appears inconsistent—some lines abruptly disappear from the plotting area, making it hard to follow their trends. The plotting area in map-based figures (e.g., Figure 2) could also be expanded to utilize the full page width available on an A4 sheet.
Thank you for your remark, we improved figure 5 by using the same y-axis, removing the color, and increasing the font size.

Regarding the figure captions: In my previous review, I highlighted the need for more detailed figure captions. Although some improvements have been made, several issues persist. For instance, in Figure 4, the meaning of the variable "Max Td2m" is unclear. Using the search tool, I found no explanation or definition of it in the text. Based on the first sentence of the discussion section, I infer that it refers to dew-point temperature. However, this acronym is inconsistent with the nomenclature used in Figure 1, where the variable is labeled as "Dewpoint temp." The authors should ensure consistency in variable naming and include clearer, more informative figure captions to aid reader comprehension. The same comment applies to Figure 3, where the reader must deduce the meaning of each variable acronym presented on the x- and y-axes. Another example: the authors could provide a brief explanation of the AUC and CSI values in Figure 2 to aid the reader's understanding.
Thank you very much for pointing this out. We made sure to make consistent use of acronyms for all the variables, as presented in Figure 1.
Following your remark, we added "Higher CSI and AUC values indicate better model performance." in the caption of figure 2. we would like to highlight that the CSI and AUC values are interpreted in the text (For example: "In other words, for 99% of the grid points (GPs), the model run on the testing data predicts at least as many true positives (TP) as false positives (FP) and false negatives (FN) together.").

Regarding the discussion and conclusion sections: I agree with Reviewer 1's comment that the "results are not informative, and the discussion section is lacking". Throughout the manuscript, the authors were limited in establishing comparisons with the work of Hermann et al. 2023, and the discussion lacks a strong theoretical foundation and engagement with relevant literature, particularly in the areas of forest physiology and ecology. Another notable concern is that the authors devote considerable space to discussing the limitations of their input data and modeling approach, while giving relatively little attention to the potential strengths and innovations of their methodology. Finally, the manuscript would benefit from a more detailed discussion on the practical implications of the results for forest management in Europe. Specifically, the authors should elaborate on how their findings could *inform actionable strategies to mitigate the impacts of low-greenness events in forested ecosystems.*

Thank you for these helpful points. Following your suggestion and suggestions from reviewer #1, we highlighted further the novelty and added value of our method in the abstract, introduction and conclusion

We have restructured the discussion section into: a comparison of findings with existing literature (5.1 Identified adverse conditions), methodological considerations (5.2 Preprocessing and predictor selection), and dataset limitations (5.3). This way we increase the attention given to the comparison with existing literature.

Lastly, we extended the conclusion section to elaborate on the use of our findings to inform actionable strategies:

“The identification of critical conditions and time periods at the local scale, combined with the use of hydro-meteorological variables available from sub-seasonal to seasonal forecast products, offers practical opportunities for proactive forest management. Regional forest agencies can leverage our methods and findings to anticipate periods of increased risk and implement targeted preventive measures, such as monitoring key variables in preceding seasons and years (e.g., autumn soil moisture or spring temperature). This approach enables more strategic allocation of resources, including irrigation, thinning, or pest control, tailored to the most influential local drivers. Such preventive measures would mitigate the economic and environmental costs of forest damage. ”

Specific Comments:

Lines 88–93: I appreciate the authors for including this excerpt in response to my suggestion. I believe these added sentences will greatly enhance the accessibility of the methods and results for readers who may not be deeply familiar with the subject.

Thank you

Lines 291-292: The authors open the discussion section by suggesting that dew point temperature was important in the models for identifying dry and hot conditions as adverse for European broad-leaved forests. However, this variable’s importance is not highlighted in Sections 4.2.1 or 4.2.2, nor is it clearly reflected in the main figures. The same comment applies to coniferous forests (line 302).

Thank you, as we removed dew point temperature from the predictors, we removed all the mentions of dew point temperature.

Lines 298-301: Just a note here. I respect authors' writing styles in my reviews. That said, this paragraph consists of only two sentences and directly complements the preceding one. Would it be possible to merge them for improved flow and coherence? The same comment applies to lines 367-368, a single-sentence paragraph.

Thank you for your comment, we restructured the discussion paragraph according to your comment.

Lines 302–309: The explanation of the results related to coniferous forests appears overly speculative (e.g., bark beetle infestations), which echoes my general comment from the previous round of review.
Thank you for your remark. We reformulated this paragraph, with additional references to existing literature:
"The observed link between dry conditions (July) and warm conditions (May-July) and the increased coniferous forest browning probability may similarly be linked to drought stress, facilitating bark beetle infestations on Picea abies and Pinus sylvestris (Dobbertin et al., 2007; Müller et al., 2022). In stressful conditions, conifers' fitness decreases, directly impacting resin production and thus the capacity to control insect infestation (Netherer et al., 2024). Moreover, warm conditions accelerate the reproduction rate of the insects and multiply the number of generations in the same summer (Wermelinger et al., 2008)"

Section 5.2: I commend the authors for their transparency in Section 5.2. However, several parts of this section read more like methodological descriptions and would be more appropriately placed in the methods section rather than in the discussion.
Thank you for your suggestion, we moved a part of the NDVI discussion (on the resolution of AVHRR data) to the methods section.

Before resubmitting this paper, I recommend that the authors carry out a careful review of the manuscript's grammar and punctuation (particularly regarding the use of the comma and articles).
Thank you for your advice, we reviewed the manuscript carefully before re-submitting.

Technical Corrections:
We are very grateful for the time and careful attention reviewer#2 dedicated to this extensive technical and grammar correction! We revised the manuscript following your comments below.
One exception: since"maximum 2-m temperature" functions as a general descriptor of a variable rather than referring to a specific instance, we did not add the article "the".

Line 20: Replace "occurences" by "occurrences"
Line 32: "for example" must be enclosed in commas or be deleted (my suggestion).
Line 57: "on-parametric algorithm, based" remove the comma.
Line 58: Include "to" before "pinpoint".
Lines 67-69: The authors must include at least one reference to support this statement presented at the end of this phrase.

Line 77: I believe the correct name for EUMETSAT is "European Organisation for the Exploitation of Meteorological Satellites", not "European operational satellite agency", as stated by the authors.
Line 78: Add a comma between “MetOp2” and “and MetOp3”.
Line 80: Add a comma between “instrument” and “as described”.
Line 95-96: Replace “extreme” by “extremely”. Perhaps the end of the sentence could be changed to "…Hermann et al. (2023), which binarized NDVI to focus specifically on explaining extremely low NDVI events".
Lines 110 and 111: Add a comma between “2012” and “and 2018” (in both lines).
Line 113: Delete the dot after "step 3".
Line 115: Exclude “in” before 1988.
Line 119: Add the article "a" before "longer time series".
Dew point: Throughout the manuscript, the term "dew point" is spelled in three different ways: dewpoint, dew-point, and dew point. Authors should review the entire document (including Figure 1) to use only "dew point." I also noticed that in most of the text, the authors use "dew point temperature", and in others just "dew point" (e.g., line 302).
Line 121: Consider adding the word "hydro-meteorological" between "adverse" and "conditions" (i.e., To capture potential adverse hydro-meteorological conditions for forests).
Line 124: This phrase is incomplete (We extract these seven variables…) and repeats much of the text provided in the previous sentence.
Line 128: Replace the preposition ‘’to” by “with” (…was associated to the same).
Line 142: Consider changing from "...large amount of data available spatially" to "...large amount of available spatial data”.
Line 145: Remove the article “a”: (…with a sufficient data)
Line 190: Add a comma in “1,059” and consider changing from “GPs over Europe” to “GPs across Europe”.
Line 200: Consider changing from "... (forest damage), and xiC are all the other predictors in the training data and n is the length of the training data." to "...(forest damage), xiC is all the other predictors in the training data, and n is the length of the training data".
Line 225: Replace the preposition ‘’4” by “four” (The four most important…)
Line 230: Add a comma between “before” and “increase”
Line 234: Replace the preposition ‘’3” by “three” (The three most important…)
Line 241: Add a comma between “small” and “and” (are relatively small, and several…)
Lines 243-249: Review/include the articles before the variables (e.g., predictor is the maximum 2-m temperature…).
Line 291: Add a comma between “i.e.” and “high temperature”
Line 295: Consider changing from “…,which increased consistently with temperature over these last decades in many regions” to “… “which has increased consistently with temperature over the last decades in many regions”.
Lines 296-297: Consider changing from “and once soil is dry, plants have to close

their stomates" to "and once the soil is dry, plants have to close their stomata"
Lines 313-314: Consider changing from "…at European scale are set on flat areas with important water reserve or available groundwater table." to "…at the European scale are set on flat areas with important water reserves or an accessible groundwater table."
Line 375: Replace "measurement" by "measurements"
Line 388: Add a comma between "that" and "given"