Response to Reviewer #2 - João Teixeira

R2.0 This study investigates the use of large ensemble climate simulations to drive the LPJmL-SPITFIRE dynamic global vegetation-fire model to better characterize extreme wildfire events under present-day climate conditions. The authors compare fire simulations driven by a 40-member Single-Model Initial-Condition Large Ensemble (SMILE), a single ensemble member, and reanalysis data, and assess their ability to capture extremes in fire danger index (FDI), burned area (BA), and fire carbon emissions (FireC). The results demonstrate that large ensembles sample a broader range of physically plausible wildfire extremes than single climate realizations and provide new insights into the relationship between extreme fire weather and fire impacts.

This manuscript addresses an important challenge and presents a valuable contribution to wildfire risk assessment and presents a novel application of large ensemble climate-impact modelling to process-based global fire simulations. The methodology is clearly described, and the manuscript is generally well written and well structured. However, the discussion would benefit from a deeper interpretation of the mechanisms driving the simulated extremes and a clearer assessment of the limitations associated with the experimental design, particularly regarding the treatment of fire ignitions and the interpretation of regions where deterministic simulations exceed the ensemble maxima. Subject to these revisions, the study would provide a useful contribution to the literature on wildfire extremes and large ensemble climate-impact modelling.

We thank the Reviewer João Teixeira for the overall assessment. Below we individually respond to the referee comments with our responses in blue. Specifically, we suggest revisions to the discussion section accordingly, providing a deeper interpretation of the mechanisms driving the simulated extremes and the limitations raised by the reviewer. The revised text is indicated in ***bold italic blue*** following the answer to each individual comment.

R2.1 The manuscript clearly demonstrates that the single-fire and reanalysis-fire simulations generally undersample the most extreme FDI, BA, and FireC compared to the SMILE-fire simulations. However, the discussion focuses primarily on that undersampling occurs rather than what drives it. While the broader sampling of internal climate variability is presented as the main explanation, the manuscript also shows that extreme BA and FireC are not solely controlled by FDI. I encourage the authors to further discuss the processes driving the undersampling, particularly given that BA and FireC exhibit substantially larger differences than FDI.

Thank you for this comment. We agree that is important to distinguish the different mechanisms the reviewer is pointing out. The first is a sampling effect: the SMILE-fire covers a broader range of internal climate variability, and therefore captures rarer, more extreme events than the reanalysis-fire or single-fire simulations can. This is the general mechanism that explains why SMILE-fire simulations sample more extremes of FDI, BA, and FireC. The second mechanism is how this internal climate variability propagates through the LPJmL-SPITFIRE model chain. The DGVM models fire through the following steps in succession (Appendix B, Drüke et al., 2019; Thonicke et al., 2001; Oberhagemann et al., 2024):

1. The potential for natural and anthropogenic ignitions is calculated based on lightning flashes and population density
2. FDI is calculated based on VPD, foliage projective cover (FPC), plant functional type (PFT), and precipitation (Drüke et al. 2019). The potential ignitions are then scaled by the FDI, and therefore transformed in potentially successful ignitions

3. These potential successful ignitions start fires, which spread by the wind-based Rothermel equation, and use fuel loads obtained from the LPJmL component
4. The resulting fire effects of vegetation mortality and fuel consumption feed back into the vegetation component for subsequent fires

In our interpretation, this fire-vegetation interaction is why propagating internal climate variability in LPJmL-SPITFIRE has a larger effect on burned area and emissions than on FDI. FDI itself fluctuates from one ensemble member to another, both because of differences in the initial conditions in temperature, humidity, and precipitation, and because FDI is scaled with the dynamic vegetation components FPC and PFT. However, burned area and emissions are affected by additional layers of the model chain: on a daily timestep, fire spread is affected by wind and fuels, and the fire itself alters the fuel bed and vegetation state that subsequent fires depend on. The cumulative effect of representing a broader range of internal climate variability with the mechanistic fire-vegetation interaction, rather than the sampling effect alone, explains the larger differences between SMILE-fire and single-realization simulations in BA and FireC compared to FDI.

We agree this requires more clarity in the manuscript, and will make improvements to the Appendix B clarifying the LPJmL-SPITFIRE components, and in the Discussion section better examining the processes behind the larger differences in BA and FireC rather than FDI:

Appendix B: *'SPITFIRE considers two types of ignition sources: lightning flashes and human ignitions determined based on population density (see Eqs. 3 and 4 in Thonicke et al. 2001).* ***Following a potential ignition event, the number of ignitions is scaled using a fire danger index (FDI), representing the proportion of ignitions that successfully become spreading fires.*** *While previous versions of SPITFIRE used the Nesterov Index (based on maximum temperature, dew point temperature and the different plant functional types) for calculating fire danger, recently Druke et al. (2019) implemented a VPD-based FDI.* ***The formulation of the VPD-based FDI by Druke et al. (2019) combines air temperature, relative humidity, precipitation, and vegetation density dynamically simulated by LPJmL.*** *The duration of the fire is determined by the FDI and assumes a longer burn time under high fire danger. The Rothermel equations are employed to determine wind-driven fire spread, which also varies according to plant functional types. Finally, the model uses the simulated fire danger, ignitions, and spread to estimate burned area and fire carbon emissions. The model restricts the simulation of wildfires occurring to natural vegetation, excluding managed land, such as agriculture, due to fundamentally different fire dynamics in this type of land (Ribeiro et al. 2024).* ***As shown in Fig. B1, the main components of SPITFIRE consist of (1) ignition module, (2) fire danger module (3) fire spread module, and finally (4) fire impacts (burned area, fire carbon emissions and vegetation mortality). Together, these 4 components determine the successful ignitions (fire occurrence), the fire danger (intensity and duration), fire spread (wind-driven Rothermel model) and the resulting impacts (burned area, emissions) which later feedback into the vegetation component of LPJmL for subsequent time steps.'***

Discussion: ***'While it is expected from sampling theory that a larger sample such us SMILE-fire samples more extremes, the contribution of demonstrating this using a mechanistic process-based vegetation-fire model carries implications beyond the sampling effect itself. From the climate hazard to the impact level, the burned areas and fire carbon emissions are more underestimated by single realizations relative to the large ensemble than FDI is,*** *suggesting that the SMILE-fire simulations are particularly valuable for studying fire impact variables (burned area and fire carbon emissions) compared to the climate driver (fire danger). In other words, smaller changes in fire danger extremes due to internal variability have the potential to trigger stronger responses in burned area and emissions.* ***We interpret this as nonlinearities and vegetation-fire interactions that originate from the SMILE-LPJmL-SPITFIRE model chain. First, the input SMILE preserving internal climate variability determines the FDI together with dynamically simulated vegetation density. Then, the potential ignitions (natural and anthropogenic) are scaled by the FDI into successful ignitions, which spread using dynamically simulated fuel loads, and the resulting fuel consumption and vegetation mortality feed***

back into the vegetation component for subsequent fires (see Appendix B). Hence, while FDI mainly reflects initial condition differences among the climate ensemble members, burned area and emissions also depend on vegetation and fuel availability. Higher fire danger can increase the likelihood of successful ignitions and, once a fire is established, favour faster and more extensive spread under favourable fuel and weather conditions. Consequently, relatively small differences in fire danger can result in larger differences in the number and extent of fires, which then produce larger differences in fuel consumption and vegetation mortality.

This same model chain effect also explains why the degree of undersampling differs among burned area and fire carbon emissions. *For most regions, the maxima of single realizations rank below those of SMILE-fire, with fire carbon emissions from single realizations showing the largest undersampling.* ***While burned areas mostly result from the Rothermel-based rate of spread, fire carbon emissions depend on the burned areas and additional modeling layers.*** *In our interpretation, fire carbon emissions are particularly more sensitive because they depend not only on burned area but also on fuel load and combustion completeness, and thus exhibit higher sensitivity to the role of internal climate variability on the process-based modeling chain. This interpretation is also consistent with the pairwise correlation structure among FDI, BA and FireC: the strongest relationship is between BA and FireC, with more modest relationships between FDI and BA or FireC, reflecting the additional modeling steps that accumulate along the model chain.'*

R2.2 Similarly, the regions where the single-fire or reanalysis-fire simulations exceed the SMILE-fire maxima (e.g., NHAF, CEAM, and BONA) deserve further discussion, as these exceptions may provide additional insight into the processes controlling extreme fire events and the limitations of the ensemble framework.

Thank you for raising this valuable point. First, we would like to clarify that, in terms of percentage relative difference in the maxima values, the few exceptions where SMILE-fire does not sample the greatest value occur only in comparison with reanalysis-fire simulations, not single-fire. Second, these exceptions also occur only in burned area and fire emissions, not in FDI (Figure4b, c). These clarifications are important for three reasons:

First, reanalysis-fire simulations are driven by what we consider observed historical climate conditions and should therefore be able to capture a genuine extreme fire year that actually occurred. In contrast, the single-fire simulation, driven by the r1i1p1f1 climate ensemble member, is more likely to simply not hit those observed extremes. Nevertheless, single ensemble members of SMILEs are what is typically used in model intercomparison projects such as ISIMIP-FireMIP, and here we show how extremes are undersampled under that setup.

Second, in our interpretation, what primarily happened in burned area, and especially fire emissions, in NHAF, CEAM, and BONA is that, despite its broader sampling of internal climate variability, SMILE-fire did not happen to generate those three regional extreme years in any of its members. This does not indicate a limitation in the ability of SMILE-fire to sample extremes in general but rather reflects the fact that any finite ensemble size, however large, cannot guarantee it reproduces every localized historical event. Nevertheless, for the remaining regions and variables, SMILE-fire does sample more physically plausible extreme values than the reanalysis-fire and single-fire simulations, emphasizing its capacity to capture extremes that could have been possible under current climate conditions, and that single-realizations did not simulate.

Third, the fact that these exceptions occur in BA and FireC but not in FDI (and reanalysis-fire and not single-fire) reinforces the point raised in our response to the previous comment R2.1: in the model, burned area and fire emissions are not solely controlled by climate, but also by fuel, vegetation states, ignitions. In these regions, the reanalysis-driven fuel and vegetation iteration may have aligned with conditions that favored unusually large burned areas and emissions, even though the SMILE-fire FDI maximum was not exceeded.

We further inspected the time series of FDI, BA and FireC in these three exceptions (Figure 1). The main conclusions are:

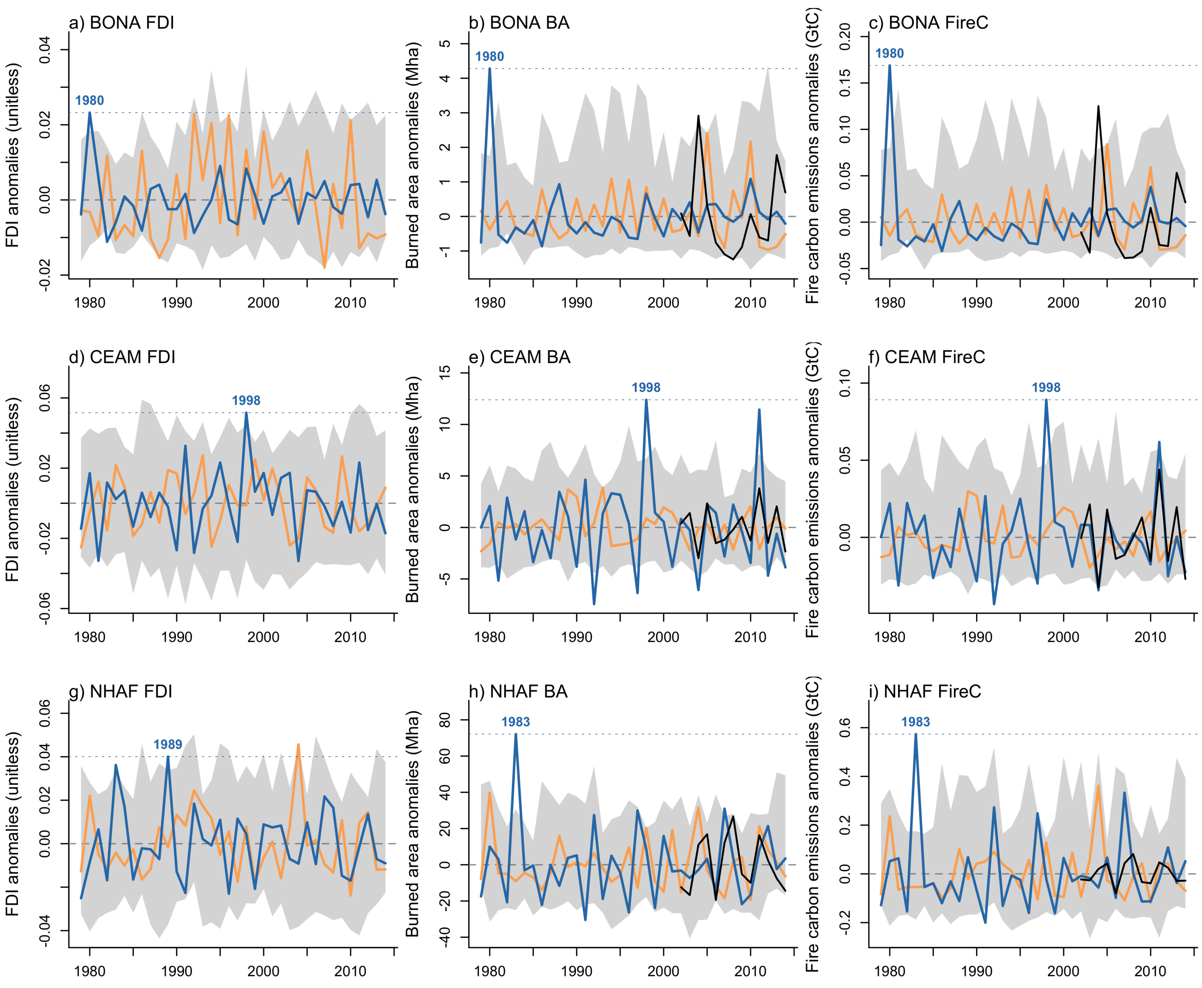

Figure 1: Same as Figure 2a-c in the manuscript, for BONA (a-c), CEAM (d-f) and NHAF (g-i). Time-series show (a,d,g) annual maximum fire danger index (FDI), (b,e,h) annual total burned area (BA), and (c,f,i) annual total fire carbon emissions (FireC). Four datasets are shown: the 40-member large ensemble labeled 'SMILE-fire' (gray shading, full ensemble range), a single ensemble member labeled 'smile-fire'(yellow), a reanalysis-fire simulation labeled 'reanalysis-fire' (blue), and GFEDv5.1 reference data (black).

1. SMILE-fire captures a broader range of FDI values than the reanalysis-fire maximum in all three regions
2. SMILE-fire does not capture the 1980 extreme BA and FireC in BONA, the 1998 extreme BA and FireC in CEAM, or the 1983 extreme BA and FireC in NHAF.
3. BONA 1980, CEAM 1998, and NHAF 1983 extremes in BA and FireC tend to co-occur with extreme FDI values (although in NHAF, the maximum reanalysis-fire FDI occurs in 1989, but 1983 is the second highest).

Interestingly, 1983 (Figure 1 h,i, NHAF) and 1998 (Figure 1 e-g CEAM) coincide with well-documented strong El Niño events, while 1980 (Figure 1 a-c BONA) was more a moderate El Nino event (`https://psl.noaa.gov/enso/past_events.html`). The year of 1983 is also popularly known for impacts in synchronized crop failures around the globe (Anderson et al., 2019), and the El Niño–Southern Oscillation (ENSO) has a strong influence in intra- and interregional synchronize extreme fire weather (Yin et al., 2026).

So, this may explain the elevated FDI anomalies in these years in the reanalysis-fire, and that SMILE is likely to capture these extreme climate patterns given its broader sampling of internal variability. However, the SMILE-fire are independent realizations of internal variability and 'SMILE years' are not synced with real world years (because the SMILE is not nudged to the observed climate, and therefore not temporally synchronized by default). So further down the model chain (i.e. burned area and emissions calculations), these exceptions may simply indicate that by chance the SMILE did not simulate the very specific extreme conditions (e.g. sufficient fuel availability, a favorable vegetation state, wind speed and ignitions at the time) of those years in those locations. A finite ensemble, however broad its climate sampling, is not guaranteed to reproduce every individual localized historical event. The main conclusion that SMILEs are a powerful tool for sampling plausible extremes beyond the historical record holds nonetheless.

We will add this figure to the Appendix and revise the discussion paragraph about these exceptions accordingly: *'The potential for extremes beyond those captured by reanalysis and a single member drawn from SMILE-fire varies markedly by region. At the regional scale, the SMILE-fire captures higher maxima of fire-related variables than both the reanalysis and single-member in nearly all GFED regions, except for boreal north America (BONA), central America (CEAM) and northern hemisphere Africa (NHAF).* ***In these three exceptions, SMILE-fire captures a broader range of FDI values than the single-realizations maximum, but does not reaches the corresponding extreme burned area and fire carbon emissions of reanalysis-fire (1980 in BONA, 1998 in CEAM, and 1983 in NHAF). These extreme years in reanalysis-fire coincide with documented El Niño events, which indicates that because SMILE-fire does not follow exactly the actual observed historical sequence, it is not guaranteed to reproduce this specific combination of extreme conditions in the LPJmL-SPITFIRE output.*** *In addition, though modest, these exceptions may be related with factors relevant for the stochasticity of fires in boreal and tropical regimes that are not incorporated in our large ensemble scheme. For example, it could be the case that for the same probability of lightning (which is not different across input ensemble members in our case), any of the SMILE members synchronizes the rare combinations of fire danger and fuels captured by the reanalysis. In some cases, the single-member simulations might also include the most extreme event of the SMILE by chance. Although a finite number of ensemble members cannot fully encompass all physically possible climate-fire interactions, the ensemble of 40 members used here sample a substantially broader range of climate trajectories than single realizations, consistently revealing stronger impacts across most regions.'*

We also noted that Figure 4C required larger interval in the xlim of panel b) to make the NHAF reanalysis-fire outlier visible, and will update the figure accordingly in a revised version.

Additional reference used for the purpose of this answer: Anderson, W. B., Seager, R., Baethgen, W., Cane, M., & You, L. (2019). Synchronous crop failures and climate-forced production variability. Science Advances, 5(7), 1–10. https://doi.org/10.1126/sciadv.aaw1976

R2.3 The experimental design prescribes lightning ignitions from the OTD/LIS climatology while only the meteorological forcing varies across the ensemble members. Consequently, natural ignitions are not dynamically consistent with the simulated weather, and climate-driven variability in lightning occurrence is not represented within the ensemble. I encourage the authors to discuss the implications of this assumption more explicitly, particularly regarding its influence on the interpretation of the simulated extremes and the extent to which the ensemble captures climate-driven variability in wildfire occurrence.

Thanks for the suggestion and the comment, a similar point was also raised by Reviewer #1 and we agree it deserves more discussion and clarification in the revised version. As the reviewer notes, (1) lightning occurrence is not dynamically consistent with the climate variables of SMILE-fire (because the SMILE is not nudged to the observed climate, and therefore not temporally synchronized by default) and (2) the raw natural ignition potential is the same across individual ensemble members, which means that the natural lightning does not change with the meteorology of the climate model exactly. Nevertheless, the raw potential ignitions are subsequently scaled by FDI (please read the response R2.1 about the LPJmL-SPITFIRE modelling steps that are improved in the revised version of the Appendix), which means that ignitions inherit some variability across ensemble members through this scaling step, even though the raw ignition potential does not. Because they are scaled with FDI, the scaled ignitions represent potential 'successful' ignitions. In this sense, the ensemble does partially represent climate-driven variability in ignitions, but only as it is mediated through fire danger, not through any climate-driven variability in lightning frequency itself.

We acknowledge that this is a simplification in the methodological set-up: ideally a fully dynamically consistent design would vary lightning occurrence with each ensemble member meteorology. However, lightning data are not readily available as standard climate large-ensemble output, and this setting is already novel in contrast to standard process-based modeling.

As of the interpretation of the simulated extremes, in our set-up they are mostly driven by the variability specific to each ensemble member in fire danger and its downstream effects on the dynamical vegetation-fire mechanisms, rather than by the variability in lightning ignitions of each ensemble member. In years or regions where lightening ignition are part of the primary controls on fire occurrence, our experimental design may therefore underrepresent the full extreme range that could arise if lightning variability were also allowed to vary dynamically across ensemble members. This is a plausible additional explanation for the exceptions discussed in our response to R2.2, particularly in regions where ignitions, rather than fire spread, are the limiting factor.

We will revise this point in the discussion accordingly: *'Beyond hydrometeorological drivers, topography and fire ignition assumptions are aspects that future work could further address. Satellite-derived lightning products have inherently limitations (Christian et al., 2003) and human-caused ignition based on population density (Thonicke et al., 2010) does not fully capture human ignition behavior driven by, for example, land management practices or socioeconomic factors (Ribeiro et al., 2024; Forrest et al., 2024. Furthermore, lightning ignition probability is held constant across input ensemble members, and thus does not respond to changes in convective activity associated with different climate states.* ***Nevertheless, because these raw ignition potentials are subsequently scaled by the FDI before becoming successful ignitions (see Appendix B1), successful ignitions do inherit some variability from each ensemble member through this scaling step, even though the underlying ignition potential does not. As a result, the simulated extremes in this study are driven primarily by the variability in FDI and the vegetation-fire feedback loops, rather than by variability in ignition occurrence itself. Therefore,*** textit while fire weather favorable conditions across the ensemble members amplify the fire outcomes of the same prescribed ignitions, lightning cannot itself act as a source of climate-driven variability in the simulations. ***Hence, in ignition-limited regions, our experimental design may therefore underrepresent the full extreme range that could arise if lightning variability were also allowed to respond dynamically to the simulated climate of each ensemble member.*** *Future work could include climate model output of lightning-related variables to allow ignition probability to respond dynamically to the climate state, enabling a more complete representation of natural climate variability in fire activity (Janssen et al., 2023).* ***However, this is usually not standard output in SMILEs.'***

Minor comments

R2.4 L22–23 – Consider replacing "global warming" with "climate change", as the latter is more appropriate in this context given that wildfire activity is influenced by multiple climatic drivers, not only increasing temperatures.

We agree with the suggestion and will revise accordingly.

R2.5 Section 2.3 – Given that the manuscript focuses on extreme wildfire events, it would be useful to briefly discuss the ability of LPJmL-SPITFIRE to reproduce the upper tail of the observed BA and FireC distributions, in addition to the mean variability presented in Figure 2.

Thank you for the comment, this question was also raised by Reviewer #1 and we agree that this is a relevant point. In addition to Figure 2, the Appendix Figures C6, C7 and C8 illustrate the annual maximum values of FDI, BA and FireC at the pixel level across the different simulations. In particular, Appendix Figures C7 and C8, compare the burned areas and fire emissions simulated by LPJmL-SPITFIRE against GFEDv5.1. Both the maps and the distributions when pooling all pixel values of the temporal maxima (C7f and C8f) show (1) a good spatial agreement in the upper tail values and (2) that, as expected, SMILE-fire maxima values are consistently larger than those in single simulations.

We agree that this information is currently not well represented in the manuscript and will revise accordingly to better highlight the ability of LPJmL-SPITFIRE to reproduce the upper tail of the GFEDv5.1BA and FireC distributions. After line 219, section 3.1, we will add a new last paragraph: ***'The ability of LPJmL-SPITFIRE to reproduce the spatial variability of the GFEDv5.1 BA and FireC extremes is shown in Appendix Figures C7 and C8. In general, a good spatial agreement is found at the pixel level. Although BA and FireC show highly skewed distributions (Figures C7e and c8e), reanalysis-fire and GFEDv5.1 agree closely in the log-transformations when pooling all the maxima values of all pixels (Figures C7f and c8f), and SMILE-fire shows a heavier upper tail consistent with its broader sampling of internal climate variability.'***

In addition, we will revise the discussion paragraph that addresses the SPITFIRE challenges in modelling extremes: ***'While LPJmL-SPITFIRE produces reasonable estimates of mean and extreme burned area and fire carbon emissions, several aspects of extreme fire behavior remain less explored.*** *In particular, SPITFIRE represents surface fire spread but does not explicitly simulate crown fire or canopy fire spread, which are characteristic of the most extreme fire events in boreal and other fire-prone ecosystems (Ward et al., 2018). Additionally,* ***as the live fuel moisture levels are calculated based on a growing season index (Oberhagemann et al., 2025)****, including the influence of wind and precipitation on live fuel moisture computation would further contribute to improve extreme fire danger conditions and to separate the influence of wind speed on fire spread from the influence of atmospheric drying of the fuel. (...)'*

R2.6 Figures 3 and 4 – The information shown in the upper maps is largely redundant with the regional boxplots below, as both present the regional values of Δmax. Instead, I suggest displaying the grid-cell values of Δmax in the maps. This would better illustrate the spatial variability within each GFED region and provide additional insight into where the largest differences between the single-fire and SMILE-fire simulations occur.

Thank you for this comment. We would like to clarify that the maps and boxplots are intended to convey complementary, rather than redundant, information. While the upper maps show the regional relative difference between the maxima values (p100) of each single realization and the pooled SMILE-fire simulations (Δmax value), the boxplots (Figure 3d-f, 4d-f) provide more statistical detail showing the full distribution of the single-fire simulations in contrast to increasing extreme percentile thresholds (p95, p97.5, and p99) of the pooled SMILE-fire.

First, the boxplots and asterisks allow to show that SMILE-fire sampling is progressively relevant toward the most extreme values. Second, the p97.5 (purple asterisk) is the closest to the expected return period of the maximum value from the 36-year single realizations (1-in-36 years). So, we believe that it's important to situate that SMILE-fire percentile in contrast to the maximum of the single realizations. In fact, in many regions, the 97.5th percentile of the SMILE-fire exceeds the maximum of the single-realizations (not just the p99, black asterisk in the boxplots, and the p100 in the maps).

We agree, however, that displaying Δmax at the grid-cell level, as suggested, would offer additional insight into spatial variability within each GFED region. We explored this at the pixel scale but found the resulting patterns too noisy to interpret meaningfully. Part of that is because at the pixel scale, not all pixels burn frequently every year, and many pixels may have near-zeros even in the maxima. Since Δmax is the relative difference between two maxima, dividing by a near-zero can produce unmeaningful ratios. The regional aggregation avoids this artifact because regional burned area totals are essentially never zero. Additionally, these GFED regions are a standard widely used regionalization in global scale fire studies and calculating Δmax at this scale is consistent with the scale at which DGVMs fire model performance is typically evaluated and compared across the literature.

We therefore suggest retaining the regional maps and the boxplots, which we believe still provide useful pyrogeographic context alongside the more detailed statistical comparison shown in the boxplots. We also suggest leaving the results section regarding the boxplots unchanged, in section 3.2 Lines 237-246.

R2.7 Figures 3 and 4 – The current map projection makes some regions difficult to visualize due to the wrap-around at the map boundaries. Consider using a projection such as Robinson, which would improve the readability of the global spatial patterns.

Thank you, we propose making changes to improve the readability of the maps in Figures 3 and 4, including switching the map projection to Robinson, removing the meridian and parallel grid lines and repositioning the regional labels.

R2.8 Figure 4 – The caption states "Same as Figure 4 but for the reanalysis-fire." This should read "Same as Figure 3 but for the reanalysis-fire."

Thank you, we will update the caption accordingly.

R2.9 L242–245 – The discussion of the p97.5 and p99 thresholds would benefit from a brief reminder that these return periods are estimated from the pooled ensemble and should not be interpreted as true return periods.

Thank you for the comment, a similar question was also raised by Reviewer #1 and we agree that this is a relevant point to clarify. We use the return period language only to facilitate the interpretation of the percentile thresholds, by translating them into an intuitive frequency framing, to highlight that *'Of these, p97.5 is the closest to the expected return period of the maximum value from the 36-year single realizations (1-in-36 years).'* (Lines 166-167). As we do not report the return period values themselves, but rather use this language as auxiliary interpretation, we will revise L242-245 and place the return period language in parenthesis instead, and clarify that they are obtained from the pooled SMILE-fire:

'Even for the ***p97.5 percentile (equivalent to a 1-in-40-year events calculated from the pooled SMILE-fire)****, which are closer in frequency to the most extreme events expected in a 36-year single realization, the SMILE-fire ensemble exceeds the upper whisker of the boxplot (defined as 1.5 × interquartile range above the 75th percentile) in several regions (despite not always surpassing the highest individual outlier).'*

Additionally, in reply to Reviewer 1 we will also add a sentence in the end of the section 2.3.1 to clarify the usage and pooling of large empirical samples to characterize extreme percentiles: ***A central advantage of using large ensembles is that they provide a sufficiently large empirical sample to characterize extreme values without relying on parametric assumptions. To calculate the percentiles, we treat the 40 members as statistically exchangeable and sufficiently independent, consistent with the standard use of SMILEs in literature (Deser et al., 2020; Maher et al., 2021; Tebaldi et al., 2021; Bevacqua et al., 2023). This approach treats each member exchangeably, differing only in their initial conditions of the forcing SMILE. Although they share the same external forcing, we assume each member can still be considered sufficiently independent because we focus on detrended LPJmL-SPITFIRE output to calculate the percentiles. In terms of ensemble size, the 40 members are within***

the range identified for robust estimation of extreme values for climate variables using SMILEs (Tebaldi et al., 2021).

R2.10 L322–329 – The discussion of the limitations of the VPD-based FDI is useful. It may also be worth briefly acknowledging that it does not explicitly account for wind, despite wind being one of the primary controls on fire spread.

Thank you, a similar question was also raised by Reviewer #1 and we agree that the manuscript can benefit from a better discussion about the implications of the VPD-based FDI.

We will adapt the description of the different modelling steps of LPJmL-SPITFIRE in Appendix B (see R2.1) and also mention the limitations associated with VPD-based FDI as the single climate driver in our analysis regarding driver-impact relationships: *'There are some limitations of the present study to note.* ***In contrast to older versions of LPJmL-SPITFIRE which used the Nesterov Index to represent fire danger, the current version of FDI is based on VPD, precipitation and plant functional types (PFT)-specific tuning parameters. Despite being one of the primary controls of fire, we acknowledge that the VPD-based FDI does not explicitly account for wind, because in LPJmL-SPITFIRE the fire spread is calculated after the FDI via the Rothermel model.*** *Thus, while LPJmL-SPITFIRE framework does account for meteorological variables beyond VPD, their contributions are not explicit in what we refer here is as 'climate driver', as we define it through FDI (VPD-based). Disentangling the relative roles of other climate drivers beyond FDI constitutes an important avenue for future work. Particularly, assessing the ensemble-based relationships between climate drivers and extreme impacts (burned area and carbon emissions) could be generalized to a broader set of climate variables beyond FDI, including wind speed,* ***soil moisture****, and large-scale atmospheric circulation patterns. Such an approach would allow exploring physically self-consistent combinations of* ***compounding*** *drivers governing extreme fires, in line with physical climate storyline methods.'*

R2.11 L352 – "furter contribute" should be "further contribute."

Thank you, will be revised.