Author Responses to Referee RC2

Manuscript: EGUSPHERE-2026-136

All page and line numbers refer to the originally submitted manuscript.

Review of Zhu et al. "Quantifying the current and future likelihood of the 2022 extreme wildfires weather conditions in France with anthropogenic climate change"

Summary

This paper describes an event attribution analysis of fire weather conditions associated with three 2022 fires in the southwest of France and how the likelihood of these conditions changes under a future warming scenario. The authors consider multiple event definitions to test the sensitivity of their results. I find the concept of this study to be worthwhile and the manuscript was generally well written and easy to follow. My main concerns are about the robustness of the results.

General comments

How do you account for the (sometimes large) difference in resolution between the CMIP6 models and the observations?

Response: We agree that the resolution difference between the observational datasets and the CMIP6 models warrants more explanation. In our analysis, the observational data (SAFRAN at 8 km) is used solely to estimate the exceedance probability (p_{OBS}) of a given wildfire; it is not used for a direct comparison with, or evaluation of, CMIP6 outputs. The attribution analysis is then carried out independently with each CMIP6 model by comparing the ALL and NAT simulations at the model's native resolution. Because the key result relies on the ratio of p_{ALL} to p_{NAT} (and not a direct model–observation comparison), no downscaling was necessary. We clarified in the Methods section.

Inserted at p. 6, L114: "For the attribution analysis, each CMIP6 model is analysed at its native spatial resolution. The observational dataset (SAFRAN) is used only to estimate the exceedance probability of the observed event (p_{OBS}), not for a direct model–observation comparison."

Revised text p. 6-7, L136-144: "To quantify the impact of anthropogenic climate change (ACC), we employed a commonly used approach to calculate the exceedance probability of each wildfire-related FWI (p_{OBS}) (Barbero et al., 2020), following the procedure applied in the previous section to the 1959–2023 SAFRAN observations. Here, SAFRAN is used only to estimate the exceedance probability of the observed event (p_{OBS}) and is not compared with the CMIP6 simulations. The attribution analysis is then performed independently with each CMIP6 model, at their native spatial resolution, by comparing the ALL and NAT experiments. We then compared the exceedance probabilities under two scenarios: (i) the ALL scenario, including all anthropogenic and natural forcings (hereafter p_{ALL}), and (ii) the NAT scenario, which includes only natural forcings (hereafter p_{NAT}). For the GEV distribution fitted to the ALL-scenario simulated annual maxima of the MA-FWI time series, we inverted its cumulative distribution function (CDF), FALL(x), to find the FWI level in the p_{ALL} scenario such that $1 - F_{ALL}(FWI_{ALL}) = p_{OBS}$. We then applied this same threshold FWIALL to the GEV distribution fitted to the NAT-scenario simulated annual maxima of the MA-FWI time series – using its CDF $F_{NAT}(x)$ – to compute $p_{NAT} = 1 - F_{NAT}(FWI_{ALL})$. Because the attribution metric relies on the within-model ratio between p_{ALL} and p_{NAT}, no spatial downscaling was required."

Revised text p. 12, L203-205: "Finally, we examined how ACC altered the probability of those FWI conditions. Figure 6 illustrates, for one model and one spatiotemporal set-up (NorESM2-LM, r1i1p1f1), how pNAT varies relative to the reference probability pOBS (= pALL by construction), from which RR and FAR were derived."

If I understand correctly, this study calculates the risk ratios for four different model realizations individually and then takes the median across these results. My concern is that the uncertainty for each is high and the range across the models is large, so when the median is reported without an uncertainty range, it conveys a level of confidence and precision that I don't think is there.

Response: We agree with the reviewer that the uncertainty within each individual model as well as the uncertainty across the models is substantial. In the original Figure 7, we chose to display only confidence intervals for the four individual model realizations, as adding an additional uncertainty band for the multi-model median in the same panel would have made the figure difficult to read. However, we added a supplementary figure (see below) showing the multi-model median RR together with its pooled ensemble uncertainty range. For each year, we pooled together the bootstrap replicates from all models into a single set of 400 samples (4 models × 100 bootstrap replicates each), computed the 5th and 95th percentiles of the pooled set as the lower and upper bounds of the uncertainty range. This pooled approach is intended to reflect both the within-model sampling uncertainty and the spread across models.

Revised text p. 7, L157: "Finally, attribution scores (RR and FAR) from individual models were summarized across models using a multi-model median. In that case, we pooled the 100 bootstrap replicates from each of the four models. This pooled ensemble range is intended to reflect both within-model sampling uncertainty and the spread across models.

The parametric bootstrap approach was implemented as follows:

1. Generate new samples of ALL and NAT scenarios from the estimated non-stationary GEVs.

2. Re-estimate the non-stationary GEVs based on these new samples and compute the RR and the FAR.

3. Repeat steps 1-2 100 times to derive model-specific parametric confidence intervals and the pooled ensemble uncertainty range described above."

New Supplementary Figure (Fig. S6):

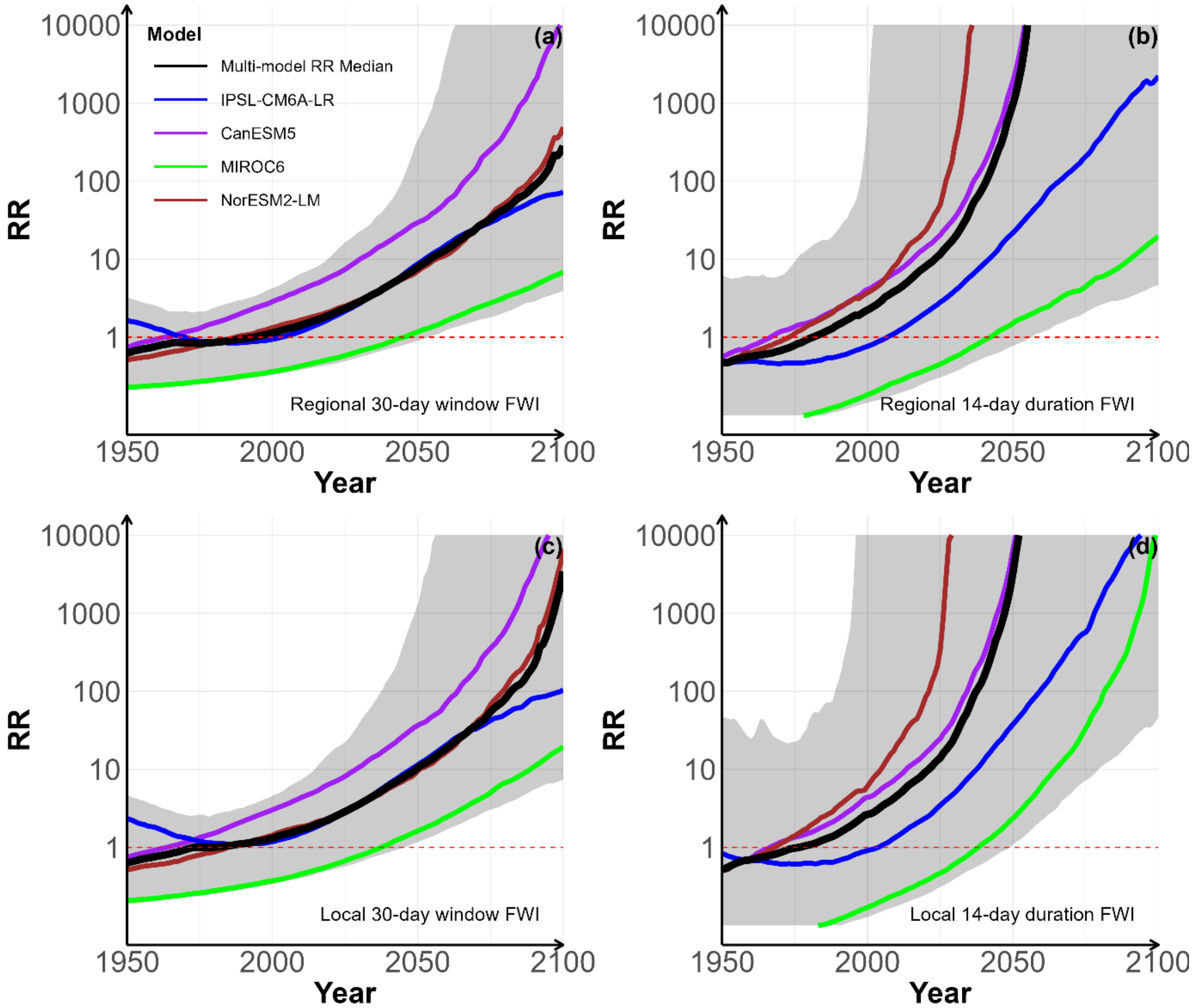

Figure S6. Same as figure 7 except that the uncertainty range was computed for the multi-model median RR. For each year, the grey shaded envelope denotes the corresponding 5th–95th percentile range of the pooled bootstrap replicates from four models (4 × 100 = 400 samples). This pooled range reflects both within-model sampling uncertainty and inter-model spread.

We also revised Figure 8. In the original manuscript, the confidence intervals for the multi-model median FAR were computed by taking, for each bootstrap replicate, the median FAR across the four models, thereby forming a set of 100 cross-model median values, and then using the 5th and 95th percentiles of that set as the confidence bounds. While this approach quantifies the uncertainty of the median itself, it does not fully retain the within-model variability and may therefore underestimate the total uncertainty. We have now replaced this with the same pooled ensemble uncertainty range method described above. The updated Figure 8, its caption, and the related text in the manuscript have been revised accordingly to clarify this treatment of uncertainty.

Revised figure (Fig. 8):

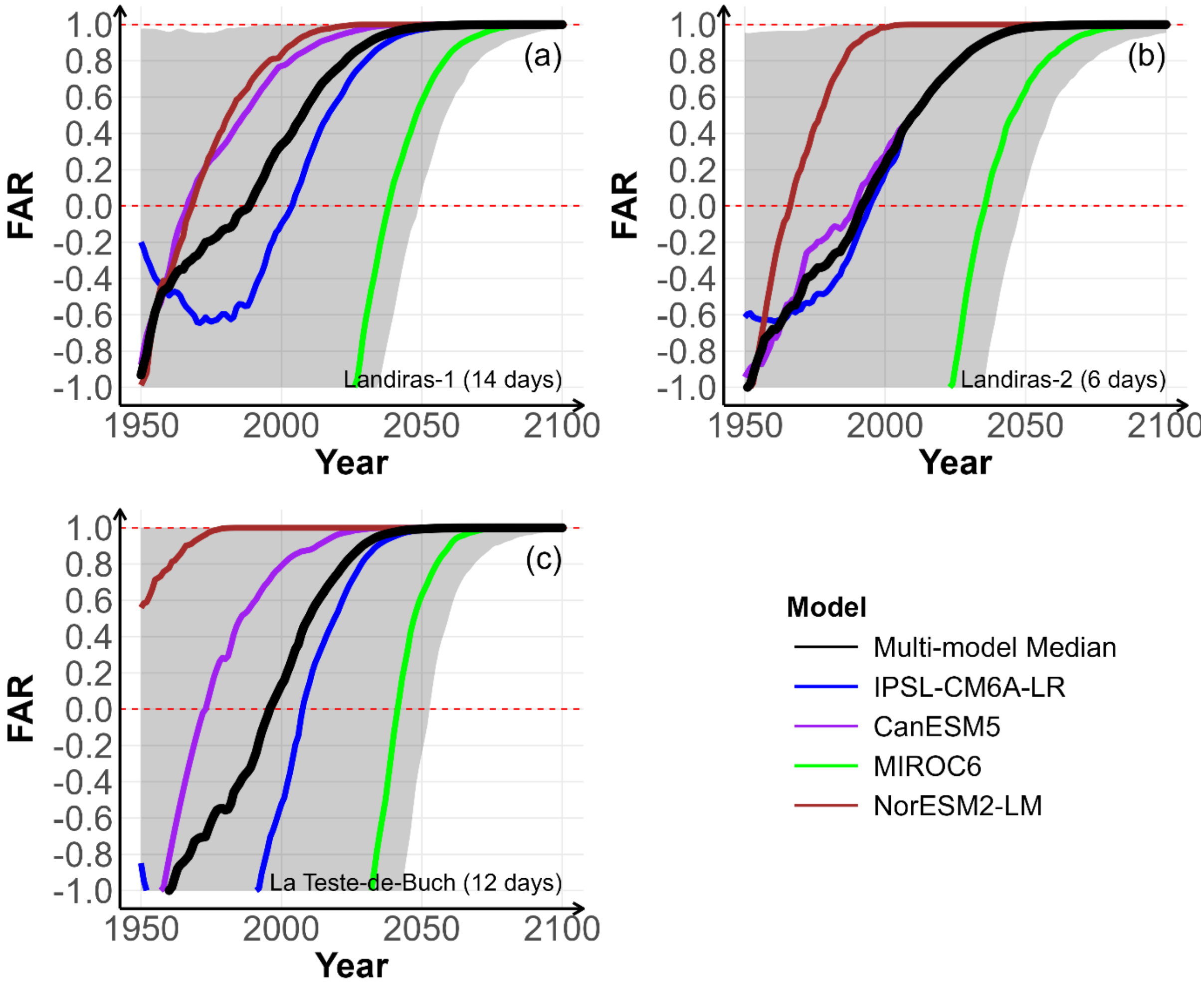

Figure 8. Fraction of attributable risk (FAR, 1 − 1/RR) using the local FWI over the event-duration window for (a) Landiras-1 (14-day), (b) Landiras-2 (6-day), and (c) La Teste-de-Buch (12-day). The black curve shows the multi-model median across models. The y-axis is truncated to [−1,1]; red dashed lines indicate FAR = 0 (no anthropogenic contribution) and FAR = 1 (fully attributable). The shaded envelope indicates the 90 % pooled ensemble uncertainty range for the multi-model median, obtained by pooling the bootstrap replicates from the four models and taking the 5th and 95th percentiles of the pooled distribution. This approach captures both within-model sampling uncertainty and inter-model spread.

As stated by the reviewer, the uncertainty (whether sampled at the model level or across models) is quite large and we believe that this deserves to be emphasized throughout the text. We mentioned it in the Results and Discussion sections.

Revised text at p. 16, L266-268: "Our study suggests that climate change increased the risk of such conditions by 2–10 times in 2022 and will continue to do so by several orders of magnitude by the end of the twenty-first century under a medium-level radiative forcing scenario. However, the pooled ensemble uncertainty range indicates substantial uncertainty across and within models."

Inserted at p. 16, L282: “However, we note that the associated pooled ensemble uncertainty remains substantial.”

It was not clear why this study was limited to only one realization from each of four CMIP6 models. This limited sample size for fitting extreme value distributions results in larger uncertainty. If calculating the FWI is limiting, there are datasets available where this computation has already been performed (e.g., [https://doi.org/10.5194/essd-15-2153-2023]).

Response: We agree that relying on a single realization per model limits our ability to characterize within-model (internal) variability, and we acknowledged this limitation more explicitly in the revised manuscript. Our choice was primarily driven by data availability and by the need to maintain a strictly comparable experimental design across models. For NorESM2-LM, the publicly accessible CMIP6 archive provides only one realization (r1i1p1f1) for the ssp245-nat experiment. To avoid mixing models with unequal ensemble sizes, we therefore used a single member for all four models; otherwise, RR/FAR curves from models with more available members would appear artificially smoother simply because they are based on more simulations. In addition, each realization represents a distinct and internally consistent climate trajectory associated with a specific set of initial conditions, and pooling members through an average or median would break this physical consistency. If multiple realizations mainly differ by initial conditions, simply averaging them would tend to smooth variability and extremes, rather than necessarily reduce the uncertainty of the GEV fit. We nevertheless acknowledge that this choice limits the characterization of within-model variability, and we clarified this point in the Data and Methods section.

Inserted at p. 6, L114: “For each model, we used a single member (r1i1p1f1) in both the ALL and NAT experiments. This choice was primarily motivated by data availability: for NorESM2-LM, only one realization was publicly available for the ssp245-nat experiment in the CMIP6 archive. For consistency and to avoid artificially smoother RR/FAR curves for models with more available members, we used a single member for all four models.”

Regarding the dataset suggested by the reviewer (Quilcaille et al., 2023), we agree that it is a valuable resource, but it does not include natural-forcing-only experiments (e.g., hist-nat or ssp245-nat), which are essential for our attribution analysis. In addition, it only provides annual-scale FWI indicators rather than the daily time series we need to estimate FWI levels across multiple temporal scales.

Related to the previous comment and considering the different climate sensitivities of the models, would using GSAT as a predictor in the nonstationary distribution instead of year allow for a more even comparison between models?

Response: We thank the reviewer for this insightful suggestion. Using GSAT as a covariate instead of year is indeed an appealing idea, because it places models with different climate sensitivities on a more common physical basis. Following this suggestion, we explored an alternative formulation of the nonstationary GAMLSS-GEV in which each model’s 30-year smoothed GSAT time series (see Fig. R1) was used as the covariate.

Figure R1:

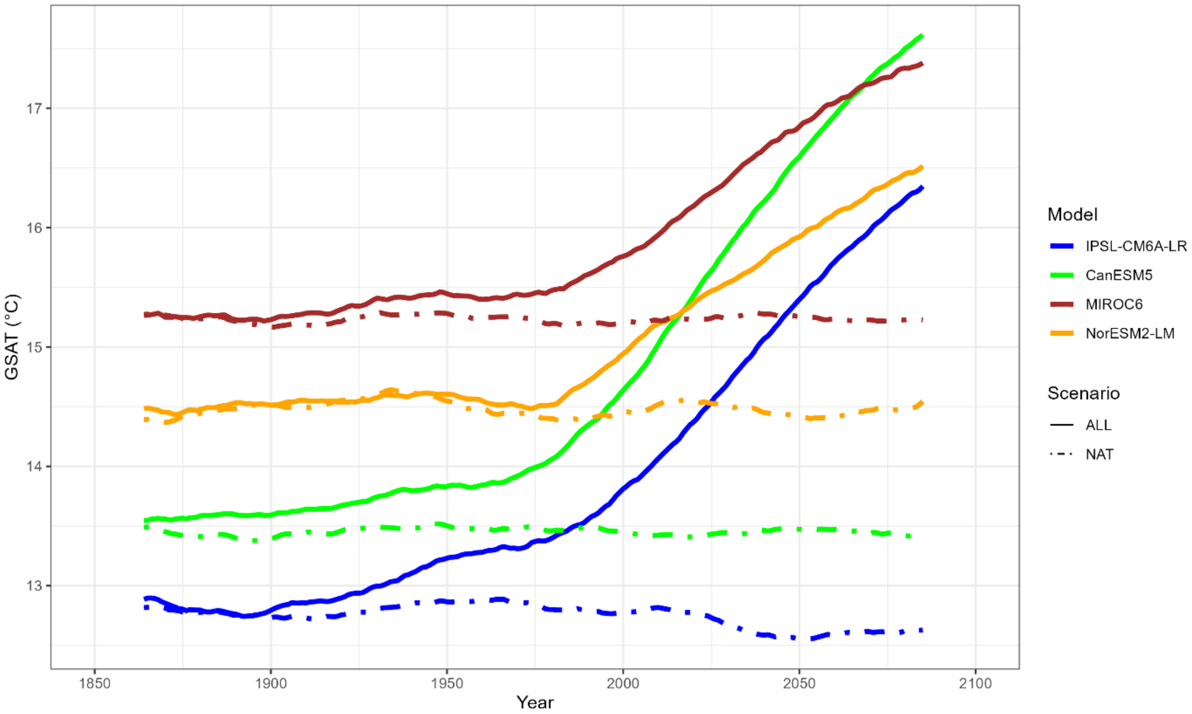

Figure R1: Global Surface Air Temperature (GSAT) across multiple CMIP6 models under the ALL (anthropogenic + natural forcing) and NAT (natural-only forcing) scenarios for the r1i1p1f1 simulation. The GSAT time series are smoothed using a 30-year moving average to highlight long-term trends only.

However, we found that RR (Fig. R2) and FAR (Fig. R3) estimates were substantially more variable from one year to another with several erratic spikes. This may arise from the asymmetry of the covariates (GSAT ALL vs GSAT NAT) whereas both ALL and NAT were fitted against the same covariate when using the calendar year. We thus prefer to stick with the year-based approach in the main paper.

Figure R2:

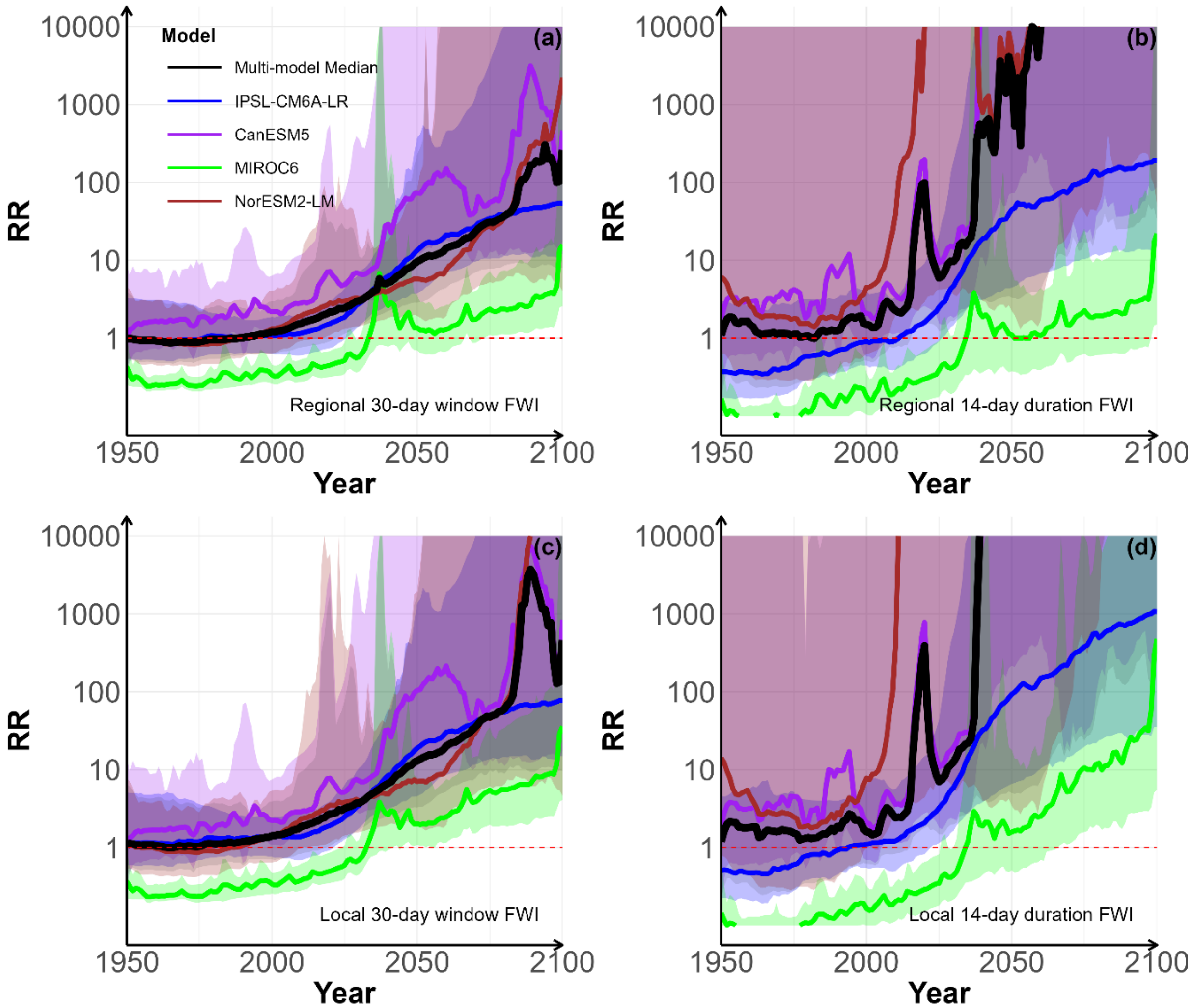

Figure R2. Same as Fig. 7, except that the risk ratio (RR) of FWI conditions associated with the Landiras-1 wildfire was estimated from GAMLSS-GEV models fitted to annual maxima MA-FWI using each model's global surface air temperature (GSAT), rather than year, as a covariate.

Figure R3:

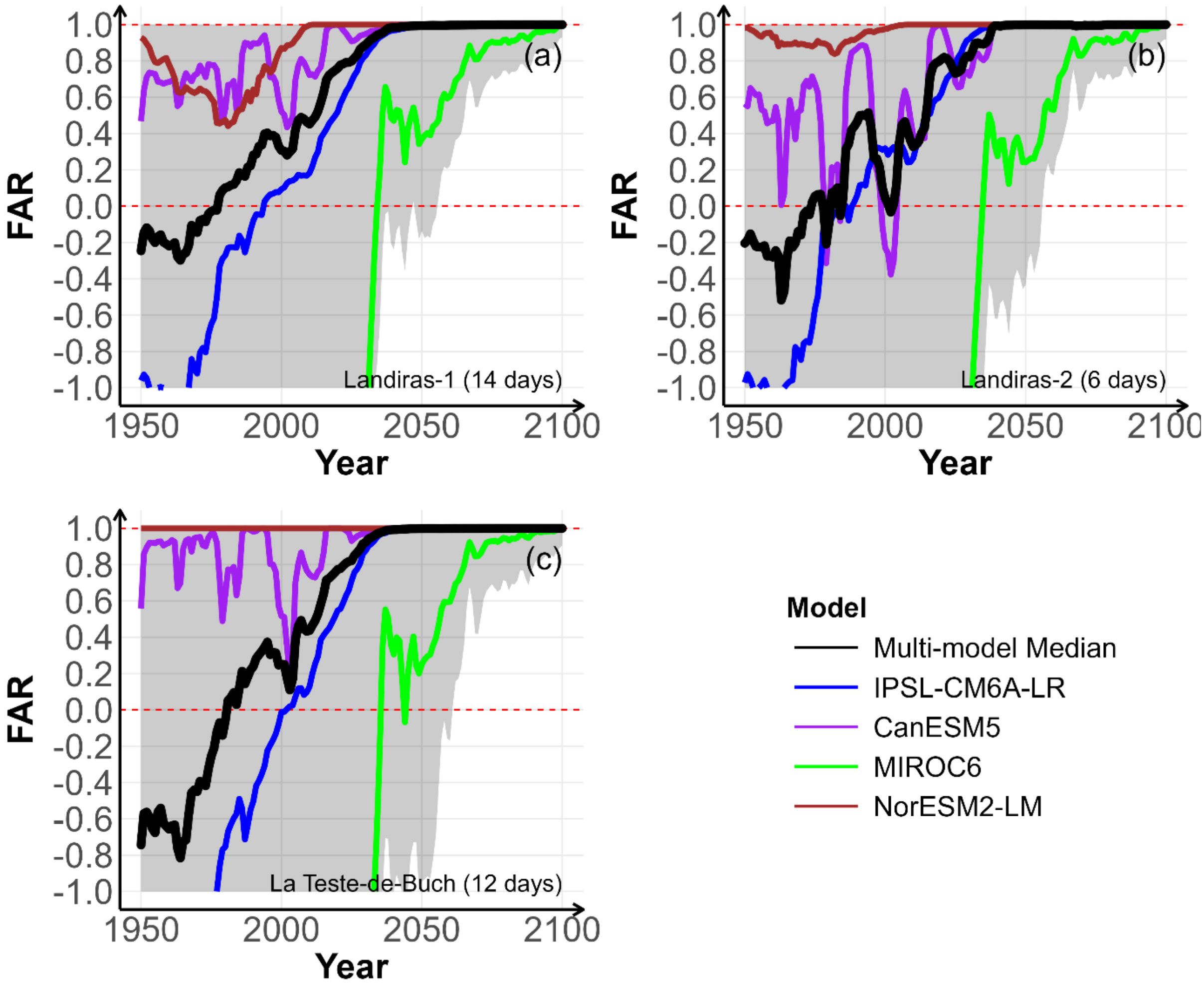

Figure R3. Same as Fig. 8 except that the fraction of attributable risk (FAR) was estimated from GAMLSS-GEV models fitted to annual maxima MA-FWI using each model's global mean surface air temperature (GSAT), rather than year, as a covariate.

For the Landiras fires, how do you reconcile conducting the analysis with the annual maximum when the maximum value could only come from one of these events?

Response: We appreciate the opportunity to clarify that the annual maxima MA-FWI for 2022 is used solely as the annual extreme input for fitting the GEV distribution, and is not intended to represent the fire weather conditions of any individual wildfire. In Figure 5, we indicated the fire weather level of each of the three wildfires (FWI aggregated over the burning duration) but those numbers were not used in the GEV fitting. They are therefore distinct from, and not equivalent to, the 2022 annual maxima MA-FWI used in the GEV fitting. We clarified in the Data and Methods section to avoid any ambiguity.

Revised text at p. 6, L133-134: "Note that the annual maxima of the MA-FWI time series were used only as annual extreme inputs for fitting the GEV distribution, and do not represent the fire-weather conditions of any individual wildfire. The FWI level observed during each wildfire is therefore distinct from the annual maxima used in the GEV fitting."

Specific comments

L30: Does "since the 1940s" mean that there was a larger fire in the 1940s or is this when the comparable records start?

Response: Indeed, this was ambiguous. Our intention was to refer to a specific fire: the Landes forest fire of August 1949 in the Gironde department, which burned approximately 50,000 hectares and remains the largest wildfire on record in France. To avoid this confusion, we have revised the text and added a reference.

Revised text: "When combined, the Landiras-1+2 wildfire burned over 19 776 ha, which makes it the largest wildfire in France since the Landes forest fire of August 1949 (Sarrau and Yagoub, 2025)."

Sarrau, J. and Yagoub, M. M.: Documentation of Historical Forest Fires and Hazard: Case of Gironde and Les Landes, France, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., X-G-2025, 771–778, https://doi.org/10.5194/isprs-annals-X-G-2025-771-2025, 2025.

L45: Should North America here be the western United States?

Response: Absolutely, the cited studies actually focused on the western United States.

Revised text: "However, fire weather conditions (combining multiple meteorological variables) have received less attention, although a number of efforts have been made in the western United States (Abatzoglou and Williams, 2016; Williams et al., 2019; Brown et al., 2023), Canada (Kirchmeier-Young et al., 2019a), Australia (van Oldenborgh et al., 2021), and France (Barbero et al., 2020; Lanet et al., 2024)."

L76: The reference for the Canadian FWI System should be Van Wagner (1987)

Response: Good catch. We corrected.

L93: What happens if a fire covers more than one grid cell?

Response: This is a good question. Because the BDIFF dataset provides only the city of the fire and not the exact fire perimeter, each fire was matched to the nearest SAFRAN grid cell. A wildfire may indeed cover multiple grid cells but this should not happen for the vast majority of fires, given the size of the SAFRAN grid cell (64 km^2 = 6,400 ha). Moreover, sampling additional neighboring grid cells is not expected to change much the results given the strong spatial autocorrelation of the FWI, including at daily scale. We clarified this in the Data and Methods section and discuss this limitation explicitly in the revised manuscript.

Revised text: "To quantify the relationship between local FWI conditions and fire events, we extracted for each wildfire the daily FWI time series from the nearest SAFRAN grid cell, over a window extending from 90 days before to 90 days after the wildfire start. Note that BDIFF does not provide fire perimeter and that multiple grid cells may potentially intersect with the actual fire perimeter. However, this effect should be limited given the size of SAFRAN grid cell (64 km^2) and the inherent spatial autocorrelation of FWI."

Figure 2: Should the "=" between FWI_ALL and FWI_NAT be an arrow instead?

Response: Good catch. We have modified Figure 2 accordingly.

Revised figure (Fig.2):

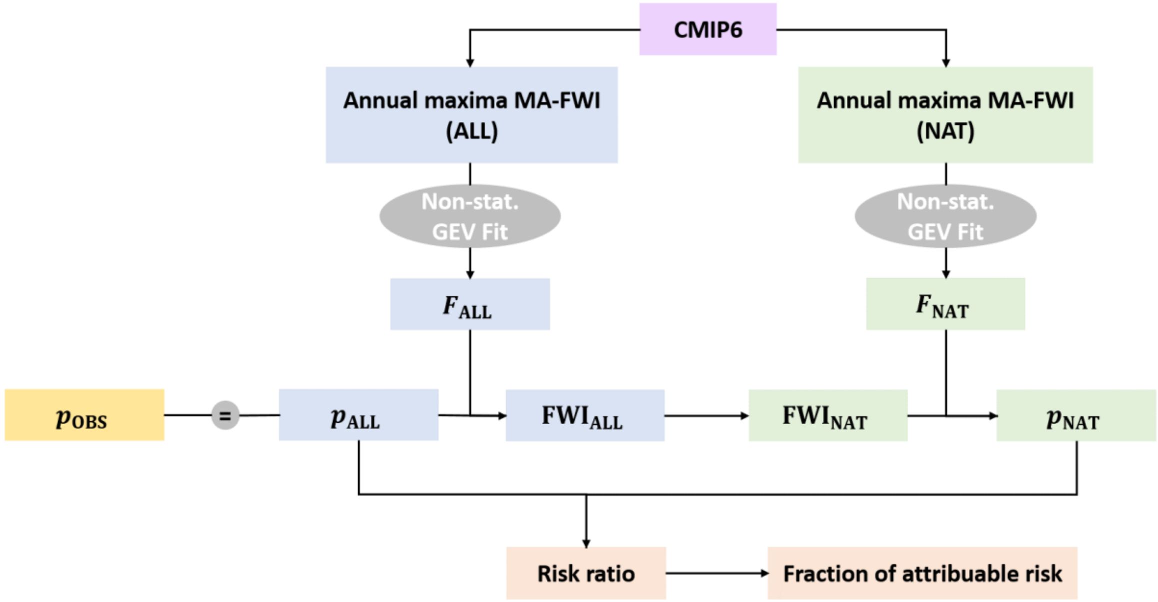

Figure 2. Schematic workflow used to estimate exceedance probabilities (denoted by p) under ALL and NAT forcings from CMIP6 annual maxima MA-FWI: a non-stationary GEV fit provides the cumulative distribution functions (CDFs) $F_{ALL}(x)$ and $F_{NAT}(x)$ (where F denotes the fitted GEV CDF); FWI_{ALL} is obtained by inverting F_{ALL} such that $1 - F_{ALL}(FWI_{ALL}) = p_{OBS}$, and p_{NAT} is then computed as $1 - F_{NAT}(FWI_{ALL})$.

L184: Could this be related to an early start to the fire season?

Response: Absolutely. The persistent positive FWI anomalies in the months preceding wildfire may partly reflect an earlier onset of the fire season, in addition to antecedent dry conditions. We clarified in the revised manuscript.

Inserted at p. 9, L186: "These persistent pre-wildfire positive anomalies may reflect not only prolonged antecedent hot and dry conditions, but also, to some extent, an earlier seasonal onset of the fire weather season."

Figure 4: Suggest making the green shading darker to increase visibility.

Revised figure (Fig. 4):

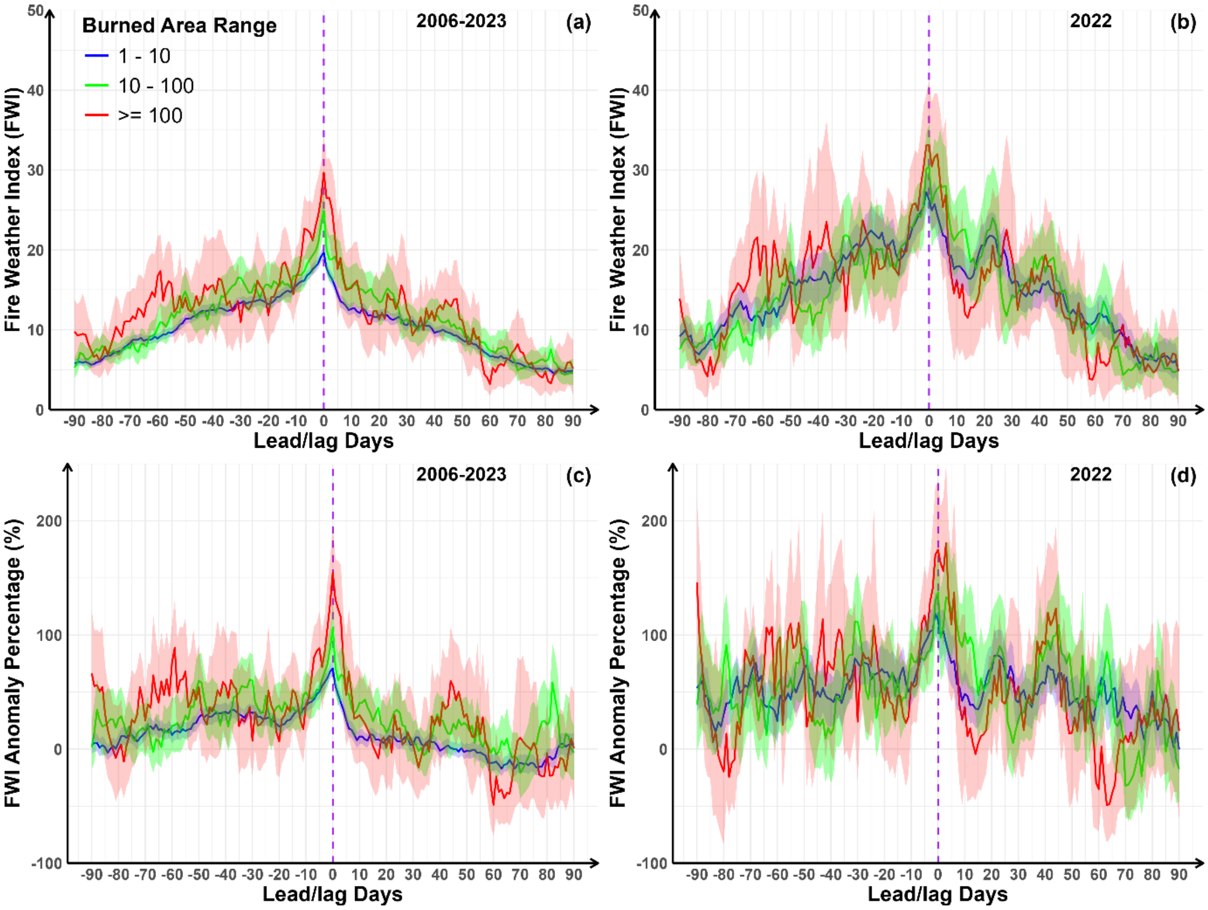

Figure 4. Lead–lag time series of FWI (a, b) and percent anomalies (c, d) relative to wildfire dates for three fire size classes over 2006–2023 (a, c) and 2022 only (b, d) in SW France. Anomalies were computed relative to the long-term (1959–2023) mean local seasonal cycle. Blue, green, and red curves denote BA = 1–10 ha, BA = 10–100 ha, and BA ≥ 100 ha, respectively. Shaded bands indicate 95% bootstrap confidence intervals. The x-axis shows lead/lag days from −90 to +90 relative to the wildfire starting day (day 0; purple dashed line).

L196: These are ranges of best estimate values from the different definitions, not uncertainty ranges. It would be helpful to clarify this in the text here.

Response: Agreed. We corrected to avoid any misunderstanding.

Revised text: "The estimated RP values of the observed annual maxima FWI associated with the three wildfires (Fig. 5, right panels) span the following ranges of best-estimate values across the four spatiotemporal scales: Landiras-1: 6–34 years;

Landiras-2: 17–89 years; La Teste-de-Buch: 6–101 years, illustrating how sensitive the RPs are to the chosen scales."

Figure 6: I found this figure and the accompanying text to be confusing. Would the observed probability (and subsequently p_{ALL}) not change through time as well? One suggestion would be to choose the FWI_{ALL} threshold based on the $p_{ALL} = p_{OBS}$ in 2022, but to then use the same FWI_{ALL} threshold for all other years. This might better reflect the probability of those fire conditions in different years.

Response: We tried to clarify. Figure 6 was designed to illustrate the logic underlying the RR computation in Figure 7, not to depict how the probability of exceeding a fixed 2022 FWI threshold evolves over time. In our framework, $p_{ALL} = p_{OBS}$ is used as a reference, and the figure illustrates how p_{NAT} varies relative to that reference, from which the attribution metrics are derived. So yes, the observed probability and subsequently p_{ALL} do not change with time. We revised the figure caption and accompanying text to avoid misunderstanding.

Revised caption (Fig. 6): Example of the NAT-only exceedance probability for the Landiras-1 wildfire (local and fire-duration set-up) using NorESM2-LM (r1i1p1f1): the NAT-only exceedance probability p_{NAT} (black; median) is shown relative to the reference probability $p_{OBS} = p_{ALL}$ (blue) that does not change with time. Shaded envelope indicates the 80 % parametric-bootstrap confidence interval.

Revised p. 12, L203-205: "Finally, we examined how ACC altered the probability of those FWI conditions. Figure 6 illustrates, for one model and one spatiotemporal set-up (NorESM2-LM, r1i1p1f1), how p_{NAT} varies relative to the reference probability p_{OBS} (= p_{ALL} by construction), from which RR and FAR were derived."

L242: "proportional" here implies a linear relationship. Was this intended or could a different word be chosen?

Response: We agree that it was a poor choice of words. Our analysis only shows that larger wildfires are associated with larger FWI anomalies, based on a composite analysis and not a formal linear fit. We therefore revised the wording.

Revised p. 9, L182-184: "Figure 4a shows that FWI increases until the wildfire day and decreases in the following days, with higher FWI values for larger wildfires."

Revised p. 15, L241-242: "The exceptionally high FWI values observed in 2022 in southwestern France, whether sampled locally or regionally, were conducive to a series of wildfires, with larger wildfires associated with larger FWI anomalies."

L296: The use of only a single ensemble member needs to be mentioned much earlier in the manuscript.

Response: We agree with this point, and we have already addressed it in the relevant part of the previous General Comments.