[Reviewer 1]

This protocol paper outlines the Fire Model Intercomparison Project (FireMIP) in the Coupled Model Intercomparison Project Phase 7 (CMIP7). It establishes a comprehensive experimental design, including specific simulations, required inputs and outputs, and a detailed analysis framework. The primary contribution is to provide a standardized, community-wide approach for evaluating fire simulations in Earth System Models, assessing past and future fire regime changes and their drivers, and quantifying the multifaceted impacts of fires on climate, ecosystems, and society within a fully coupled modeling framework. The protocol is clearly written and well structured. However, while the paper well describes the technical specifications and experimental design, the accompanying discussion of fire associated in-depth scientific analyses remains comparatively general and would benefit from a stronger theoretical framework or clearer metrics. Given the critical importance of fire in the earth system and the global need to reduce model uncertainties, this FireMIP paper is a substantial contribution that merits careful revision before publication.

Reply: We thank the reviewer for the positive assessment and constructive comments. We have revised the manuscript accordingly. In particular, Section 5 (Recommended analyses) has been revised to improve the framework and organization as well as to provide clearer evaluation metrics. A point-by-point response is provided below.

Major Remarks:

The "recommended analyses" section requires further strengthening. Currently, its three subsections align well with the three scientific questions outlined in the paper: (1) How do models perform? (2) How do fire regimes change? (3) What is the impact? Yet the subsection titles remain overly general, limiting their ability to reflect the specific features of a FireMIP. To enhance the clarity, I recommend introducing structured subsections that explicitly reflect the unique scope and design of FireMIP. For instance, under the first question (How do models perform?), subsections could include: "Regional evaluation (e.g., permafrost fires, Amazon fires, and African fires)", "Evaluation of burned area, fire frequency, and carbon emissions", "Variability and sensitivity analysis". Such structuring would better articulate the key dimensions of fire model assessments, ensuring the recommendations are tailored to the project's objectives and actionable for future implementations.

Reply: Thanks for the helpful suggestion. We agree that Sec. 5 benefits from clearer, more FireMIP-specific structure. We kept the top-level subsection titles (5.1 Fire simulation evaluation; 5.2 Fire changes and drivers; 5.3 Fire impacts) to align with the experiment-group names and scientific questions, but have strengthened each subsection by (1) adding structured run-in headings that reflect FireMIP-specific dimensions, (2) clarifying key metrics and definitions, and (3) adding introductory text to clarify the framework and scope.

The revised Sections 5.1–5.3 are now structured as follows:

- 5.1 Fire simulation evaluation
  - Target variables and experiments
  - Evaluation metrics (totals, spatial pattern, seasonality, interannual/decadal

variation, extreme fire frequency, links to local factors (climate and socioeconomic), links to teleconnections, and sensitivity tests)

Evaluation focus (CMIP6 limitations)

Key regions (Africa, Amazon; Arctic-boreal zone including permafrost fires)

Benchmarks (present-day, long-term, regional)

5.2 Fire changes and drivers

Fire changes

Uncertainties in simulated fire changes

Fire drivers and causal pathways

5.3 Fire impacts

Impacts of fire and fire aerosols (historical)

Impacts of changes in fire and fire aerosols (future)

Underlying mechanisms

Benefits of the coupled CMIP7 framework

Minor Comments:

(1) Line 116-129: Both "scientific question 1" and "critical issue addressed by Li et al., (2024)" use "(1)". Consider using different notations to avoid confusion.

Reply: Thanks for pointing out this. (1), (2), and (3) for the critical issues addressed by Li et al. (2024) have changed to (i), (ii), and (iii).

(2) Line 130: The key point of the second scientific question is unclear, as multiple aspects are combined without clear distinction. While Figure 1 provides a good summary of Q2, the corresponding text description (Line 132-136) lacks clarity.

Reply: According to your suggestion, we have revised the description of Q2 to "This question focuses on two linked aspects: (i) characterizing fire changes in the past, present, and future, and (ii) attributing these changes. Regarding fire changes, CMIP6-simulated fire changes still have large uncertainties, and CMIP7 offers an opportunity to revisit them with improved models and scenarios. The uncertainties in fire change simulations can be quantified and separated into contributions from model uncertainty, internal climate variability (initial-condition uncertainty), and scenario uncertainty. Regarding attribution, earlier studies focused either on local, direct fire drivers (e.g., local weather and climate) or on attributing climate changes to anthropogenic versus natural forcings. CMIP7 FireMIP integrates the two by analyzing how anthropogenic and natural forcings shape global fire changes by altering local, direct fire drivers.".

(3) Line 163-165: How to generate the initial condition for offline land model is unclear.

Reply: Lines 163-165 refer to the coupled ESM simulations rather than offline land-model simulations. We have added "For coupled simulations," to make this explicit.

For Groups 1–2 (CMIP7 DECK and FastTrack experiments), initial-condition ensembles follow the corresponding experimental design (e.g., branching from different times in the piControl run, or applying small perturbations to an atmospheric state variable such as temperature). Details of the initial conditions for Group 3 (FireMIP-specific) experiments are provided later in the dedicated Group 3 description where Group 3 experiments are introduced.

(4) Line 175: Should the "Fire impacts" group 3 in Table 1?

Reply: Yes, we have corrected the typo.

(5) Line 183-187: The description of 2.3 and 2.4 are different with Table 1. What is "overshoot"? Could be more specific than Table 1.

Reply: We have revised the descriptions of 2.3 and 2.4 to match Table 1.

In addition, in descriptions of vllo and vlho, we have added "Here, "overshoot" denotes the magnitude by which global warming temporarily exceeds the 1.5°C target level during the 21st century (Van Vuuren et al., 2025)."

(6) Line 190: Should "failed-policy futures" be "future failed-policy"?

Reply: "Current-policy" and "failed-policy" are adjectival modifiers of "futures", so "current-policy and failed-policy futures" is correct.

(7) Line 221-223: By comparing different experiment results, we are able to answer different questions. To summarize these potential comparisons in Figures or Tables would be helpful to enhance the clarity.

Reply: Thanks for the suggestion. We have added a new table (Table 2) to summarize the potential experiment comparisons and their associated scientific purposes.

(8) Line 226-234: How these inputs could be used in any example models could be demonstrated in Figures.

Reply: We have added a new table (Table 3) illustrating how each fire-specific input is used in ESMs and when it is required.

(9) Line 239-240: Not sure how common these (2-5) inputs are used by different models. If very common, standardized inputs will help reduce uncertainties.

Reply: Around half of the ESMs participating in CMIP7 FireMIP may use inputs (2-5). We have revised the manuscript to note that using standardized, consistent datasets (e.g., the CESM3 input set) is recommended to reduce uncertainty.

(10) Line 243 &249: "driver" should be "drivers". Not clear what does "impact variables" exactly mean.

Reply: The "driver" and "impact" in "fire driver and impact variables" are used as attributive nouns modifying "variables". We have changed it to "variables of fire drivers and fire impacts" to avoid confusing.

(11) Line 272-273: This sentence is unclear for me. Do you mean "Variables required by CMIP7 FireMIP are all listed in CMIP7 DECK and AFS experiments."? I feel confused about "no additional requests specifically for FireMIP". Aren't burned area fraction and fire carbon emission specifically for FireMIP? If these two variables are already included in CMIP7 list, I don't think it is necessary to mention

this sentence.

Reply: The two variables are already included in CMIP7 list, so we have removed the sentence as you suggested.

(12) Line 297-304: Showing the data uncertainties of fire counts and carbon emissions in Figures would be helpful to learn the quality of these benchmarks.

Reply: Thank you for the suggestion. The large uncertainty across fire benchmark products is well reflected by the spread in global totals of burned area and fire emissions listed in Table 5 (Table 7 in the revised version), which we believe shows the inter-product differences more clearly than figures. A quality assessment of benchmarks would require dedicated analysis which is beyond the primary objectives of this CMIP7 FireMIP protocol paper, and the fire observational community has not yet reached a consensus on quality ranking of these global products.

(13) Line 284: Figures could be used to help demonstrate potential analysis.

Reply: Potential analyses are demonstrated in Figure 1 as well as new Table 2.

(14) Line 285-321: Only data sources were introduced in this section. What kind of metrics can be used or what aspects can be focused on should be also introduced?

Reply: We have added dedicated headings in Sec. 5.1 to clearly distinguish evaluation metrics, and expanded the text to specify the key metrics/aspects that should be focused on. In Fig. 1, we have revised the label "characteristics" to "metrics" to align with the revised text made in Sec. 5.1.

(15) Line 322: I prefer to a clearer title "Fire-related changes and corresponding drivers".

Reply: We agree that "Fire-related changes and corresponding drivers" may be clearer, but we prefer to keep the shorter sub-section title "Fire changes and drivers" for conciseness and consistency with the terminology used in Figure and Tables throughout the manuscript.

(16) Line 329-339: Categorizing drivers of fire-related changes would strengthen the clarity of the potential analysis.

Reply: Agree. Categorized.

(17) Line 341-353: Categorizing the impact would enhance the clarity of the potential analysis.

Reply: Agree. Categorized.

[Reviewer 2]

General Comments:

This manuscript presents the well-designed protocol for integrating the Fire Model Intercomparison Project (FireMIP) into CMIP7, establishing a crucial framework for advancing global fire modeling within fully coupled Earth System Models. It clearly outlines the key scientific questions and a logical multi-experiment structure for evaluating fires, attributing changes, and quantifying impacts. The study effectively integrates fire processes into the broader Earth system perspective and demonstrates a strong awareness of observational uncertainties. However, providing further justification for key methodological choices and clarifying how consistency will be maintained across diverse modeling approaches would strengthen the protocol. Overall, this work constitutes an essential and timely reference that will significantly advance coordinated fire modeling and its integration in climate projections. The paper is worthy of publication after minor revision.

Reply: We sincerely appreciate your positive assessment and helpful comments. The point-by-point responses are provided below.

Specific Comments:

(1) Title Suggestion: The current title reads as a results report. To better reflect the paper's nature as a design and protocol paper, consider revising it to: "Conceptions and Scientific Design of the Fire Modeling Intercomparison Project (FireMIP) for CMIP7".

Reply: Thanks for this suggestion. The manuscript is submitted to the "CMIP7 scientific objectives, experimental design, and organization" special issue and is a protocol paper describing ongoing and future activities. We therefore prefer to retain the concise title "The Fire Modeling Intercomparison Project (FireMIP) for CMIP7", which is consistent with the naming convention of protocol papers of other CMIP7 MIPs in this special issue (…MIP for/contribution to/towards CMIP7). We have revised the Abstract to explicitly indicate that this is a protocol paper by replacing "This paper …" with "This protocol paper …".

(2) Explanation for Diverging Fire Indicators (P2, Lines 31-35 & P3, Lines 39-42): The text notes that while global burned area has declined in recent decades, extreme fire frequency and forest fire emissions have increased. This apparent contradiction is not explained. A brief, accessible mechanistic explanation should be added (e.g., in the Introduction or a dedicated paragraph).

Reply: After the sentence, we have added the mechanistic explanation: "The reduction in global total burned area is largely due to decreased burning in tropical savannas and grasslands, which is mainly driven by human fire suppression (e.g., agricultural expansion increases landscape fragmentation and thus reduce fire spread, fire management) (Andela et al., 2017). By contrast, increases in extreme-fire occurrence and forest-fire emissions are mainly linked to climate change and variability that enhance fuel drying and, in some regions, to fuel accumulation resulting from long-term fire suppression (Zheng et al., 2021; Cunningham et al., 2025).".

(3) P5, L123-L25: The manuscript states that FireMIP in CMIP7 aims to improve upon CMIP6; a more explicit and structured comparison is needed. It should be clarified which specific deficiencies identified in CMIP6 fire simulations are explicitly addressed, and whether there are concrete examples where new experiment designs or outputs directly respond to past limitations.

Reply: Before this sentence, we summarized key deficiencies identified in CMIP6 fire simulations: "CMIP6 ESMs still underestimate the recent decline in global total burned area, fail to capture the spring fire peak in the Northern mid-latitudes, and perform poorly in the Arctic–boreal zone.". In CMIP7 FireMIP, we will use outputs (e.g., burntFractionAll and fFire) from Experiment 1.1 (historical/esm-historical) to conduct a comprehensive evaluation analogous to that performed for CMIP6 (Li et al., 2024). To assess whether CMIP7 addresses the deficiencies identified in CMIP6 fire simulations, we will place particular emphasis on (i) recent trends in burned area, (ii) fire seasonality in the Northern midlatitudes, and (iii) the spatial and temporal variability of burned area and fire carbon emissions in the Arctic–boreal zone. These are now explicitly stated in Section 5 (Recommended analyses).

(4) The manuscript strongly emphasizes evaluation against observations, but the link between evaluation metrics and process-level understanding is sometimes implicit. It needs to be explained how the proposed analyses help diagnose why models differ, rather than merely how much they differ.

Reply: We revised the text to make explicit how the evaluation is used for process-level diagnosis, not only to quantify model-observation differences. Specifically, we interpret standard fire outcome metrics (totals, spatial patterns, trends, seasonality, variability, and extremes) together with driver-response diagnostics that relate fire to climate, fuels/biomass, and socioeconomic factors, which helps identify the dominant controls behind inter-model spread and likely sources of biases. We also note that attributing biases to specific processes requires (i) knowing key differences in fire parameterizations across ESMs, and (ii) complementing the coupled-model evaluation with offline land-model simulations (e.g., FireMIP experiments outside CMIP and/or land-hist in CMIP7) which are forced by the same prescribed atmospheric reanalysis and CO_2 concentrations.

(5) Rationale for Experimental Priority Tiers (P8, Table 1 & P9, Lines 190-200):

Reply: We have added the Rationale in Sec. 3 (Experimental design) as "The experimental priority tiers in CMIP7 FireMIP follow the CMIP7/CMIP data-request convention. Priority 1 (required) is the minimum set needed to meet the FireMIP objectives (fire evaluation; fire changes and drivers; fire impacts) and to maximize broad community utility. Priority 2 (recommended) includes important experiments that provide a more complete and robust assessment (e.g., improved attribution by separating anthropogenic forcings; impact analyses of fire aerosols, fire changes, and fire aerosol changes), but may not be feasible for all modeling centers due to computational cost or other constraints. Priority 3 (optional) includes additional experiments that further support the FireMIP objectives and are of interest to the community, which are undertaken as resources allow."

(6) The manuscript labels some experiments (e.g., 2.3, 2.4) as lower priority. The scientific rationale for this prioritization is not fully explained. Please add a brief justification, such as their relative contribution to the core scientific questions or resource considerations, to guide modeling groups.

Reply: We have added a brief justification "Experiments (2.3) (hist-aer, hist-GHG, hist-lu) can provide a more detailed attribution of historical fire changes to specific anthropogenic drivers (aerosols, greenhouse gases, and land use), thereby strengthening the assessment of Scientific Question 2 (Fire changes and drivers). However, the core separation of natural versus anthropogenic influences can be addressed using (2.1) hist-nat relative to the historical experiment, so (2.3) is lower priority. Similarly, experiments (2.4) (scen7-vlloc/esm-scen7-vllo and scen7-vlhoc/esm-scen7-vlho) are designed to add overshoot-specific insight, but the core assessment of future fire changes across alternative scenarios and associated drivers can be addressed with (2.2) (scen7-h/m/ml/l), and (2.4) is therefore also assigned a lower priority.".

(7) P9-P10, Group 3 experiments: For the hist-no-fire experiment, models with prescribed emissions and models with interactive fire modules implement "fire suppression" through different technical methods. It should be addressed whether the climate and carbon cycle responses to these two methods are comparable, and how this technical heterogeneity should be accounted for in the multi-model comparison of fire impacts.

Reply: The hist-no-fire experiment is an idealized "no-fire world" historical simulation used to quantify fire impacts. For comparability, all models run 1850–2021 no-fire simulations initialized from an 1850 no-fire spin-up, regardless of whether fire emissions are prescribed or interactively simulated. In ESMs with prescribed fire emissions, burned area in the code and prescribed fire-emission fluxes in the forcing dataset are set to zero; in models with interactive fire modules, burned area is set to zero and fire emissions are consequently diagnosed as zero. We now request that modeling groups document whether they use prescribed or interactive fire emissions in Sec. 3 (Experimental design). We also recommend in Sec. 5 (Recommended analyses) stratifying results by configuration to assess the sensitivity of inferred fire impacts; consistent results would increase confidence, whereas discrepancies would motivate further investigation.

(8) P10, L230-L234: The protocol notes that models differ in their inclusion of explicit human fire suppression. It should be elaborated how FireMIP ensures meaningful intercomparison given this heterogeneity.

Reply: Thank you for raising this point. Whether and how models represent human fire suppression will affect simulations of global and regional totals and trends. Such differences among models are treated as a source of model uncertainty within the FireMIP framework, and used to understand bias and uncertainty sources when evaluating fire-related simulations, rather than as a barrier to intercomparison.

(9) P16, L289: Given the differences in spatial resolution and parameterizations among models, the manuscript should clarify how "extreme fire" is consistently defined across CMIP7 ESMs and discuss the sensitivity of inter-model comparisons to different definitions of extreme fires.

Reply: Thank you for this suggestion. We have revised the manuscript to clarify that the definition of "extreme fire" should be applied consistently across models for inter-model comparison, and we now explicitly recommend sensitivity tests to assess how inter-model conclusions depend on the chosen definition. We also added two example, percentile-based definitions that can be applied consistently across CMIP7 ESMs: (1) pool all values of a fire metric (e.g., annual, fire-season, or peak-month burned area fraction) within a given region/biome over a common reference period; define extremes as the top X% of this pooled distribution; then compute, at each grid cell, the fraction of years/seasons/months exceeding this threshold; (2) apply percentile thresholds at the grid-cell level, but only for grid cells that exceed a minimum annual burned-area-fraction threshold, to avoid defining extremes in locations with negligible burned area.

These choices of extreme-fire definitions and sensitivity tests are part of the analysis stage (after the protocol paper) and do not affect how the simulations are conducted. As stated in the revised opening of the Recommended Analyses section, this protocol provides a minimal set of recommended analyses and general guidance on how to conduct them. The specific implementation details and any additional analyses will be determined in subsequent community analyses based on the openly available CMIP7 FireMIP outputs.

To reduce the influence of spatial resolution, we have also suggested regridding both model outputs and benchmarks to a common grid (e.g., 1°) before evaluation in Sec. 5. Furthermore, after applying a common extreme-fire definition (and a common analysis grid), inter-model spread caused by parameterizations could be quantified as model uncertainty, and comparison with benchmarks can also help diagnose which parameterization performs better.

(10) P17, L297-L298: The manuscript correctly notes large uncertainties among burned area and emission products, but this issue deserves deeper discussion. Guidance is needed on how discrepancies among different benchmark products should be handled in evaluation, and whether there is a recommended strategy to avoid over-interpreting model–data mismatches.

Reply: We have added "To reduce the influence of large uncertainties in fire observations, we suggest the following evaluation strategy. For global and regional totals, a simulation is considered reasonable if it falls within the spread across available benchmarks (e.g., the full benchmark range or the multi-benchmark mean ± 1 standard deviation). For temporal variability (seasonal to decadal) and long-term trends, we recommend evaluating variabilities or changes normalized by the corresponding global or regional mean/total. For spatial patterns, we suggest spatial (pattern) correlation between simulations and each benchmark; simulations are considered skillful if the correlation is statistically significant ($P<0.05$) for at least one benchmark, with smaller P and higher correlation with more benchmarks indicating better performance."

(11) P18, Lines 330-340: Section 5.2 discusses soil moisture but does not explicitly mention precipitation, which is a primary driver of soil moisture and fuel dryness. Can the analysis framework list precipitation as a key variable for diagnosing fire regime drivers and changes? If can't, why?

Reply: We have added precipitation as a driver of fuel wetness.

Minor Comments

(1) P8, the header row in Table 1 reads "Group 2. Fire Impacts", but it should be corrected to "Group 3. Fire Impacts".

Reply: Corrected.

(2) Could the authors provide a clearer operational definition for the frequently used term "fire activity" and "fire regime"?

Reply: We have removed the term "fire activity" and replaced it throughout the manuscript with explicit fire-related variables (e.g., burned area, fire emissions).

We have also added the definition of "fire regime" as the statistical characteristics of fire variables such as burned-area seasonality, spatial patterns, totals, and trends.

(3) P8, Table 1: Could the authors clarify the criteria used to determine the priority for the reader?

Reply: Clarified. Please see the reply to your Specific Comments 5 and 6.

(4) P11, L243: Does "fire carbon emissions" refer exclusively to CO_2, and could the authors briefly clarify whether additional fire-related variables are considered?

Reply: No, "fire carbon emissions" refers to the total carbon emitted by fires across carbon-containing species (CO_2, CO, CH_4, NMVOCs, carbonaceous aerosols), rather than CO_2 alone. We have clarified it.

In CMIP, burned area fraction and fire carbon emissions are default fire outputs that are most consistently available across ESMs and thus are required variables. Additional fire-related variables (e.g., fire counts, fire intensity, fire size, fire speed, fire duration, land-cover-type–specific fire carbon emissions, and speciated trace-gas and aerosol emissions) are encouraged where available. We have added this in the revised manuscript.

(5) P12, L272: The term "AFS" appears without a prior definition. The authors are encouraged to define this acronym when it is first introduced.

Reply: "AFS" is a typo. We have corrected it to "AFT", which is defined earlier in the manuscript.

(6) Selection Criteria for Input Variables (P10, Lines 235-240): A list of fire-specific

inputs is provided, but the criteria for selecting these particular variables are not stated. Please add a sentence or two explaining the basis for their selection.

Reply: We have clarified the six variables listed in Paragraph 1 represent the maximum set of fire-specific inputs that ESMs may use. Their roles, associated fire processes, and whether a given variable is required by an ESM are listed in the new Table 3.

(7) Suggestion for Additional Observable Outputs (P11, Line 245): If satellite-observable variables like fire count and fire radiative power (FRP) could be included as model outputs. Adding these (if feasible for models) would greatly facilitate direct comparison with remote sensing products and should be considered.

Reply: Added. Note that no ESM output FRP, we therefore include fire intensity as the closest available proxy.