This manuscript presents an interesting ensemble-based framework for characterizing extreme wildfire events using a process-based global fire model forced by large climate ensembles. The methodological setup is promising and timely. However, the current manuscript remains largely descriptive, and several methodological assumptions, statistical choices, and interpretations require stronger justification before the conclusions can meet the standard of a top-tier journal.

Major comments:

1. The central conclusion that a 1440-year pooled ensemble captures more extreme wildfire events than a 36-year single realization is largely expected from sampling theory. While important, this result alone offers limited novelty. The manuscript would be stronger if it provided deeper physical insight into what specifically drives the most extreme wildfire events.
2. The study treats 40 ensemble members as a single pooled 1440-year sample. This assumes the ensemble members are statistically exchangeable and sufficiently independent. The authors should better justify this assumption and discuss possible implications for extreme-value estimation.
3. Return periods are inferred directly from empirical percentiles (e.g., p99 as a 1-in-100-year event). This approach is somewhat simplistic for rare-event analysis. A formal extreme-value framework such as GEV or peaks-over-threshold would provide more robust tail characterization.
4. The manuscript compares model output with GFED mainly using mean variability and correlations, but the study focuses specifically on extremes. Model skill in reproducing upper-tail burned area and extreme fire-emission events should be more explicitly evaluated.
5. Using a VPD-based fire danger index as the primary climate driver may be overly simplistic. Wildfire extremes are often controlled by compound hot, dry, and windy conditions, along with antecedent fuel moisture. The manuscript should better justify this choice and discuss its limitations.
6. Lightning and human ignition assumptions are largely fixed across ensemble members. This may suppress an important source of climate-driven wildfire variability, since ignition probability itself can vary substantially with atmospheric conditions.
7. The conditional metric (Δcond) based on minima of conditional and unconditional distributions is not intuitive and lacks standard statistical interpretation. More conventional dependence metrics could improve clarity and robustness.
8. The authors attribute fire danger–impact decoupling to fuels, ignitions, and vegetation feedbacks, but these mechanisms are not quantitatively decomposed. As a result, the mechanistic interpretation remains largely speculative.

9. The analysis uses only one climate model and one fire model. Therefore, the study assesses internal variability but not structural uncertainty, which may be equally or more important for wildfire projections.
10. Regions such as Europe and Boreal Asia show strong under sampling of extremes, but the manuscript provides limited physical explanation for why these regions behave differently. Stronger regional interpretation would improve scientific value.
11. All variables are linearly detrended prior to analysis. This may remove physically meaningful climate signals relevant to wildfire extremes. The rationale for detrending should be better justified.
12. The manuscript frequently refers to extreme wildfire events, but a precise operational definition is lacking. It should be clearly stated whether extremes are defined by annual maximum burned area, emissions, percentile thresholds, or another metric.

Minor comments

1. Figure 4 caption appears incorrect (Same as Figure 4) and should be corrected.
2. Figures 3–4 are visually dense and difficult to interpret at publication scale.
3. Several equations and symbols are inconsistently formatted.
4. Language editing would improve clarity and readability.