(Note: The reviewers' comments are in gray and the author's responses are in blue. Unless specified otherwise, the line numbers quoted in our responses are with reference to the revised manuscript.)

Dear Editor,

We thank both reviewers for their positive feedback on our manuscript. A detailed response to their individual comments is provided below:

Line 7-8: Is this relationship based on observation?

We assume that this comment was regarding the correlation between annual time series of both fire variables at the ecoregion level. The relevant sentence in our abstract has now been amended to clarify that the correlation is between the **modeled** time series and **observations**.

L7-8: *Moreover, the modeled annual time series of both fire variables exhibit strong correlations ($r \geq 0.6$) with observations in 16 out of 18 ecoregions.*

Figure 1: What test method and any threshold are used to define statically significant?

We used the Student's t-test to determine statistically significant trends (i.e rejecting the null hypothesis that there is no trend) with the p-value threshold, $p < 0.05$. We have now mentioned the threshold clearly in the Figure 1 caption as well as included the following line in the main text,

L111-112: *We also indicate all statistically significant ($p < 0.05$) trends, which were determined using the Student's t-test.*

Line 158: Suggest change "urban" to "urban fraction" if the fraction is used.

We have changed the predictor name from "Urban" to "Urban fraction" (L159)

Line 159: Suggest change to "..., Pop10_dist defined as ..."

We have improved the writing around population predictors (L161-162) as suggested by the reviewer. Now, the Pop10_dist line reads: "*distance from the nearest area with population density greater than 10 people per square kilometer (Pop10_dist), …*"

Line 190: I disagree with the notion that neural network models are always better than gradient boosting models or other ML models, particularly when it comes to inferring out-of-sample climates such as future climate states or different fire regimes as mentioned by the authors. The limitations of NN for future inference were previously discussed on [1,2]. I recommend to modify the sentence.

We have now modified the sentence in L190 to reflect the reviewer's comment about highlighting the limitations of ML models.

L190-193: *Recent work (Levin et al., 2022) has also shown that neural network models are more powerful at learning feature representations than gradient-boosted trees. However, generalizing the learned relationships between input predictors and fires to out-of-sample data from future climate states or different fire regimes remains a challenging problem for most ML approaches, including neural network based models (Rasp et al., 2018; Yuval and O'Gorman, 2020).*

Figure 4, 5, 8-10: it is hard to compare lines in the monthly scale evaluation. It would be nice to adjust line thickness or the plot design to facilitate comparison.

To facilitate the comparison between observations and modeled fire frequency and burned area at monthly timescales, we have now reproduced Figs. 3-5 and 8-10 with two changes:

i) higher line thickness for observed frequency and burned area,

ii) higher color transparency for the modeled uncertainty.

We hope the new plots allow the results from SMLFire1.0 to be compared more easily with observations.

We would like to thank all the reviewers for their detailed comments as well as critical feedback on our manuscript, and we look forward to a favorable response.

On behalf of the authors,

Jatan Buch