We sincerely thank Dr. Kasoar for the valuable and insightful feedback on our manuscript. The comments provided have been instrumental in enhancing the quality and clarity of our work, serving as a critical resource for refining our methodology, data interpretation, and overall presentation. We have carefully considered each comment and have conducted a thorough revision of our manuscript to comprehensively address all concerns raised.

This response document details the specific changes made in response to each reviewer comment. For ease of reference, reviewer comments are presented in black, our responses are highlighted in blue, and modifications to the manuscript are indicated in italic font. The line numbers correspond to the clean version of the revised manuscript.

The authors present a new two-way coupling of an XGBoost machine learning fire model over the contiguous US, with the ELM land surface model, a derivative of the widely-used CLM land-surface and dynamic vegetation model, which can be run as an alternative to the process-based Li et al. scheme currently used within ELM (and CLM). The XGB fire model performs very well at reproducing the observation-based training dataset (GFED5 burned area) over the CONUS. The authors also compare against BA simulations from several process-based models, and note that agreement is better over regions of the CONUS where burned area is mainly driven by climate, and poorer over regions where there is a strong human influence on the burned area signature, e.g. due to crop and pasture fires, thereby highlighting another potential use of ML models to help identify key process that should be better represented in their process-based counterparts.

ML methods show a lot of promise when it comes to more accurately parameterising sub-grid phenomena, including wildfire prediction which is notoriously uncertain among current process-based and simple empirical models, and so I really welcome initiatives like this to interactively couple data-driven fire models with a dynamic vegetation model to provide an alternative to the existing process-based scheme, depending on the desired application. I have some high-level concerns detailed below about the current presentation of the model description and validation; hopefully most of these can be resolved through additional discussion and clarifications - and very happy to be corrected if I'm mistaken or have misunderstood anything! I also have some recommendations for additional analysis and validation that I think could be beneficial. I then finally list a few very minor technical comments/clarifications.

In terms of the big picture motivation of the paper, the main new development is the interactive coupling of the XGB fire model with the ELM land surface and dynamic vegetation model. Therefore, it seems strange that no results or analysis are presented showing the feedbacks that are possible because of this coupling. All the model outputs presented are focused purely on burned area validation - where in fact, the coupled ELM-ML4Fire-XGB model performs slightly worse than just using offline XGB, presumably due to the coupling feedbacks which influence the vegetation distribution. So as it stands, the paper doesn't really motivate why coupling these models is desirable; if you just want burned area accuracy, it's better to use the XGB model offline. The key benefit is presumably the feedbacks on vegetation distribution, carbon fluxes, etc. One would imagine that the interactive vegetation distribution in ELM is improved when it's impacted by a more realistic fire distribution, or that the feedbacks on vegetation due to changing fire regimes over time are better simulated. So, it would be nice to see some results showing how vegetation-related variables are impacted by the coupling, as is this presumably the main advantage of having such a coupled model.

We appreciate the reviewer's recognition of the value in our hybrid modeling approach. We agree that wildfire prediction, given its inherent unpredictability and sensitivity to numerous climatic, ecological, and anthropogenic factors, stands to benefit substantially from data-driven methods. In response to the reviewer's concerns, we have thoroughly revised our manuscript, specifically enhancing the model description section to clarify the underlying architecture and the integration between the ML fire model and the ELM land surface model. We have also expanded our validation section to address the reviewer's recommendations.

Regarding the model description (Section 2): Though I appreciate that the underlying land-surface model (ELM) and XGBoost wildfire model have been documented previously (though, the current XGB implementation appears slightly updated re. the datasets used, time period etc.), given that the coupling between these models is the central development of this manuscript I felt that the details of the models (particularly XGB) and the coupling were a bit brief, and it was hard to figure out the answer to certain relevant details. In particular:

- What are the respective model timesteps, and what is the coupling timestep? The XGB model was trained (I think?) to predict GFED5 monthly BA, so does this mean it runs on a monthly timestep? But, on P6, L62-63, the authors say "All the datasets are resampled to 0.25×0.25 spatial and annual temporal resolution" - so does this mean that it actually runs on an annual timestep? But then, in the coupled model, it's described that the output of XGB is passed to ELM to affect land surface properties at the next timestep, and vice versa - I don't fully understand how this works if XGB is being trained with annual inputs to predict monthly or annual GFED5 BA. ELM (I would assume?) has a much shorter timestep than annual/monthly, at least for properties like surface fluxes, soil moisture, LAI etc., as well as for the meteorological driving data used as input to the coupled model?

We apologize for the confusion. The ML fire model is trained at a monthly scale. While the ELM-BGC is driven by hourly meteorological forcing, the simulated surface meteorology and vegetation conditions are aggregated (averaged or accumulated) to a monthly scale. At the end of each month, these aggregated variables are fed into the ML fire model to predict burned areas, which are then used to update the vegetation properties. Section 2.3 has been rewritten for clarification. Please see the end of the response to this comment.

- What was the spatial resolution of ELM - does it match the 0.25 degree GFED5 grid that the XGB model (presumably?) provides output on?

Correct, both ELM-BGC and XGB file models have 0.25-degree resolution. We have clarified the model spatial resolutions in Section 2.3.

- There's insufficiently clear information about the XGB training process - the details are spread out in different parts of Section 2, and it's hard to work out exactly what were the inputs (including the time resolution, any dimension reduction that was applied, etc.), what was the target output, and what data was reserved for training vs validation.

We have rewritten Section 2.3 to clarify the input and output and the training process.

- "To reduce overfitting, we build a separate ML model for each year from 2001 to 2020 using the remaining 19 years' data" - I would stress that I'm not an ML expert, so maybe this is a simple lack of subject knowledge on my part. But it's unclear to me what is meant here - how is a separate model trained for each year, using data from other years? If the model is trained to predict the BA in one year based on the meteorological data of other years, it's not clear how it would learn the correct relationships. Or do the authors mean, that it is trained to predict BA relationships for all the other years, and then the trained model is applied to the one year that was left out, as the validation data? I'd appreciate if this could just be clarified a bit. Additionally: does this mean that the model(s) can only be used for years between 2001-2020? If so, that would seem to greatly limit it's usefulness for exploring future scenarios.

This is an excellent point, which was also raised by Reviewer #1. To address this, we have modified our training approach to build a canonical model. In the revision, we randomly split the 20-year monthly data into training and validation datasets, accounting for 80% and 20% of the entire dataset, respectively. The offline-XGB model is trained using only the training dataset to learn the relationship between the predictors and the burned area. Then, the offline-XGB is applied to the validation dataset to evaluate model performance on data not used during training.

- As with the points above - the nice schematic (Fig 1) shows the same meteorological and fire-specific input datasets being passed to ELM and the process-based fire model as to the XGB model, but it's unclear whether these inputs are provided at the same temporal and spatial resolution to the respective models, or whether there are intermediate pre-processing steps. I'm not sure how easy this is to depict in the schematic, but as mentioned it's also a little unclear from the text and table of inputs as well.

Fire-specific inputs, such as lightning, GDP, and population density, are aggregated annually. ELM-BGC, which runs with hourly meteorological forcing data, retrieves this information based on the year of the current timestep. The process-based fire model is called to simulate the burned area every hour, while the XGB model is trained and used at monthly intervals. To achieve this, we average the hourly meteorological conditions and simulated vegetation properties to obtain monthly means, and interpolate fire-specific variables from annual to monthly intervals using the nearest neighbor method. The XGB model is then trained at monthly intervals. When the offline-XGB model is coupled with ELM-BGC, the XGB model is called at the end of each month, while the ELM-BGC runs at hourly timestep. The hourly variables are accumulated internally to calculate the monthly mean at the end of each month.

We found it challenging to illustrate this in Figure 1; however, we have clarified it in the accompanying text.

- As I understand it, the XGB model is initially trained using PFT distributions diagnosed from a prior run of the ELM model using its processed-based fire scheme. However, the process-based scheme predicts a different fire distribution to the XGB model. Does this therefore introduce an inconsistency, i.e. that the XGB model is trained on PFT distributions that are predicated on the wrong spatial distribution of fires? Could this be improved by e.g. repeated iterations of running the ML4Fire-XGB coupled model to update the PFT distributions, and then re-training the XGB model? It would be good if the authors could comment on this.

This is another great point. The reviewer also suggested an approach that could enhance consistency between the XGB model and ELM-BGC. Currently, ELM is configured in the "biogeochemistry" (BGC) model, with PFT distributions prescribed based on satellite products. We have clarified this in the revised manuscript. Additionally, we are implementing our hybrid approach to ELM-FATES (Functionally Assembled Terrestrial Ecosystem Simulator), which will enable updates to account for the impacts of dynamic vegetation and plant demography.

Overall, regarding the comments about Section 2. We have rewritten this section as follows to address the comments.

Initialized with the quasi-equilibrium state from the spin-up simulation, we conduct transient simulations with the process-based fire model in the ELM-BGC, driven by hourly NLDAS-2 meteorological forcings at a 0.25º resolution from 2001 to 2020. The process-based fire model operates on an hourly basis, matching the frequency of the meteorological inputs, while the ML fire model is trained and applied at a monthly interval, consistent with GFED5 data intervals. For training the offline-XGB model, the ELM-BGC outputs, including LAI, surface soil moisture, and PFT fractions, are averaged to monthly intervals, combined with monthly mean meteorological conditions, socioeconomic variables (GDP, population density), and lightning (as detailed in Table 1) to learn the relationship between predictors *and burned area. To reduce overfitting, the 20-year dataset is split, with 80% used for training and 20% for validation. During training, grid cells with fewer than 30 months of non-zero burned area (~two-thirds of the total number of grid cells) are masked. This step is important to avoid feeding the ML model distinct predictor combinations that all correspond to zero burned areas, which could skew the model's learning process. Model performance was evaluated based on its accuracy in predicting the spatial distribution and temporal variation of burned areas. Validation metrics included root mean square error (RMSE) and the coefficient of determination (R2).*

We then integrate the offline-XGB to ELM-BGC, forming the coupled model ELM2.1-XGBfire1.0. The coupled model runs at 0.25º and hourly resolutions, where the hourly model predictions are accumulated to calculate monthly means. At the end of each month, the ML fire model is called to predict the monthly burned area, updating the land surface properties (LAI, PFT fraction, and soil moisture).

Regarding the comparison of burnt area results against four FireMIP models:

- Why those particular 4 models? E.g., the authors note that none of the models they compare against included a crop fire scheme, which is potentially one reason for poor performance over central US. However a couple of the FireMIP models that are not included here, did have crop schemes - so it seems odd to omit these. To be clear, I fully expect the ML4Fire-XBG model to outperform all the FireMIP models, it just seems a bit arbitrary why the comparison is made only against these four, out of nine FireMIP models that were included in the Rabin et al. (2017) paper. If, for instance, these were the four best performing models over the CONUS, then it could make sense to compare against just these rather than against all of them. But if that is the rationale, I couldn't see it mentioned anywhere (happy to be corrected though).

- All the figures comparing burnt area are described as an average over 2001-2020. However, the FireMIP experiments that are cited in Rabin et al. (2017) only went up to 2013. Even the most recent round of FireMIP (aka ISIMIP3a) I think only goes up to 2019. So as far as I can understand, the FireMIP data can't be for the same time period.

We understand the reviewer's concern. In our manuscript, we used outputs from FireMIP Phase II (also known as ISIMIP3a), in which one of this paper's coauthors participated. Although there is no specific protocol paper for ISIMIP3a, two relevant publications have recently been made available (Burton et al., 2024; Park et al., 2024).

We selected ISIMIP3a data because the fire simulations are conducted with updated fire models and with an extended simulation period to 2019. In alignment with the ISIMIP3a simulation period, we have adjusted our comparison period to 2001–2019. At the time we conducted this research, only four models had data available. We have now updated our analysis to include seven FireMIP models, as used in the latest FireMIP benchmarking study (Burton et al., 2024).

Reference:

Burton, C., Lampe, S., Kelley, D. I., Thiery, W., Hantson, S., Christidis, N., Gudmundsson, L., Forrest, M., Burke, E., Chang, J., Huang, H., Ito, A., Kou-Giesbrecht, S., Lasslop, G., Li, W., Nieradzik, L., Li, F., Chen, Y., Randerson, J., Reyer, C. P. O., and Mengel, M.: Global burned area increasingly explained by climate change, Nat Clim Change, 10.1038/s41558-024-02140-w, 2024.

Park, C. Y., Takahashi, K., Fujimori, S., Jansakoo, T., Burton, C., Huang, H., Kou-Giesbrecht, S., Reyer, C. P. O., Mengel, M., Burke, E., Li, F., Hantson, S., Takakura, J., Lee, D. K., and Hasegawa, T.: Attributing human mortality from fire PM2.5 to climate change, Nat Clim Change, 10.1038/s41558-024-02149-1, 2024.

- The authors don't mention or discuss (as far as I could see) some very important caveats which really need to be attached to the comparison with FireMIP models. In particular, it should be noted that the process-based FireMIP models were run with different reanalysis driving data, at a much coarser spatial resolution. I appreciate that being able to run much higher resolution is a potential advantage of using an ML model. But, it needs to be acknowledged that it's not a like-for-like comparison of pure model process accuracy. The different driving data (from a different, global reanalysis product, provided at a much lower resolution than the North America-specific reanalysis data that the XGB model is driven by) is potentially a very important factor - a fairer comparison of performance would be to run the XGB model driven by the FireMIP driving data.

Thank you for pointing this out. In our model simulation, we adopted the same lightning, CO_2, population density, and GDP data used in ISIMIP3a, with the exception of the climate forcing data. To focus on fires in CONUS, we applied the upscaled hourly NLDAS-2 climate forcing at a spatial resolution of 0.25º, rather than the daily GSWP3-W5E5 forcing at 0.5º used in ISIMIP3a. This different reanalysis data source and differences in the spatial and temporal resolutions of the climate forcing could contribute to variations in burned area predictions.

Besides ISIMIP3a models, we also conducted ELM-BGC (with built-in process-based fire model) simulations driven by the same set of climate and socioeconomic forcing data as used to drive the coupled model ELM2.1-XGBfire1.0. The results show that the burned area simulation in ELM-BGC remains unsatisfactory, indicating that changes in climate forcing alone do not account for all limitations in burned area simulations in process-based models (at least for ELM-BGC). We have added the following discussion to the revised manuscript to clarify this point (Line 360-363).

The ISIMIP3a models were driven by daily GSWP3-W5E5 forcings at a 0.5º spatial resolution. Differences in forcing data could lead to variations in burned area predictions. However, since both ELM-BGC and ELM2.1-XGBfire1.0 are driven by the same set of forcings, this suggests that limitations in physical understanding may significantly hinder the performance of the process-based model.

Regarding the discussion and model validation:

- As mentioned above, it would be good to have some more quantitative discussion on the advantages of having the coupling, e.g. for vegetation distribution

Thank you for highlighting this point. First, we would like to clarify that in our current configuration, ELM is set up as a "biogeochemistry" (BGC) model, with PFT distributions prescribed based on satellite observations; we have added this clarification in the revised manuscript. Additionally, we are working to implement a hybrid approach within ELM-FATES (Functionally Assembled Terrestrial Ecosystem Simulator), which will allow the model to dynamically update vegetation and plant demographics, incorporating the impacts of changing vegetation structure over time.

Second, within the CONUS, fires primarily impact the terrestrial carbon cycle at localized scales, and their broader influence across the entire region is limited. ELM-BGC, like many terrestrial models, currently exhibits a significant bias in gross primary production (GPP) predictions across the CONUS (refer to the figure below). The coupled model, ELM2.1-XGBfire1.0, provides a significant improvement in fire prediction and slightly reduces the GPP underestimation compared to ELM-BGC, though this effect remains limited. Furthermore, converting burned area into carbon loss involves several uncertain parameters, which we did not optimize in this study.

Nonetheless, the coupling remains valuable, especially for higher-resolution model runs, examining fire-induced tree mortality, post-fire recovery, fire emissions, and fire-related air quality issues. This importance is amplified when ELM is run with its atmospheric component, E3SM, where the influence of fires on air quality, cloud, and surface meteorology becomes more significant.

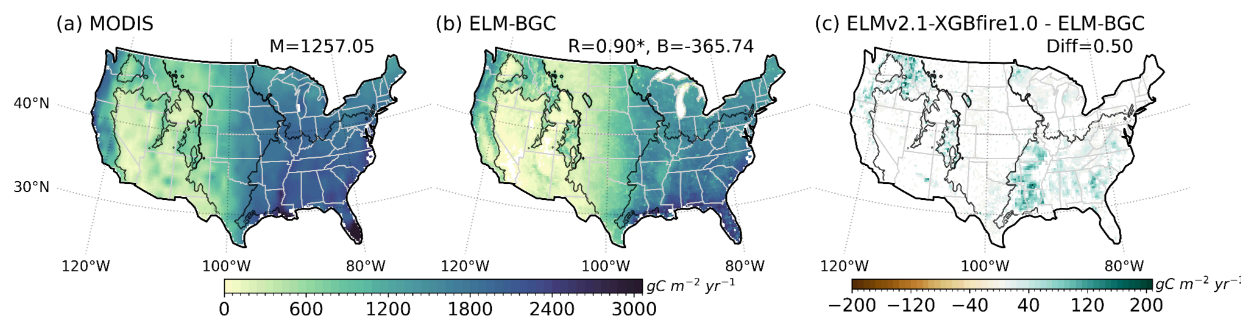

Figure. Observed and simulated GPP (gC m-2 yr-1) averaged over the period 2001–2019. The dataset names are listed at the top of panels a and b. Panel c shows the difference between ELMv2.1-XGBfire1.0 and ELM-BGC. The numbers indicate the mean (M), bias (B), pattern correlation (R) against MODIS, and difference (Diff) between the two models in panel c. Black contours outline the eco-regions.

The following paragraph has been added to the discussion in our revision (Line 382-390).

Although ELM2.1-XGBfire1.0 significantly improves the simulation of burned areas, its impact on terrestrial carbon fluxes remains limited. Within the CONUS, fires primarily affect the terrestrial carbon cycle at localized scales due to the relatively small burned areas. ELM-BGC, for instance, underestimates gross primary production (GPP) by approximately 30% (figure not shown). With more accurate fire predictions, ELM2.1-XGBfire1.0 helps to slightly reduce this negative bias (less than 1%). Additionally, while ELM-BGC using prescribed PFT distributions can moderate the effects of fires on the ecosystem, it does not account for fire-induced shifts in vegetation species, where species with greater resistance or fire-adaptive traits may gradually dominate.. Nonetheless, the coupling remains valuable, especially when the model is configured at higher resolutions. It is particularly important for evaluating fire-induced tree mortality, post-fire recovery, fire emissions, and their subsequent impacts on air quality, cloud formation, and surface meteorology, particularly when ELM is run as part of the E3SM.

- All the comparison of BA performance is performed against GFED5, which is the same data that the model was trained on. Arguably, it's quite unsurprising that an ML model trained to predict GFED5 over CONUS from 2001-2020, would do better at predicting GFED5 over CONUS from 2001-2020, compared with global process-based models that were not specifically optimised to do this. Ideally, performance would be evaluated with out-of-sample tests - for example, by running the ML4Fire-XGB model with the FireMIP inputs as mentioned previously, or by comparing against alternative datasets and/or over different time periods to the training period.

We thank the reviewer for this valuable comment. We agree that the global process-based models were not specifically optimized for GFED5 and, importantly, not tailored for the CONUS region. Historically, the CONUS has not been a fire-prone area, and it has unique fire characteristics that may not be well represented in global models. Our analysis underscores that critical processes relevant to fire activity in CONUS may be missing in these global fire models. Improving the physical understanding of these processes and refining model parameters could enhance the performance of process-based models in capturing fire regimes over the CONUS.

In response to the reviewer’s suggestion, we have added a new figure (Fig. 3, also shown below) to include GFED4s and FireCCI5.1 as additional reference datasets to account for observational uncertainties. Both GFED4s and FireCCI5.1 show a comparable spatial distribution to GFED5, with a spatial correlation coefficient exceeding 0.66. GFED5 includes more small fires, which are not detected in GFED4s and FireCCI5.1, leading to a burned area estimate that is 110% larger than the other two datasets. All process-based models overestimate burned areas when compared to GFED5, suggesting an even greater overestimation relative to GFED4s and FireCCI5.1.

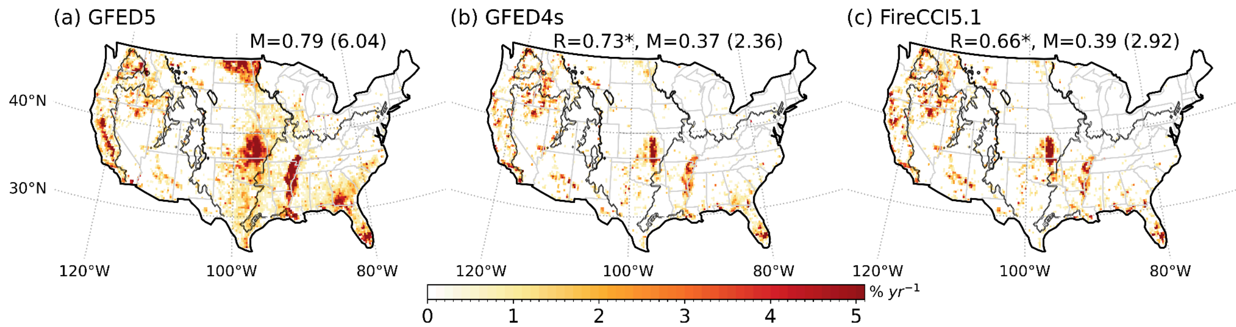

Figure 3: Observed burned area fraction (% yr-1). (a) GFED5 (2001-2019), (b) GFED4s (2001-2016), and (c) FireCCI5.1 (2001-2019). The numbers indicate the mean (M) burned area fraction and burned area (in Mha) in brackets for each dataset. The pattern correlation (R) against GFED5 is also shown, with an asterisk (*) denoting significance at the 0.01 level. Black contours outline the ecoregions.

- On that note: the authors assume GFED5 is the ground truth in evaluating that the XGB model outperforms process-based models, but it should be acknowledged that there is a large uncertainty in the observation-based BA data. This observational uncertainty should also be addressed, for example by including comparisons not just against GFED5, but against alternative BA datasets that are available, for example the USGS Landsat Burned Area product for the CONUS, or one of the FireCCI BA products.

We have added GFED4s and FireCCI51 in the revision. Please see the response above.

- P16, L51-53: "However, the ML-fired process exhibits high accuracy, as demonstrated by the Offline-XGB model, making it a reliable tool for evaluating the fired area under different warming scenarios" - the authors show that the XGB model captures well the trend in GFED5 BA due to warming during the 2001-2020 period. However, this is the same period that the model was trained to perform well on. How reliable an indication is this that the model will still be accurate under high-end future warming scenarios, where the degree of climate change over the US will substantially exceed that observed over 2001-2020? This should be discussed.

We have removed the warming-temperature experiment from this paper, as suggested by Reviewer #1. We appreciate the reviewer's insight, and we are evaluating the model responses to high-end future warming scenarios in a separate study.

Minor/technical clarifications:

P2, L27-29: "Over the globe, climate change has contributed to a 16% increase in the global burned area over the past two decades, while human influences, including ignition and suppression, have reduced by 27% (Burton et al., 2023; Jones et al., 2022)" - as it reads, I don't think this isn't an accurate paraphrasing of the studies being referenced.

Currently it reads (to me) like: there has been a 16% increase in BA over the last two decades, to which climate has been a main driver, while the influence of humans has reduced by 27%. This isn't what either of these studies showed. Burton et al. find (based on FireMIP model data) that climate change since 1901 has made average BA (median over the 2003-2019 period) 16% higher than it would have been if climate stayed fixed at circa ~1901. However, they also find that human influences have made median BA 19% lower in the present compared with early 20th Century, suggesting the net effect over the 20th Century is a small decline in BA. Jones et al. show that, in MODIS BA data, total global BA has declined by 27% over the last two decades. This is similar to previously reported results from GFED4 and GFED5, which both show ~24% declines in total global BA over the last two decades. The reason for this decline has been attributed mainly to human influences (c.f. also Andela et al. 2017).

(As an aside: since the present manuscript was submitted, the Burton et al. (2023) pre-print that is referenced has now been published as a final article, and so the citation should be updated accordingly: https://www.nature.com/articles/s41558-024-02140-w)

Thank you for pointing this out. The abovementioned studies did show that direct human influences (suppression, agriculture expansion) have played a negative role in the increased global burned area. We apologize for the confusing words in the previous sentence. This clarification has been made in the revised manuscript (Line 27-31) with the paper citation updated.

Globally, modeling studies show that climate change since the early 1900s has contributed to a 16% increase in the total burned area; however, human activities have led to a 19% decrease, resulting in a slight net decline in burned area over the 20th century (Burton et al. 2024). In the past two decades, satellite-derived data suggest that the global total burned area has declined by over 20%, with this trend primarily attributed to human influences (Jones et al. 2022; Andela et al. 2017).

Section 2.2.3: From the looks of it, the authors use existing Level 1 EPA ecoregions for their analysis regions 1, 2, and 3, but then for their regions 4 and 5 they split the Eastern Temperature Forest Level 1 EPA region into two. What was the rationale for subdividing this region but not the others? Also, this section seems oddly located in the middle of the model description, in between the description of the individual model components (2.1) and the description of the coupling (2.3), even though it relates only to the analysis of the final results and is not part of the model description. It would be better to have this section on analysis methods after all the description of the models and model coupling, I think.

We used a combination of Level I and Level II EPA ecoregions. In the Level II data, Regions 4 and 5 are further divided: Region 4 (southeastern U.S.) includes the Southeastern Plains and Appalachian Forests, while Region 5 (northeastern U.S.) includes the Mixed Wood Plains and Atlantic Highlands. These two regions exhibit distinct fire patterns, as shown in Figure 3 of the manuscript.

Following your suggestion, we have moved this section to follow the descriptions of the models and model coupling. Thank you for the helpful feedback.

P8, L04-05: "According to the GFED5, the CONUS experiences an averaged burned area fraction (BAF) of 0.6–0.9% yr-1 (4.8–7.1 Mha yr-1), which is consistent with Chen et al., (2023)" - Not quite sure what the authors intended here. Chen et al. (2023) is itself the GFED5 burned area description paper, so trivially GFED5 burnt area is consistent with itself.

“Which is consistent with Chen et al. (2023)” has been removed from the revision.

P8, L05-06: "High-burned areas are predominantly observed in the WUS" - this seems an confusing statement, since the authors then go on to list other areas which have higher BA than the WUS, and indeed Figure 3 seems to show other areas of the US where BA is higher and more widespread.

This sentence has been rewritten as:

The BAF over the WUS (Western Forested Mountains and NA Desert) ranges between 0.4–0.9% yr^{-1} (1.1–2.3 Mha yr^{-1}).

P10, L33-34: "indicating that the ML model effectively describes crop fire thereby utilizing data on crop fraction and LAI" - Is this referring to the crop PFT fraction in the ELM model? (Rather than agricultural land use fraction, which isn't listed as an input)?

Yes. It refers to the crop PFT fraction in the ELM model.

P10, L52-53: "Notably, none of the process-based models has activated the explicit cropland fire model. That says all vegetation models treat pastures as natural grasslands." - This statement is slightly confusing and conflates two things. Pasture is not the same as cropland, and they are usually represented as different land cover types in DGVMs. Similarly crop residue burning is a very different fire regime to pasture fires.

Thank you for raising this point. We agree that there are fundamental differences between cropland fires and pasture fires. To clarify, all FireMIP models in this study exclude cropland fires (Burton et al., 2024; Extended Data Table 1). Additionally, all models except the LPJ-GUESS DGVMs do not explicitly represent pasture as a separate land cover type and, therefore, do not include pasture fires (Teckentrup et al., 2019). We have revised this sentence for clarity in Lines 276-279.

As noted by Teckentrup et al. (2019) and Burton et al. (2024), none of the process-based models has activated the explicit cropland fire model. Fires are allowed in pastures. While LPJ-GUESS-SIMFIRE-BLAZE incorporates harvesting in pastures, reducing biomass and *influencing fire dynamics, all other process-based vegetation models treat pastures as natural grasslands for both vegetation growth and fire processes.*