Response to comments

Manuscript ID: essd-2024-556

Title: *Reconstructed global monthly burned area maps from 1901 to 2020*

Journal: *Earth System Science Data*

Dear editor and reviewers,
In the revision, we carefully addressed the reviewers' concerns (see point-by-point responses below) and revised the Main Text and Supplementary Information with blue for newly added and black for unchanged.

<u>Reviewer #2:</u>

General Comments

Comment #1

Earth System Science Data – Guo et al. – Feb 2025
This study reconstructed global monthly burned area from 1901 to 2020 at 0.5°×0.5° resolution using machine learning models trained on satellite data (2003–2020). Separate models were developed for extreme and regular fires based on climate, vegetation, and human activity data. The models accurately captured spatial patterns, seasonal trends, and long-term changes in burned area. Results show a global decline in burned area from 1901–1978, an increase from 1978–2008, and a stronger decline from 2008–2020. The reconstruction aligns well with charcoal records, offering a valuable tool for historical fire analysis.

The manuscript is well written, and the simulation work—including data preparation, model selection, training, validation, and regional application—is rigorously conducted. It can be shown that the authors put quite a lot effort to build such machine learning and data driven pipeline. The use cases from such data asset construction are wide. The results are clearly presented through both text and figures, and the discussion addresses the burned area distributions, trends, limitations, and uncertainties. However, I have a few specific comments regarding the methods that I believe should be clarified in more detail in the Methods section. Therefore, I recommend a minor revision to further strengthen the manuscript prior to publication.

Response #1

We thank the reviewer for the constructive and positive comments. Following your insightful suggestions, we have carefully revised the manuscript. Please find the point-by-point responses below.

Specific Comments

Comment #2

Line 95 Please add more content / citations on how the burned area on cropland is excluded to eliminate agricultural fires in this study.

Response #2

As suggested, we revised the description in the **Section 2.1 of Main Text**.

Line 95-96: "We excluded all burned pixels overlapping cropland classes in the CCI land-cover layer provided with FireCCI51 (Lizundia-Loiola et al., 2020) to remove agricultural fires from our analysis."

Comment #3

Line 120 Please clarify how the reclassification is done to split them into five land use types (forest, shrub, natural grass, cropland and others). Was it via machine learning classifier?

Response #3

As suggested, we added description accordingly in the **Section 2.1 of Main Text**.

Line 121: "The above five land use types were converted from the ESA CCI land cover maps based on the cross-walking table (Li et al., 2018). For the LUH2 dataset, we reclassified land use by summing forested primary land (primf) and potentially forested secondary land (secdf) to create a single "forest" category, and by summing all crop types (c3ann, c3per, c3nfx, c4ann, and c4per) to form the "cropland" category. To define natural grass and shrub, we first combined non-forested primary land (primn) and potentially non-forested secondary land (secdn) into a unified grass + shrub type. We then allocated this combined area back into separate grass and shrub categories based on their proportional distribution. For historical years before 1992, the proportional distribution was set the same as ESA CCI land cover in 1992, and for years in 1992-2020, the proportional distribution was set according to the corresponding year of ESA CCI land cover."

Comment #4

Line 148 Is there reasoning why an initial classification has to be conducted to classify regions into non-BAF, moderate and extreme BAF? Why couldn't the parameterization be applied to all regions without this initial classification. Conducting such classification could introduce extra errors / uncertainties. Please add more content to clarify.

Response #4

Thanks for the reviewer's comment. We need this initial classification because of imbalanced samples. We explained it in **Line 96-98 in Main Text**: "The distribution of monthly burned area within half-degree grid cells manifests that split of regular and extreme burned area amplifies the kernel density of extreme burned area for better model training (Fig. S2)."

Due to the imbalanced samples, too many grid cells with no BAF and regular BAF would lower the weights of extreme BAF in the model training, leading to underestimation of the large BAF and the total burned area. We thus excluded zero-BAF and conducted separate regression models for regular and extreme BAF. We added further clarification in the **Section 2.1 of Main Text** as follows:

Line 96-98: "The monthly distribution of burned area within 0.5°×0.5° grid cells (Fig. S2) shows that if regular and extreme fires are modeled together (black curves), the abundant moderate values drown out the extremes (orange curves), causing total area to be underestimated. We thus first conducted classification and then trained separate models for regular and extreme burned fractions to enhance the representation of extreme events and improve regression performance."

Comment #5

Line 152 I am not quite clear why 90th percentile instead of other percentiles is used to determine the extreme BAF. Please explain and clarify.

Response #5

As suggested, we added further clarification in the **Section 2.1 of Main Text** to explained why 90^{th} percentile was chosen in the classification.

Line 96: "We used the 90^{th} percentile of all burned area fractions in 0.5°×0.5° grid cells within a region as the threshold to define extreme fires. This percentile was chosen based on previous literature (Bowman et al., 2017; Cunningham et al., 2024; Lannom et al., 2014). It is high enough ($\geq 90^{th}$) to distinguish moderate from extreme samples to train separate models for each category. Meanwhile, it is not too high (e.g., 95^{th} or 99^{th}) in regions with limited data (such as Europe and the Middle East) to ensure sufficient extreme samples for model training and evaluation."

Comment #6

Line 167 From the content there are only 16 features plus some others in NHAF are used in the model. I am not sure why a recursive feature selection has to be performed. Recursive feature selection is usually applied when there is 1000+ variables used in GBM models. This process is necessary because the larger model package size would directly cause latency issues in the live production env. In this case 16+ variables won't cause such issues plue the model will not be deployed onto live env. Is such feature selection necessary?

Response #6

We agree that recursive feature selection is essential when working with a small sample size and many candidate predictors. Here, we used it to prevent performance degradation that can occur when adding irrelevant features (Guyon and Elisseeff, 2003). Moreover, reducing the feature set enhances interpretability—facilitating the analysis of feature interactions via SHAP—and conserves computational resources. We added explanation in the **Section 2.2 of Main Text** as follows:

Line 170: "The recursive feature elimination cross validation was applied to prevent model performance degradation if irrelevant features were added (Guyon and Elisseeff, 2003). Moreover, reducing the feature set could enhance model interpretability and conserve computational resources (Lundberg et al., 2020)."

Comment #7

Figure 2 For the global map of burned area difference please make it a separate figure as it is hard to see the small dots. And what is the unit? Is it a relevant difference? Please specify in the figure.

Response #7

As suggested, we split Figure 2 of the previous version into Figure R2 and Figure R3. The unit of the global map in Fig. R2 is fraction difference in 0.5°×0.5° grid cell, and it is not a relative difference. We clarified it in the caption of Fig. R2 **in Main Text**.

Comment #8

Figure 2 What does N mean here? Is it the data sample size for validation grids per year? Please specify.

Response #8

N represents number of grid cells with multi-year averaged BAF>0. We removed dots with both predicted and observed burned area fraction equal to 0 in the scatter plots of Fig. 2 because we aimed to emphasize the evaluation of model performance among grid cells with BAF>0. Therefore, N, R2, slope, p and RMSE denoted in scatter plots could be different from Fig. 2 in the original version. We clarified it in the caption (now **Fig. 3 in Main Text**).

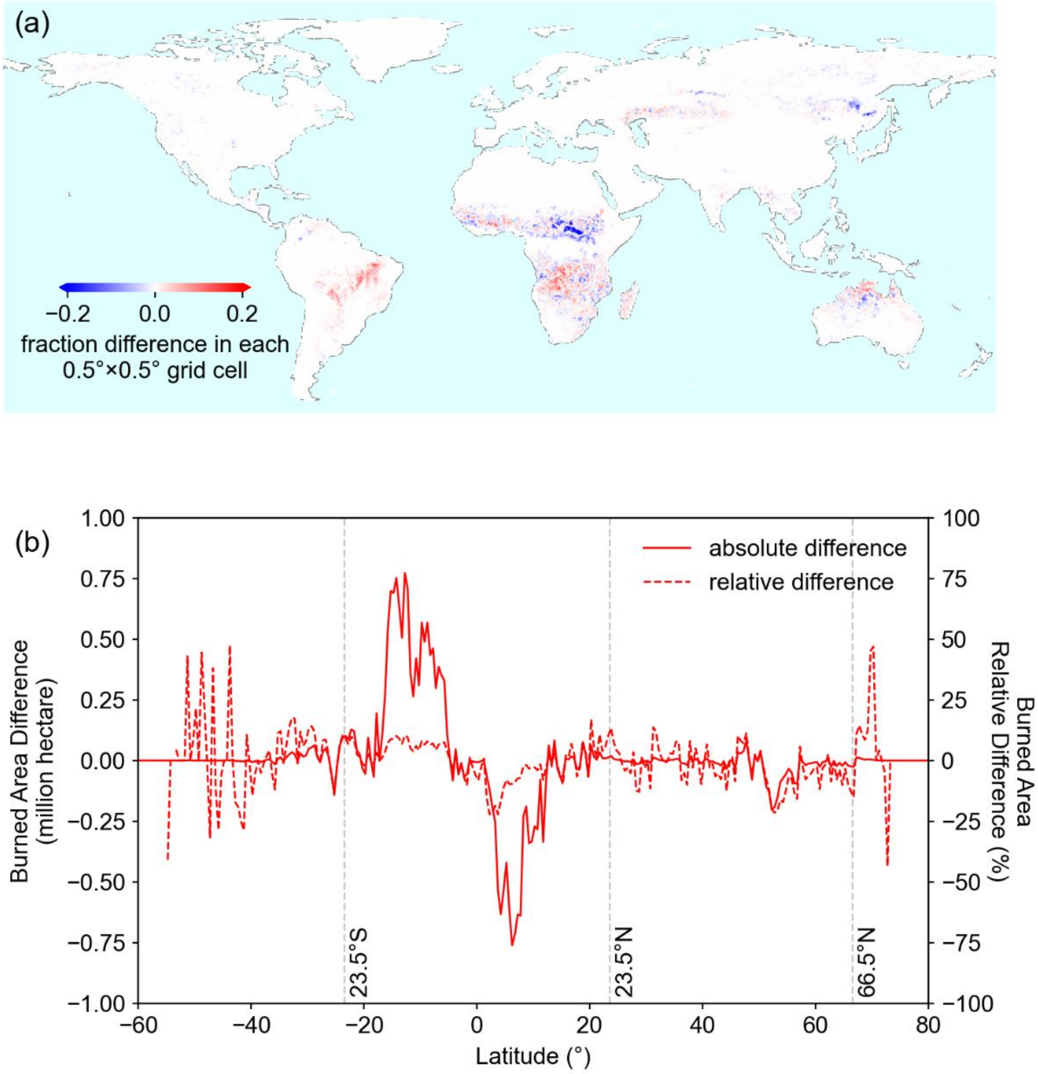

Figure R2 (revised as Figure 2 in Main Text): Multi-year (2003-2020) averaged burned area difference between our predictions by the leave-one-year-out method and FireCCI51 observations (predictions minus observations). (a) Map of burned area fraction difference in each 0.5°×0.5° grid cell. Burned area fraction difference is the ratio of burned area difference to total grid area within each 0.5°×0.5° cell, making it unitless and bounded between 0 and 1. (b) Latitudal sum of burned area difference using the burned area fraction difference map from (a) multiplied by the area of each 0.5°×0.5° grid cell. Both absolute (solid line) and relative (dashed line) differences are shown.

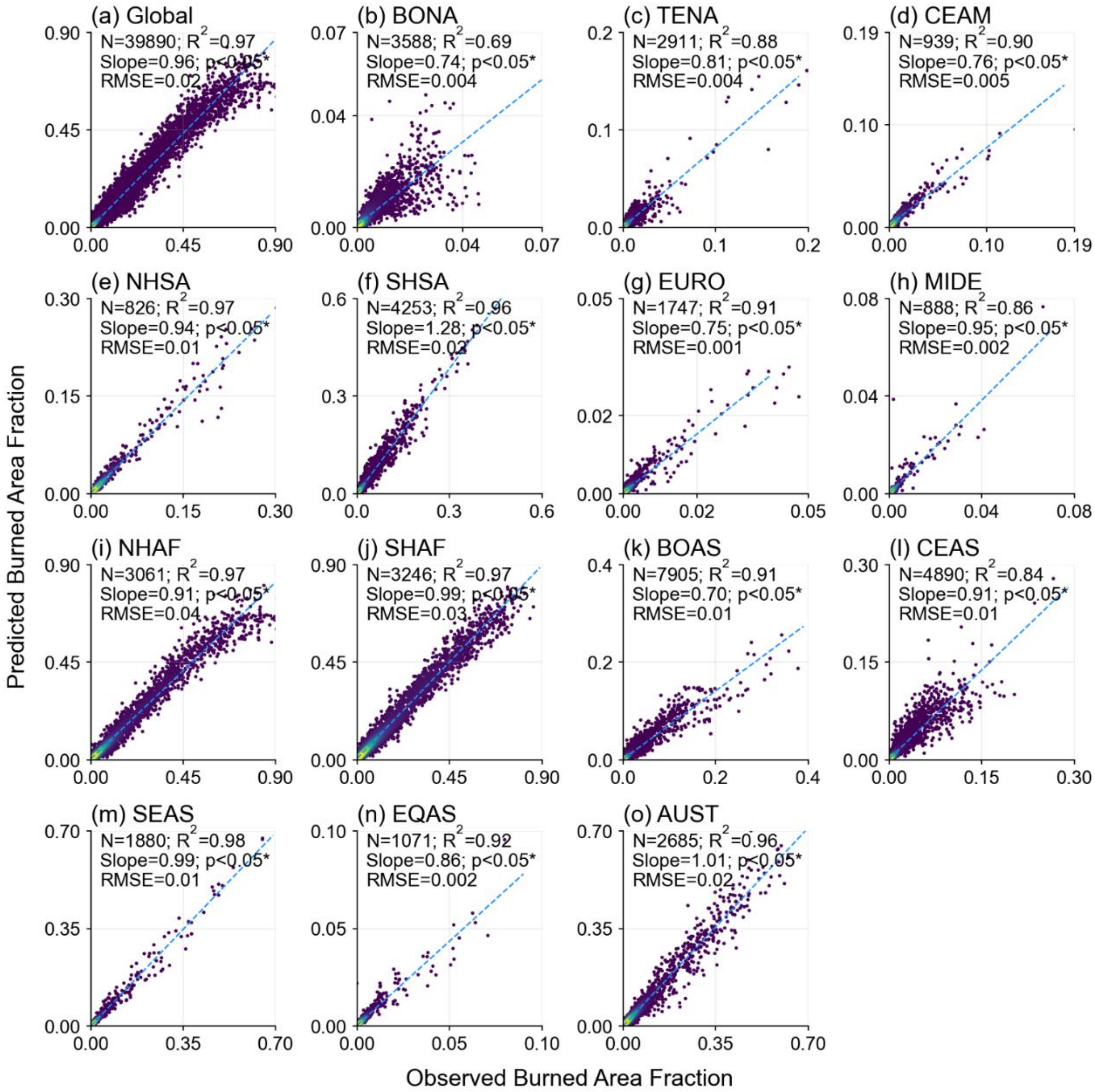

Figure R3 (revised as Figure 3 in Main Text): Scatter plots of multi-year (2003-2020) averaged burned area fraction (BAF) in each 0.5°×0.5° grid cell from predictions by the leave-one-year-out method and FireCCI51 observations for each region (a-o). Dots represent grid cells with BAF>0 averaged over 2003-2020. N, R^2, slope, p and RMSE respectively represent number of grid cells with multi-year averaged BAF>0, coefficient of determination, linear slope, p-value for linear correlation and rooted mean squared error between BAF from our predictions and observations. Burned area fraction is the ratio of burned area to total grid area within each 0.5°×0.5° cell, making it unitless and bounded between 0 and 1.

Comment #9

Figure 3 It looks like figure (a) (c) (e) vs figure (b) (d) (f) are giving the same messages. Also it is hard to find any insight from figure (a) (c) (e). Consider removing or merging them with (b) (d) (f).

Response #9

As suggested, we merge (b), (d), (f) into (a), (c), (e) as shown in **Fig. R4**. We revised the caption of this figure **in Main Text** accordingly as follows.

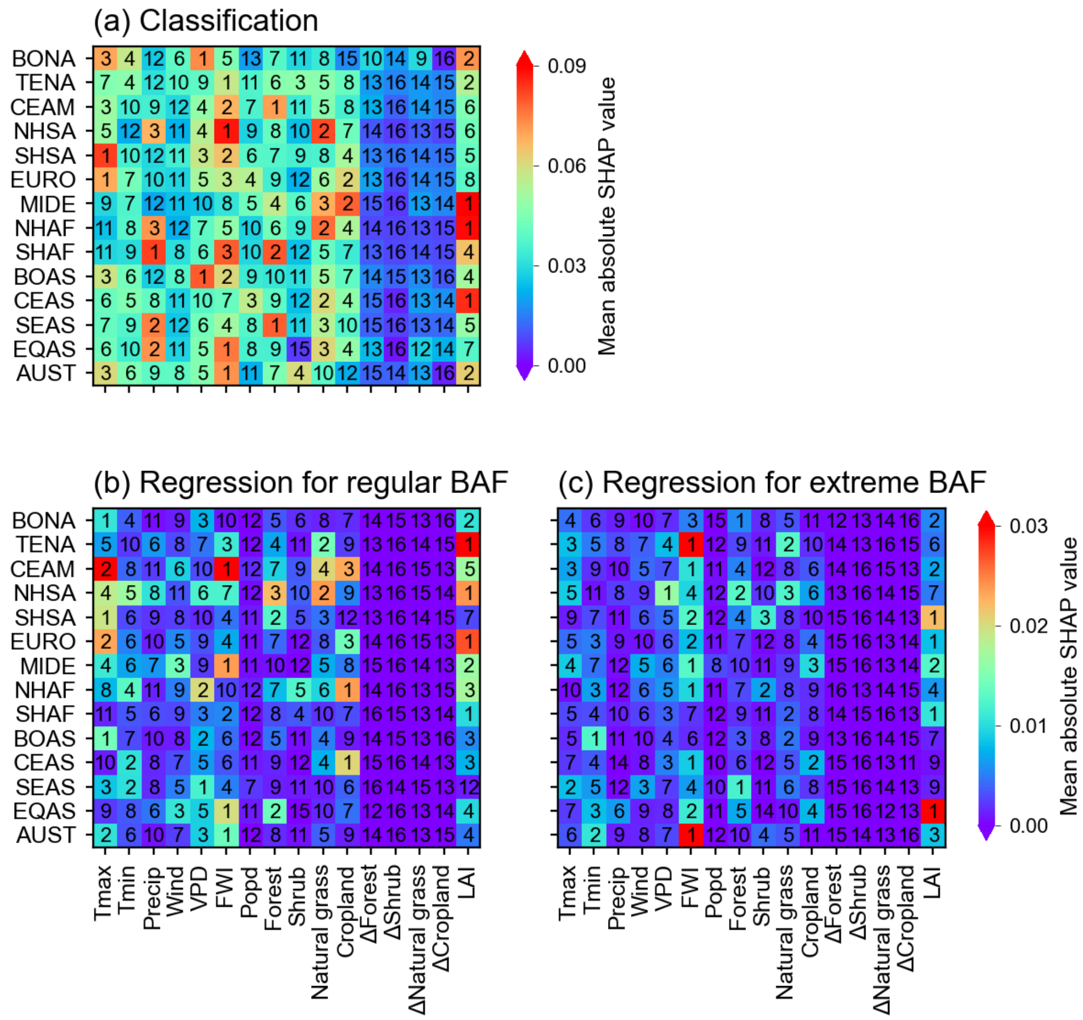

Figure R4 (revised as Figure 4 in Main Text): Mean absolute SHAP value and the ranking of all input variables (Table 1) using the random forest classification models (a) and the LSTMs regression models for regular (b) and extreme (c) BAF, respectively, in each region. Numbers denoted in grids are the ranking of variables, and higher ranking represents relative higher mean absolute SHAP value in the corresponding GFED region.

Comment #10

Figure 5 It is hard to tell the accuracy difference in (a) (b) as most of them are dark red (are most of them 100% or 70%?). Please consider using a different color scale to make them more distinguishable.

Response #10

We changed the color scale as suggested (**Fig. R5**) **in Main Text**.

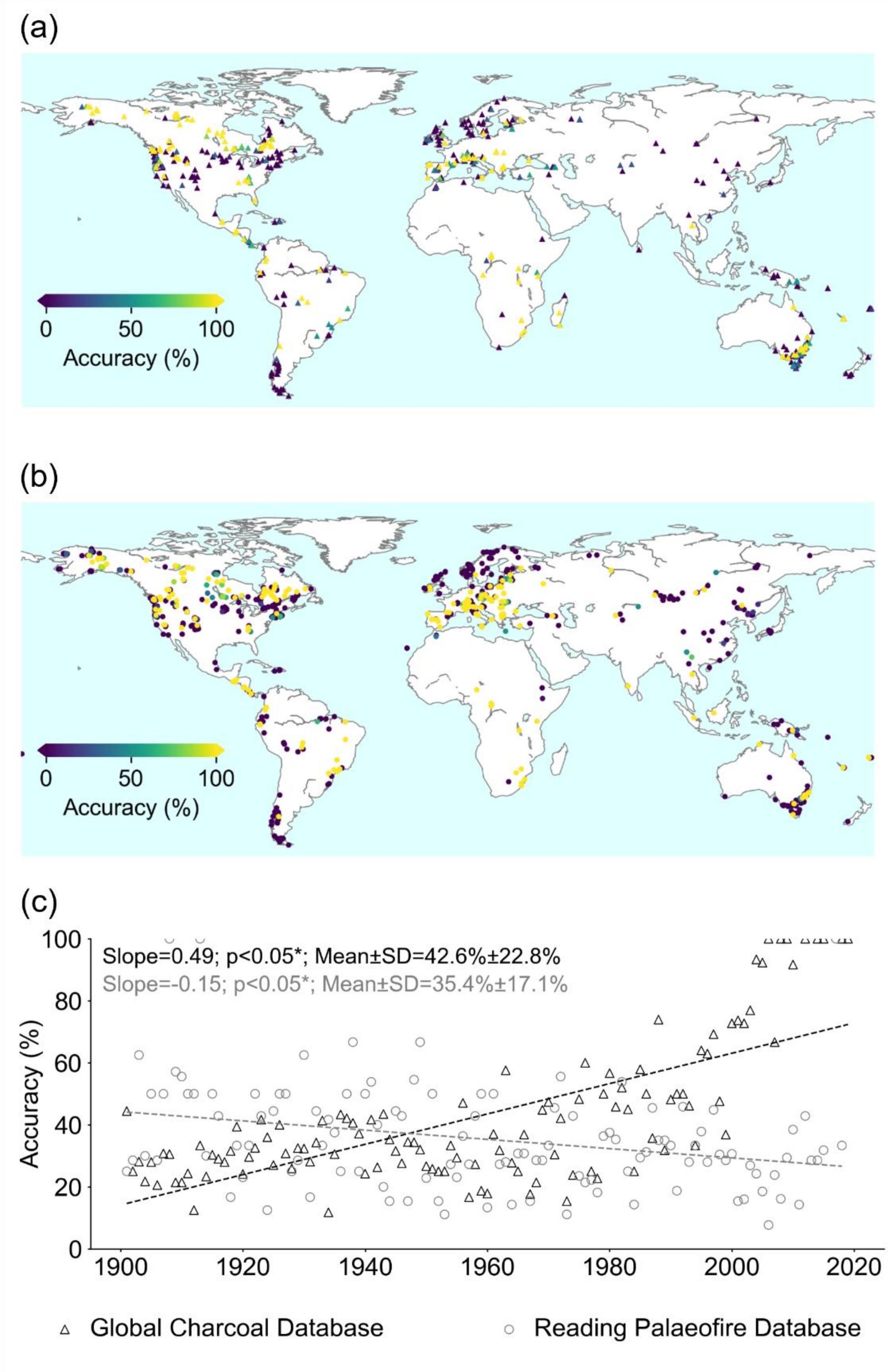

Figure R5 (revised as Figure 6 in Main Text): Fire occurrence comparison between two charcoal record databases and our prediction from 1901 to 2020. (a) Site accuracy map using Global Charcoal Database. The site accuracy (%) is equal to the number of records with predicted burned area dividing the number of all records multiplying by 100%. (b) The same as (a) but using Reading Palaeofire Database instead. (c) Accuracy time series using Global

Charcoal Database and Reading Palaeofire Database respectively. Note that only records with record year ± record age uncertainty overlapping with 1901-2020 are taken into consideration.

Comment #11

Figure 6 Again, (a) (b) (c) convey important messages of spatial distribution of burned areas. but they are too small in terms of figure size. Please make them bigger / clearer.

Response #11

We split the **Fig. 6** in the previous version into **Fig. R6** and **Fig. R7** in the current version to make them bigger and clearer. We also revised their captions and citations accordingly **in Main Text** as follows:

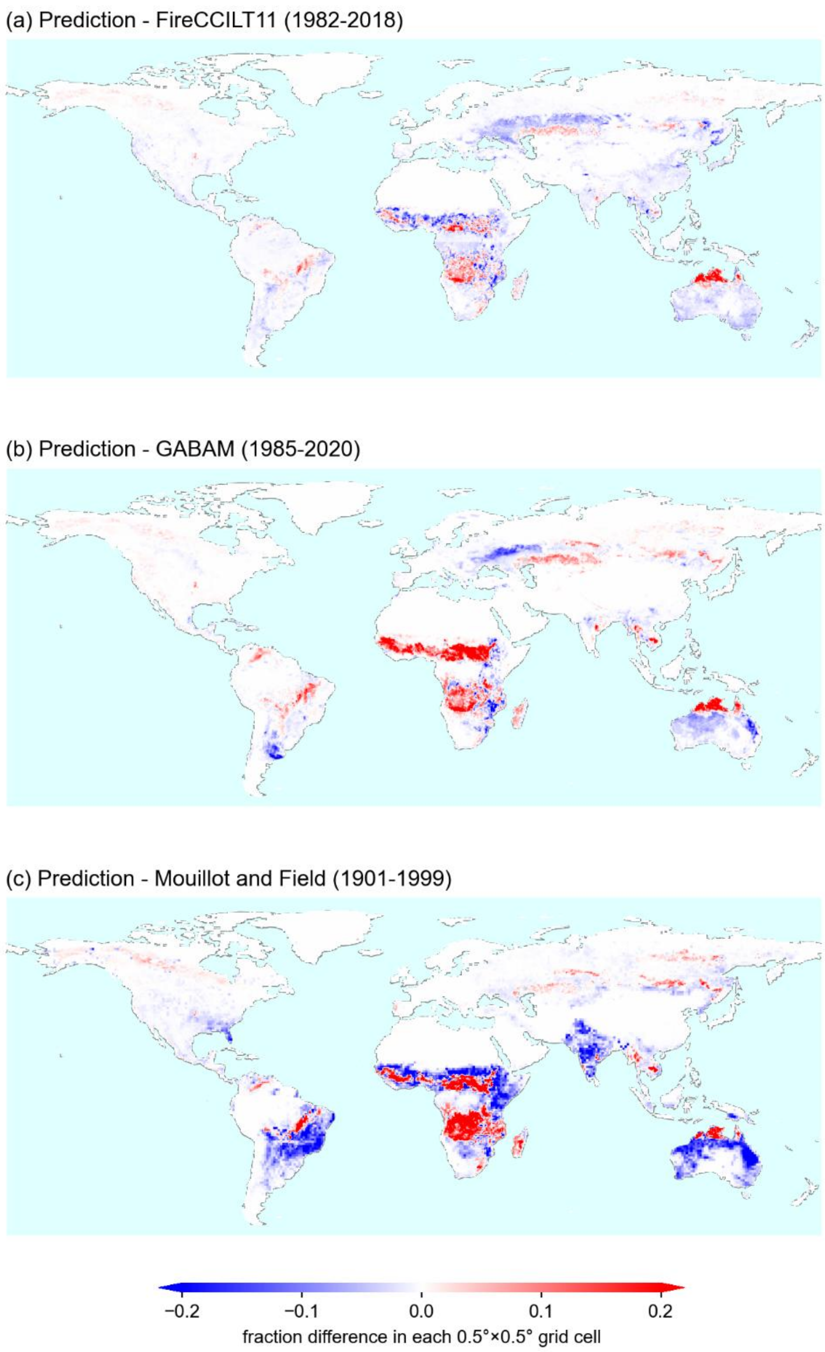

Figure R6 (revised as Figure 7 in Main Text): Maps of burned area fraction difference between our predictions and other global burned area datasets. (a) Map of multi-year average (1982-2018) burned area fraction difference map between our predictions and FireCCILT11 (the former minus the latter). (b, c) Same as (a) but using Global Annual Burned Area Maps (GABAM) (1985-2020) and Mouillot and Field (2005) (1901-1999) instead respectively. Note that there are several years (1986, 1988, 1990, 1991, 1993, 1994, 1997 and 1999) without available data before 2000 in GABAM.

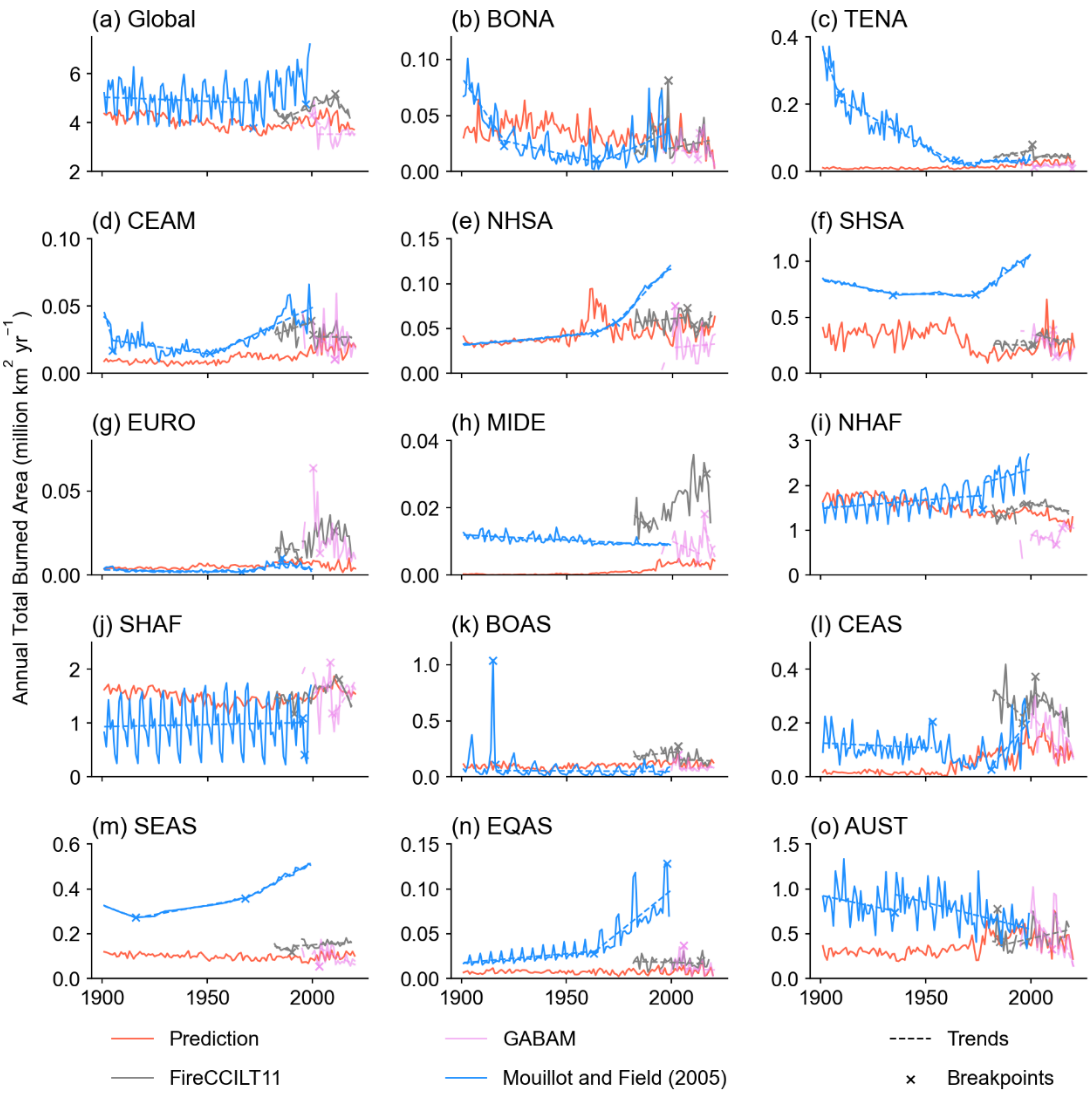

Figure R7 (revised as Figure 8 in Main Text): Time series of annual total burned area across the globe (a) and in each region (b-o) from our predictions (red lines), Mouillot and Field (2005) (blue lines), FireCCILT11 (grey lines) and GABAM (purple lines). The breakpoints and significant slopes (p-value<0.05) were calculated by methods mentioned in Sect. 2.2. Note that there are several years (1986, 1988, 1990, 1991, 1993, 1994, 1997 and 1999) without available data before 2000 in GABAM, and thus breakpoint detection and linear slopes were applied after 2000 for this dataset.

References:

Bowman, D. M. J. S., Williamson, G. J., Abatzoglou, J. T., Kolden, C. A., Cochrane, M. A., and Smith, A. M. S.: Human exposure and sensitivity to globally extreme wildfire events, Nature Ecology & Evolution, 1, 0058, 10.1038/s41559-016-0058, 2017.

Cunningham, C. X., Williamson, G. J., and Bowman, D. M. J. S.: Increasing frequency and intensity of the most extreme wildfires on Earth, Nature Ecology & Evolution, 8, 1420-1425, 10.1038/s41559-024-02452-2, 2024.

Guyon, I. and Elisseeff, A.: An introduction to variable and feature selection, J. Mach. Learn. Res., 3, 1157–1182, 2003.

Lannom, K. O., Tinkham, W. T., Smith, A. M. S., Abatzoglou, J., Newingham, B. A., Hall, T. E., Morgan, P., Strand, E. K., Paveglio, T. B., Anderson, J. W., and Sparks, A. M.: Defining extreme wildland fires using geospatial and ancillary metrics, International Journal of Wildland Fire, 23, 322-337, 2014.

Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., Houghton, R. A., and Peng, S.: Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015), Earth Syst. Sci. Data, 10, 219-234, 10.5194/essd-10-219-2018, 2018.

Lizundia-Loiola, J., Otón, G., Ramo, R., and Chuvieco, E.: A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data, Remote Sensing of Environment, 236, 111493, https://doi.org/10.1016/j.rse.2019.111493, 2020.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I.: From local explanations to global understanding with explainable AI for trees, Nature Machine Intelligence, 2, 56-67, 10.1038/s42256-019-0138-9, 2020.