Reviewer 2:

The authors seek to quantify how geoengineering might be expected to affect wildfire occurrence, extent, and carbon emissions towards the end of the 21st century. They compare four different scenarios: two of the world without any geoengineering (one "cooler" - SSP2-4.5 - and one "warmer" - SSP5-8.5), one in which idealized (solar shade) geoengineering is applied to force global mean radiative forcing (RF) in the "warmer" scenario to follow that in the "cooler" scenario, and one in which stratospheric sulfates are emitted to achieve the same objective. To do so, they use the fully-coupled global chemistry-climate model WACCM, which includes an interactive treatment of wildfires. They evaluate wildfire outcomes principally through the metrics of total burned area and total carbon emissions, finding that - by both metrics - geoengineering "overcompensates" for the increase expected under SSP5-8.5, bringing total wildfires below the amount projected by SSP2-4.5. They also note some smaller differences in outcomes between the two geoengineering scenarios, and that there is some nonuniformity in the spatial distribution of outcomes.

The question asked by the authors is interesting, important, and timely, and the methods applied are appropriate. WACCM is a world-renowned climate model and the scenarios chosen are both well described and reasonably well known, making them more relevant to the community at large. The paper is incremental in nature, as it does not produce any astonishing new findings, but the content is nonetheless novel. The conclusions drawn are also supported by the data produced.

With this in mind, I have no major concerns regarding the content or framing of this paper. I have listed below some minor tehcnical and presentation concerns. Once these are addressed, I believe the manuscript will be appropriate for publication in ACP.
Response: Thank you for the comments and suggestions. Please see below for our response to the specific comments.

Minor technical issues

1. The authors appropriately caveat that the estimates of geoengineering's impacts are calculated based on small, variably-sized ensembles. In one of the central figures of the paper (Figure 6), correlations are presented between each "fire" variable (burned area and fire carbon emissions) and several "driver" variables (e.g. surface temperature). Values which are not significant at the $p \leq 0.1$ level are not shown. However, it is not clear how significance is evaluated. Similarly, I am concerned by the differing numbers of simulations between all of the various scenarios. Both are important to this manuscript, and I recommend that the details be not only listed but also discussed in the manuscript main text.
Response: Thank you. All the calculations in the revised manuscript are now based on ensemble mean values regardless of the ensemble spread. Significance of the correlation coefficients is therefore calculated and evaluated based on the ensemble mean values. We added the following statement in Section 2.4 for clarification:
"To be consistent, for scenarios with multiple simulations, only ensemble means are shown and analyzed. I.e., ensemble means are calculated before any analyses or calculations, and hence a scenario with multiple simulations is treated in the same way as a scenario with only one simulation by only using the mean value of the ensemble members."

2. With regards to the question of the number of ensemble members used, I would recommend that the authors perform an analysis where a consistent ensemble size is used. Currently the manuscript appears to be biased by this discrepancy; for example, in Figure 2, the full range of burned areas between ensemble members is shown. However, smaller ensemble sizes will generally result in smaller ranges. Similarly, it would be useful to know how these ensemble members were initialized (e.g., in the geoengineering cases where there are 2 ensemble members, were they initialized with the same initial conditions as 2 of the 5 members from SSP5-8.5?).

Response: We agree with the reviewer that different ensemble sizes may result in biases in the ensemble ranges. I.e., a scenario with larger ensemble spread may be due to larger ensemble size rather than larger variabilities. Section 3 (Future trends of fires) shows ensemble spread, which is subject to this issue. In the contrast, in Section 4 (Mechanism of geoengineering impacting fires) we only use the ensemble mean values and the analyses does not involve ensemble spread therefore different ensemble sizes are unlikely to be an issue. To be consistent among scenarios and avoid differences in the range introduced by differences in the ensemble sizes, we changed the Section 3 and corresponding figures and tables (Tables 2 and 3) to only show ensemble mean values rather than spread. We added the following clarification in Section 2.4 where we describe the simulations:

"Different ensemble sizes could result in differences in ensemble spread. To be consistent, for scenarios with multiple simulations, only ensemble means are shown and analyzed. I.e., ensemble means are calculated before any analyses or calculations, and hence a scenario with multiple simulations is treat in the same way as a scenario with only one simulation by only using the mean value of the ensemble members."

We also added the following statement in Section 2.4 to describe the initialization of the ensemble members:

"The future projection simulations analyzed in this study were initialized with the ensemble WACCM6 historical simulations. Therefore, the initial conditions of different ensemble members are different."

3. On lines 209 to 216, the authors analyze how the estimated total burned area differs between SSPs. One specific claim stands out to me. The authors state that the changes under SSPs 8.5 (~20%) and 7.0 (~10%) are the largest, whereas those in the others scenarios are "relatively small". I would first request that qualitative claims such as this one be replaced with quantitative statements where possible, as I am unsure what "relatively small" means. However in this specific case I am also concerned that this conclusion may be driven more by the form of the evaluation than by a meaningful difference. Visually, it is unclear from Figure 2 that there is any significant difference between the trajectories taken by SSP1-2.6 and SSP3-7.0; however, these are also the two cases that had just one ensemble member. Providing simple quantitative statements (e.g. "increases of less than 4%" is better than "relatively small") would help, but better yet - as discussed above - would be a quantitative discussion of how confident the authors can be when comparing results from a single ensemble member to multi-member averages.

Response: Thank you. We have revised lines 209 to 216 to be more specific and quantitative. The revised Section 3.1 is as follows:

"The global total wildfire burned area in these simulations is projected to increase under all the SSP scenarios (Figure 1a). The largest increases (averages for the 2091-2100 period relative to the 2021-2030 period) in the global burned area are seen in the SSP5-8.5 scenarios (~20%). The changes in SSP1-2.6 and SSP2-4.5 are less than 4% (see Table S2 for projected regional and global change of burned area and fire carbon emissions in 2091-2100 relative to 2021-2030 (%) under different scenarios). In terms of the spatial distribution, the 40°N–70°N latitude is the only latitude band in which the burned area consistently increases under all the SSP scenarios (Figure 1b). In the 10°S–5°N latitude band (tropical region), the burned area consistently decreases under all scenarios to a diverse extent. While global total burned area is expected to increase under most global warming scenarios, burned area may decrease in some regions due to changes in anthropogenic activities or reduced 2-m relative humidity and/or reduced soil moisture. A more detailed discussion on future trends of fire activity under the SSP scenarios are provided in the Supplement."

We also added the following discussion on the potential uncertainties to Section 2.4:

"Comparing results from a single simulation to multi-member averages could introduce potential uncertainties as ensemble mean values are in general different from values from a single member. However, the analyses and comparisons here are as useful as comparing single simulations, if not more, because in our approach we attempted to improve model projection for several scenarios by using ensemble means to replace single simulation values when possible."

Minor presentation issues

4. In several figures the color scales used are either confusing or misleading. The two most problematic examples are Figures 5 and 6. In Figure 5, the same color scale is used for 5 different quantities. Worse, for 4 of the quantities, "yellow" (i.e. the central value) is zero; however, for temperature, dark red is zero. This discrepancy is visually confusing. I would suggest using, at the very minimum, a different, one-sided color scale for the temperature change.

Response: Thank you for pointing this out. We have revised color scales in Figures 5 and 7 to make it clear. Please see below.

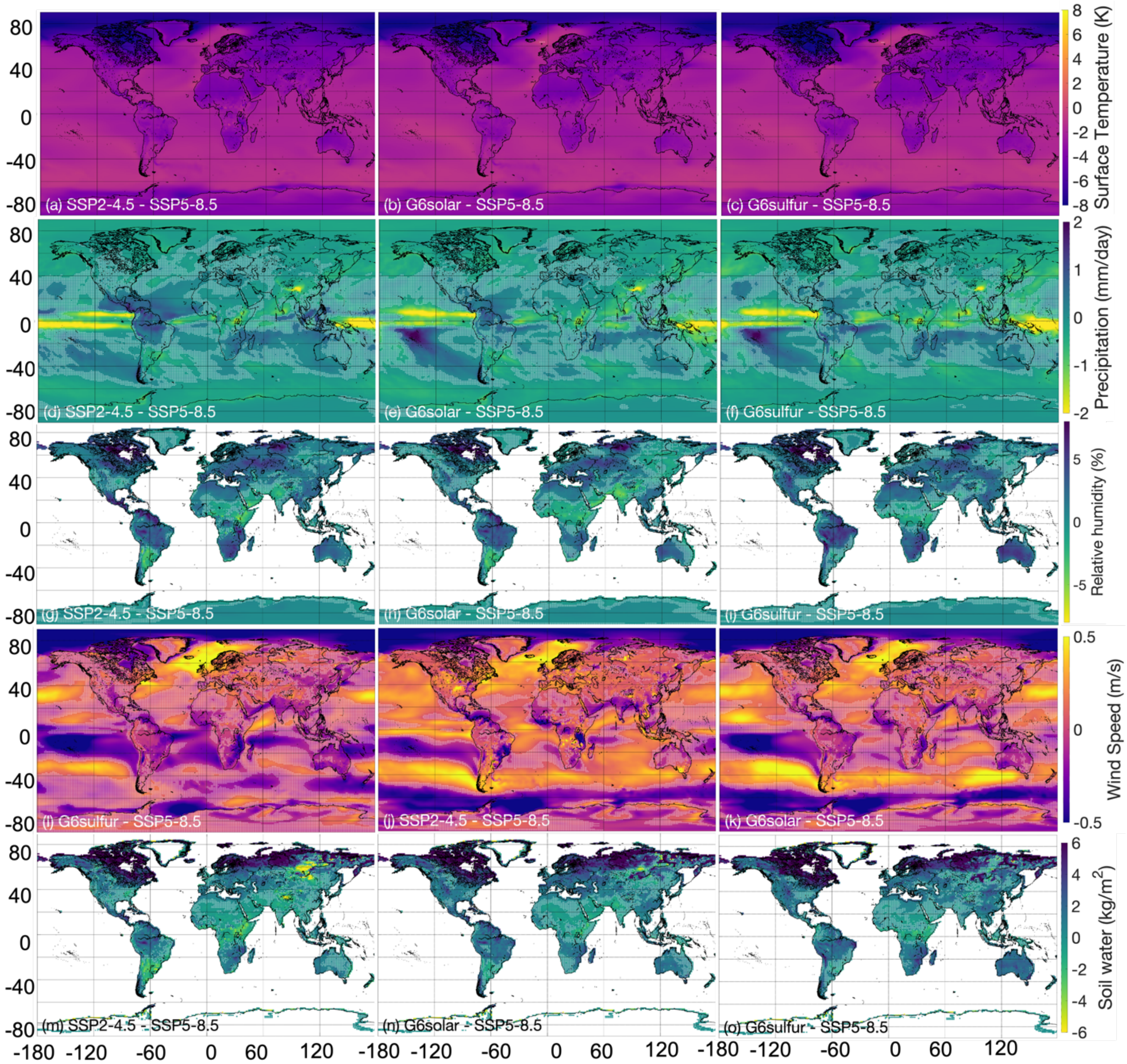

Figure 5. The difference in surface temperature (K) of (a) SSP2-4.5 from SSP5-8.5 (b) G6Solar from SSP5-8.5, (c) G6Sulfur from SSP5-8.5 averaged for 2091-2100. (d-f) are the same as (a-c) but for precipitation (mm/day). (g-i) are the same as (a-c) but for 2-meter relative humidity (%). (j-l) are the same as (a-c) but for 10-meter wind speed (m/s). (m-o) are the same as (a-c) but for soil water content at top 10 cm (kg/m^2). The grids where SSP2-4.5, G6Sulfur, or G6Solar is not significantly different from SSP5-8.5 is marked with white shade. Taking precipitation of SSP2-4.5 as an example, the significance for each model grid is calculated by student t-test (p value is 0.1) using 10 years of SSP2-4.5 precipitation data during 2091-2100 (10 data points) and 10 years of SSP5-8.5 precipitation data during 2091-2100 (10 data points).

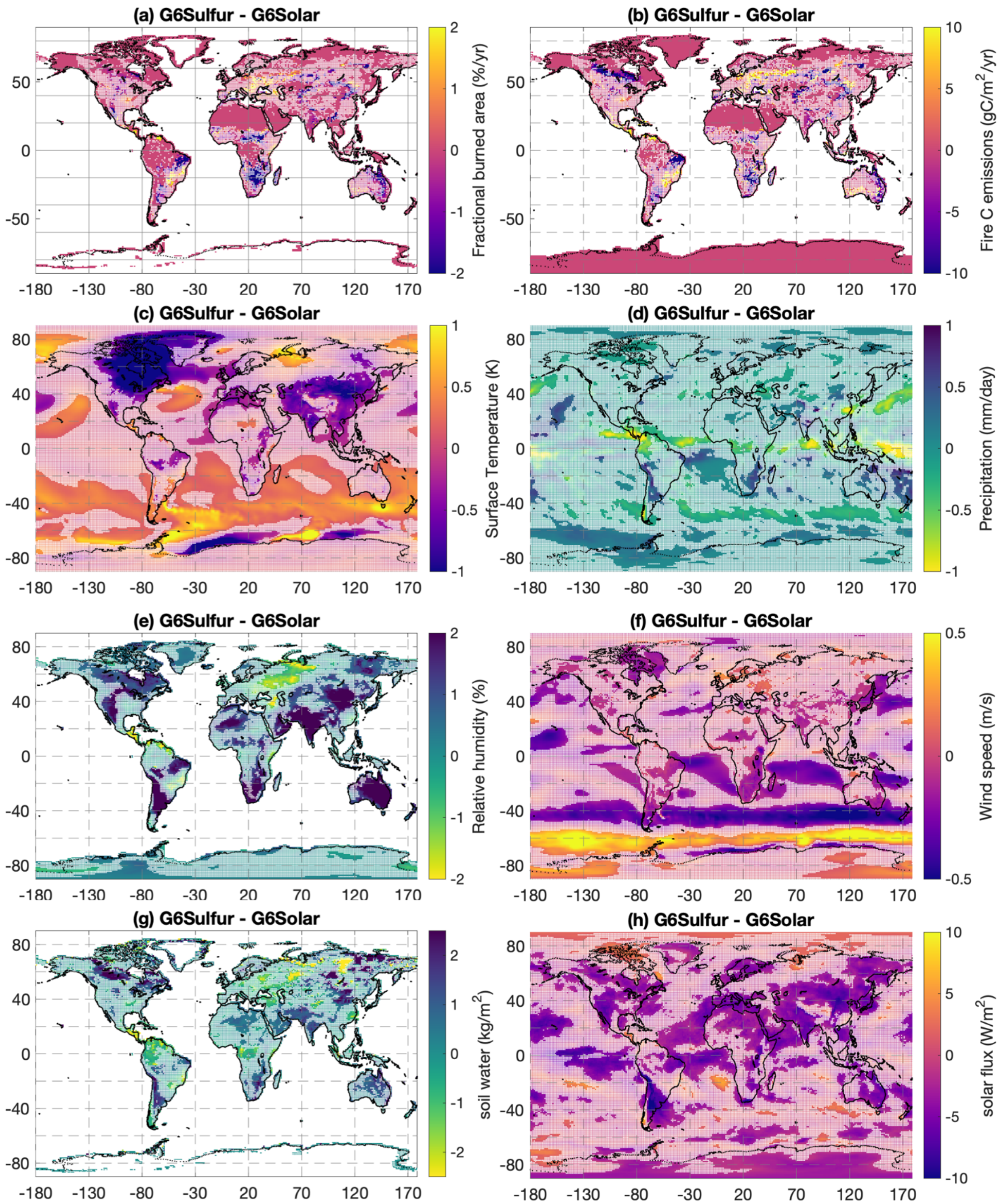

Figure 7. The difference between G6Sulfur and G6Solar in (a) burned area fraction (BA; %/yr), (b) fire carbon emissions (Cemis; gC/m^2/yr), (c) surface temperature (TS; K), (d) precipitation (Precip; mm/day), (e) 2-meter relative humidity (RH; %), (f) 10-meter wind speed (U10; m/s), (g) soil water content at top 10 cm (Soilwater; kg/m^2), and (h) downwelling solar flux at the surface (FSDS; W/m^2) averaged for 2091-2100. The grids where SSP2-4.5, G6Sulfur, or G6Solar is not significantly different from SSP5-8.5 is marked with white shade. Taking precipitation of SSP2-

4.5 as an example, the significance for each model grid is calculated by student t-test (p value is 0.1) using 10 years of SSP2-4.5 precipitation data during 2091-2100 (10 data points) and 10 years of SSP5-8.5 precipitation data during 2091-2100 (10 data points).

5. Similarly, the color scale used in Figure 6 is visually misleading. The extremes of the scale not only change from panel to panel, but the darkest shade changes from being a positive correlation of ~+0.5 (.e.g. 6a) to a negative correlation of -0.6 (6i). The color white also changes - while usually being 0, sometimes it is not (e.g. 6g, 6h). A consistent, two-sided color scale - for example that used in Figure 5 - running from -1 to +1 (or perhaps -0.6 to +0.6) would greatly help comprehension. Furthermore it is very difficult to read some of the values for the darkest backgrounds (see e.g. the EQAS/SSP2-4.5 value in 6i). Strangely the presentation in Figure S5 is much improved, although a standing issue with this manuscript is that a color scale with both red and green present is likely to cause issues for those with colorblindness.

Response: Thank you for pointing this out. We have revised color scales in Figures 6 (please see below) to make it consistent. And we have also revised all other Figures to make them colorblind friendly. The figures either use colorblind-friendly color scheme (e.g., Figures 1 and 2) or included information other than color for demonstration (e.g., numbers in Figure 6 and shapes in Figure 8).

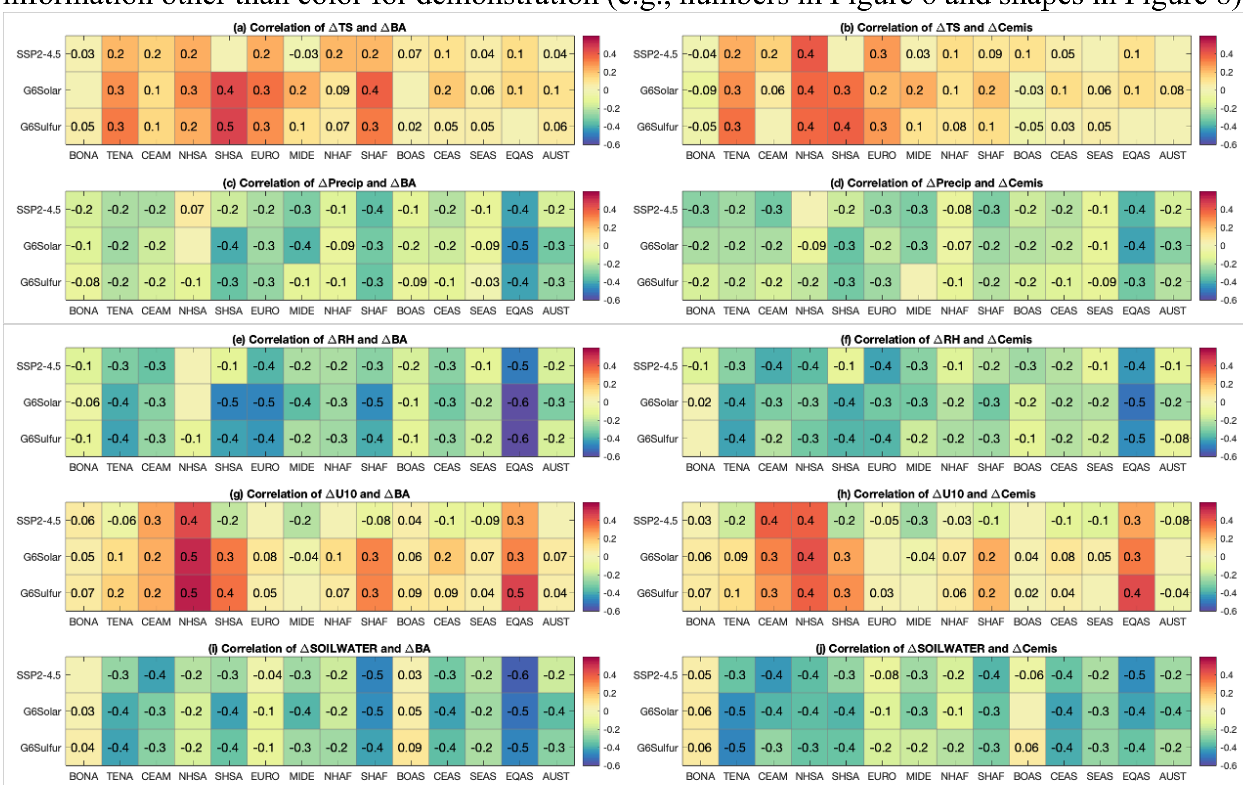

Figure 6. Correlations of (a) surface temperature change (ΔTS) and burned area change for SSP2-4.5, G6Solar, and G6Sulfur, and (b) ΔTS and fire carbon emission change (ΔCemis) for SSP2-4.5, G6Solar, and G6Sulfur. Only correlations that are significant are labeled (p value <= 0.1). For SSP2-4.5, ΔTS is calculated for individual model grids within the region and annual values. It is defined as TS of SSP2-4.5 minus TS of SSP5-8.5 (the reference case). For G6Solar and G6Sulfur, ΔTS is defined in the same way as SSP2-4.5. ΔBA and ΔCemis are defined in the same way as ΔTS. (c-d) are the same as (a-b) but for precipitation change (ΔPrecip). (e-f) are the same as (a-b)

but for relative humidity change (ΔRH). (g-h) are the same as (a-b) but for 10-meter wind speed change (ΔU10). (i-j) are the same as (a-b) but for the change in soil water content at top 10 cm (ΔSOILWATER). Correlations are calculated for 14 fire regions (x-axis), following Giglio et al. (2010), namely Boreal North America (BONA), Temperate North America (TENA), Central America (CEAM), Northern Hemisphere South America (NHSA), Southern Hemisphere South America (SHSA), Europe (EURO), Middle East (MIDE), Northern Hemisphere Africa (NHAF), Southern Hemisphere Africa (SHAF), Boreal Asia (BOAS), Central Asia (CEAS), Southeast Asia (SEAS), Equatorial Asia (EQAS), and Australia and New Zealand (AUST). The definition of the regions can be found in Figure S3.

6. Why is the p-value of significance different between Figure 6 and Figure S5?

Response: Thank you for pointing this out. We updated Figure S5 so that its p value is consistent with Figure 6. Please see the updated figure below:

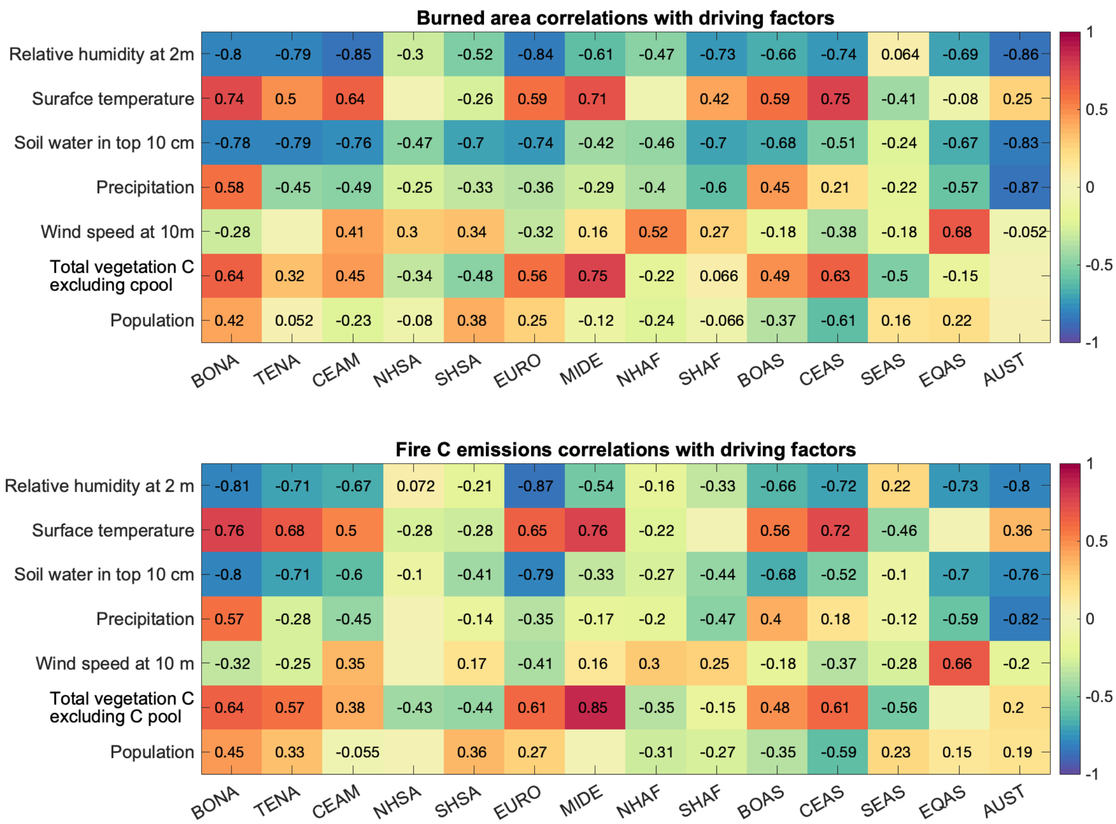

Figure S5. Correlations of burned area, fire carbon emissions with the driving factors (surface temperature, precipitation, relative humidity at 2 m, wind speed at 10 m, total vegetation carbon excluding carbon pool, and population) over 14 regions. The 14 regions are BONA (Boreal North America), TENA (Temperate North America), CEAM (Central America), NHSA (Northern Hemisphere South America), SHSA (Southern Hemisphere South America), EURO (Europe), MIDE (Middle East), NHAF (Northern Hemisphere Africa), SHAF (Southern Hemisphere Africa), BOAS (Boreal Asia), CEAS (Central Asia), SEAS (Southeast Asia), EQAS (Equatorial Asia), and AUST (Australia and New Zealand). The values of the correlations are labeled in the figure unless the correlation is not significant (P value > 0.1). The correlations are calculated based on the annual mean values of the variables, and all simulations are included in the calculation regardless of their scenarios.

7. The caption for Figure 8 mentions that the shaded region is relevant to a comparison between delta-BA and delta-C emis, but that does not seem correct based on the axes. I suspect this was incorrectly copied from the caption for Figure 9?

Response: Thank you for pointing this out. There were typos in captions of both Figure 8 and 9. We have corrected them.

8. Lines 495-504: This phrasing is incorrect. The comparison being made is between strengths of correlation, not magnitudes of effects. Saying that "impacts of the shown variables (...) on burned area are in general stronger than their impacts on fire carbon emissions" implies that a conclusion is being drawn regarding the size of the impacts, not their degree of correlation with other variables. I would suggest rephrasing for clarity.

Response: We revised lines 495-504. Below is the revised paragraph:

"For G6Solar and G6Sulfur, the correlations of the shown variables (especially for ΔTS, ΔRH, ΔU10, and ΔFSDS) with burned area are in general stronger than their correlations with fire carbon emissions (as shown by more data points that fall into the shaded area). This is expected because these variables directly impact burned area, whereas fire carbon emissions are determined by both burned area and fuel availability. Fuel availability is further directly or indirectly impacted by many variables including but not limited to the shown ones here. Therefore, the correlations between the shown variables with fire carbon emissions are not as strong as their correlations with burned area. The patterns in G6Solar and G6Sulfur and closer to each other when using SSP2-4.5 as a reference (Figures 6). This is not only because their approaches to reducing forcing from SSP5-8.5 to 4.5 W/m^2 are different, but also because the scenario configuration of SSP2-4.5 is different from SSP5-8.5 and SSP5-8.5-based G6Solar and G6Sulfur (e.g., LULCC)."