We would like to thank both reviewers for their time and effort in reviewing the manuscript. The comments have been extremely helpful and we believe that the comments have significantly improved the manuscript. At the reviewers' request, we have re-ordered the sections to make a more compelling narrative. Because of this re-ordering there are some substantial changes to the text that are difficult to capture using the standard 'track-changes' algorithm. We therefore provide an additional 'clean' document where the changes are incorporated into the text.

We address each of the specific comments in the red text below:-

Reviewer Comment 1 (submitted on 11 Sept)

This manuscript investigates the methodology for incorporating biomass burning emissions within the UK Earth System Model due to their importance on radiation, clouds, and climate. A series of modeling experiments investigated the use of no scaling factor, a doubling of emissions, and a scaling based on the dry matter consumption. AOD from these simulations were compared to observations from AERONET and VIIRS. There are a few concerns with the present version that would warrant major revisions prior to publication.

Diurnal cycles can be present in emissions and the AOD, which were neglected in the comparisons between the model simulations and the observations. A daily mean was used for the model, but VIIRS will only have a single observation per day and AERONET will only be representative of cloud-free daytime observations. Better care needs to be taken to ensure these are proper comparisons.

Thank you for raising these points about the potential diurnal cycle of fire activity. This has indeed been noted for certain fire regimes and depend on a range of factors including the timing of ignitions and diurnal variations in meteorology. As also noted, the observations from MODIS, VIIRS and AERONET to detect fire activity and AOD are available only at discrete times when overpasses or retrievals are available. As such there is an absence of information about the diurnal cycle of emissions to drive the model and the daily mean data from GFED4.1s is spread through the day and interpolated smoothly from one day to the next. Bearing this in mind, and the context of the study, which is to explore ways of improving emission scalings in a coarse-resolution global Earth System model, we do not attempt hour-by-hour temporal co-locations with VIIRS. Since the evolution of individual fires and fine-scale structure of smoke plumes cannot be resolved, we build statistics from daily mean AODs that coarse-grid models should be able to reasonably capture, on average, for regions and time periods characterised by prolific fire activity and broad regional-continental scale smoke plumes.

It could also be made clearer as to why the authors would recommend using the FIRE_DM approach when the case studies demonstrate that it is not necessarily any better. Finally, the manuscript would benefit from better organization and higher quality figures. More detailed comments are below.

Whilst no modelling approaches are ever perfect, the FIRE_DM simulation outperforms the standard (FIRE_STAN) model configuration by moderating emissions in regions and time periods when extreme fire activity occur, yet it does so without forcing a detrimental reduction of AOD in regions such as Africa and Southern Amazonia where an approximate doubling is still required (as in FIRE_STAN). This is demonstrated in the mean bias error (MBE) between model and observations (0.07 for FIE_DM, -0.08 for FIRE_1X, and 0.93 for FIRE_STAN). We have reorganised the discussion of results in section 3 and clarified many sections of the text to better guide the reader through the evidence indicating where, how and why the FIRE_DM approach improves the simulation of BBA in the model. Plots have also been improved with thicker lines and larger font sizes.

Responses to more detailed comments:

-Line 91: Should the California fires be included with in this sentence? Either way, the sentence is missing “and”.

Amended

-Line 111: I do not see how you can distinguish radiative forcing based on the imagery, though I would honestly recommend removing figure 2 and its associated text.

We consider Figure 2 extremely useful in framing the study as it provides the reader with a clear visual image of the scale and evolution of the smoke plume. We have already referenced the study by Keil and Haywood (2003), that clearly shows that smoke aerosol above highly reflective cloud leads to a positive radiative forcing, but a negative radiative forcing when over cloud-free regions of the ocean. To satisfy the reviewer, we have amended the text and added a further, rather more up to date, reference (Peers et al., 2018) that clearly shows that absorbing smoke aerosol will have positive radiative forcing over cloud.

-Line 150: If there are two years of model data, why is only one year shown in the results? Even if there are no extreme fires, wouldn’t the additional data be beneficial for evaluating the weaker fires?

The focus of the paper is on extreme wildfires; the title states that this is the focus of the paper. We state the rationale for examining 2020 in detail earlier on in the introduction

“The year 2020 was an exceptional year for extreme fire events, with the Australian Black Summer fires extending into January 2020, the northeast Siberian fires taking place throughout June to October, and the Californian wildfires of August-September 2020 (Nolan et al., 2022)”

We ran simulations from 2019 mainly to capture the extreme wildfires of S.E. Australia that occurred in December 2019-January 2020. We now clarify our approach in the revised text:-

“We generally restrict our analysis to 2020 which, as we have noted, was an exceptional year for extreme fires (Nolan et al., 2022), but extend our analysis to November & December 2019 to allow for analysis of the extreme wildfires in S.E. Australia (e.g. Damany-Pearce et al., 2022).”

Damany-Pearce, L., B.T. Johnson, A.F. Wells, M.J. Osborne, J. Allan, C. Belcher, J.M. Haywood, Australian Wildfires cause the largest stratospheric warming since Pinatubo and extends the lifetime of the Antarctic ozone hole, Scientific Reports, 12:12665 https://doi.org/10.1038/s41598-022-15794-3, 2022.

-Line 159: change to “a month of spin up”….”for the period of 2019 through the end of 2020”

Done

-Line 161: reanalysis should be singular

Done

-Line 185: Are biomass burning emissions included for other species in the model (like SO2 and NH3) that can contribute to the total AOD? If so, are they scaled in the same manner as carbon?

In our model, biomass burning aerosol emissions in GLOMAP-mode are represented by OC and BC components. The rationale behind this is that global coupled models that are used for centennial simulations in e.g. CMIP6/7 require the number of tracers to be minimised for computational efficiency. OC (and associated elements comprising organic matter) make up the majority of the scattering components and BC makes up the principal absorbing component of biomass burning smoke in both tropical/sub-tropical (e.g. Wu et al., 2020), and boreal fires (e.g. Sarnio et al., 2010). We now explicitly state this assumption:-

"In common with the majority of GCMs that simulate biomass burning smoke, emissions of inorganic species are not included on the basis that observations of the chemical composition of smoke aerosols indicate that the refractory component is dominated by OC, while the aerosol absorption is predominantly from BC for both sub-tropical (e.g. Wu et al., 2020), and boreal fires (e.g. Sarnio et al., 2010)."

Saarnio, K., Aurela, M., Timonen, H., Saarikoski, S., Teinilä, K., Mäkelä, T., Sofiev, M., Koskinen, J., Aalto, P.P., Kulmala, M. and Kukkonen, J., 2010. Chemical composition of fine particles in fresh smoke plumes from boreal wild-land fires in Europe. *Science of the Total Environment*, *408*(12), pp.2527-2542.

Wu, H., Taylor, J. W., Szpek, K., Langridge, J. M., Williams, P. I., Flynn, M., Allan, J. D., Abel, S. J., Pitt, J., Cotterell, M. I., Fox, C., Davies, N. W., Haywood, J., and Coe, H.: Vertical variability of the properties of highly aged biomass burning aerosol transported over the southeast Atlantic during CLARIFY-2017, Atmos. Chem. Phys., 20, 12697–12719, https://doi.org/10.5194/acp-20-12697-2020, 2020.

-Figure 1/Section 1.3: Consider reducing the length of the text and combining the figure into figure 3, making figure 3 have three panels. These sections are related and the reader would then have already the description for GFED.

As requested by the reviewer, text in section 1.3 has been edited to make points more concise. We have rearranged the order of the figures, so combining the figures no longer makes sense.

-Section 3.2: I am a little confused about the emphasis on the California wildfire event in the section title if the figures show all of the US for the entire month of September 2020.

This now refers to our new section 3.4. Figure 2 shows that, after a few days, the emissions in California are advected eastwards by prevailing westerly winds over the continental USA. This is where high quality measurements of the AOD are available from AERONET. It is therefore logical to expand the analysis region from the source region to the continental scale to encompass analysis of the advected smoke and enable analysis using AERONET.

-Figure 4: Do any of the yellow dots have an inverse gradient that are well below 1 such that it would be better to scale down? And likewise for the purple dots for scaling beyond 2?

We have extended the contour range of this figure to show the full range of gradients, which illustrates that the majority of sites within the plume have an inverse gradient of approximately 1, with a small minority below this, including NASA_Ames (site b). There are also a number of sites with an inverse gradient greater than 2.

-Line 318: observation should be plural

Fixed, and changed all similar cases of ‘observation’ to ‘observations’

-Section 3.2.1: What is the temporal resolution of the model output? The reason I ask is because there is a potential application of imposing a diurnal cycle on the daily emissions, in addition to expanding the number of data points for the statistics as a one-month sample is rather short.

As noted above, diurnal cycles may occur e.g. burning of agricultural waste which are anthropogenically initiated and occur predominantly during daylight hours, and wildfires that are essentially uncontrolled. However, given the limited sampling frequency of observations on active fires, the model was not configured to include or impose a diurnal cycle of emissions but was driven by the daily mean emissions available from GFED. Given the lack of diurnal variation of these model inputs, model outputs such as AOD were also produced as daily means. We now note this in the text describing how emissions were implemented:

“No diurnal cycle is assumed when implementing these in our model.”

The focus on a single month or single case study, aids physical interpretation of the model performance, whereas other sections expand to multiple regions and longer periods to enable the more statistical assessment that is also required to evaluate large-scale models.

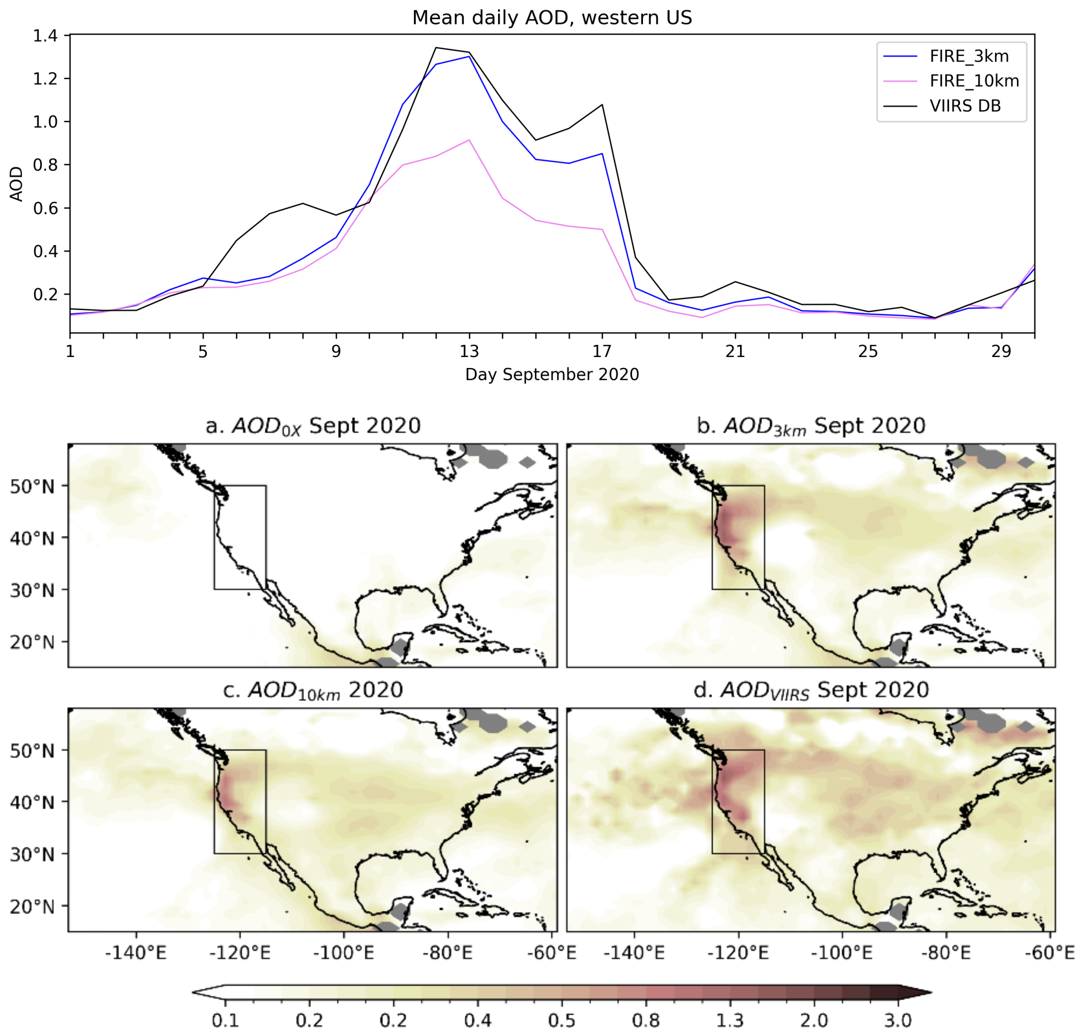
-Line 349: Is this confirmed with the satellite observations?

Yes - this is now noted in text at line 471 ‘The modelled plume appears to be largely missed here when compared to satellite observations’

-Line 352: Another possibility is vertical placement of the emissions is incorrect, which would then alter the dynamical flow.

We have explored this possibility by completing an additional model run where unscaled GFED4.1s emissions from boreal and temperate forest and deforestation fires are injected uniformly between 0 and 10 km based on lidar observations of the smoke plume (FIRE_10km), comparing this to the unscaled GFED4.1s emissions injected uniformly between 0 and 3 km, based on a simplification of AeroCom recommendations (FIRE_3km). These comparisons are in the supplementary figures (Fig. S1-S3). We found that the 10 km ejection height worsened the relationship between model and observations over the western US region (Fig S4 and S5). These simulations were done in HadGEM3-GA8, an atmosphere-only configuration of UKESM1.1, therefore there are some differences between these and the UKESM1.1 simulations. We have added this to the methods on line 168

‘An additional test simulation where unscaled GFED4.1s emissions from boreal and temperate forest and deforestation fires were injected uniformly between 0 and 10 km was run, based on lidar observations of the western US smoke plume (Winker et al., 2009). However, due to tropical Savannah emissions making up a large part of the global BBAOD this had little effect on the AOD distribution globally, and worsened the relationship between model and observations over the western US region, compared to a 3 km injection height. (See figures Fig S1 to S3 in the supplement).’



-Line 386: Does the figure show the straight up monthly mean AOD or was it subsampled for the availability of VIIRS observations. I am guessing it wasn't subsampled as the other panels have gray shading (I am assuming) for missing data. It is best to show an apples to apples comparison.

At the request of the reviewer, we have amended the figure to show subsampled modelled AOD for availability of VIIRS observations. Explanation of grey areas to represent missing data has been added.

-Figure 7: Why no panel for FIRE_1X?

FIRE_1X has now been added to figure

-Figures 8, 9, and 11: It is not appropriate to use daily mean values here. SNPP has a single snapshot each day.

The standard GFED4.1s product has daily fractions and there is no observational data about how those fire emissions were spread through the day. Higher temporal resolution model data would only be relying on assumed diurnal cycles. We have added a note in section 2.3 Observational data at line 221 'While it is not ideal to use daily snapshot values, this is the best compromise for having a high spatial and temporal availability of data.'

-Line 427: Typo in VIIRS

Fixed

-Line 510: Could also be anthropogenics

Anthropogenic aerosols added as a potential explanation

-Line 553: There is no measure of significance here.

Changed to 'modifies'

-Line 573-574: should be "due to"

Corrected

-Figures 14 and 15: Can CERES be added here to give a sense of where these simulations lie with respect to observations?

CERES would only evaluate TOA radiation flux not and would not distinguish the aerosol IRF forcing.

-Line 603: These are not novel results. If this section is going to remain in the paper, it should cite relevant references that have already shown this.

Corrected- added to line 638 'This is in agreement with previous studies (Marlier et al., 2014)'

Marlier, M. E., Voulgarakis, A., Shindell, D. T., Faluvegi, G., Henry, C. L., and Randerson, J. T.: The role of temporal evolution in modeling atmospheric emissions from tropical fires, Atmospheric Environment, 89, 158–168, https://doi.org/10.1016/j.atmosenv.2014.02.039, 2014.

-Paragraph beginning on Line 625: I do not think it was sufficiently demonstrated that this is the case, particularly because there was no measure of statistical significance showing that the FIRE_DM simulation was any better than FIRE_1X. Another important consideration for the conclusions section is to specify that this is specifically for a period that uses MODIS for fire detections. The footprint for VIIRS is smaller, meaning that smaller background fires can be detected and that these results may not be applicable for future versions of GFED. I recognize this work was likely started prior to the beta release of GFED 5, but nevertheless, the implications should be mentioned.

These caveats have been added to the conclusion:

'The results presented here are from a period that used MODIS for active fire detections and may not be applicable to other GFED versions or time periods relying on different methods to estimate emissions. For instance, the footprint for VIIRS is smaller (used for GFED5 from 2023 onwards), meaning that smaller active fires can be detected.'

-Throughout: missing spaces in W m^{-2} and g m^{-2}

Fixed

-Throughout: The quality of the figures could be improved by using bigger and bolder text and thicker lines for the line plots.

Font size and line thickness of plots has been increased

-Consider reorganizing the paper such that the global AOD is discussed first, and then you go into more detail for the other regions, ending with California.

Thank you for this suggestion. We have reordered some of the sections within the results, so that the global analysis is discussed first. We believe that this significantly improves the rationale for the work and the narrative for the paper.

Citation: https://doi.org/10.5194/egusphere-2025-3936-RC1

Reviewer Comment 2 (submitted on 27th October)

The authors present a novel emission-scaling approach based on daily dry matter (DM) burnt thresholds and evaluate it in the context of the 2020 California wildfires using the UKESM1 atmosphere-only model. The proposed daily-resolution emission framework, combined with a DM-based scaling scheme, aims to correct for under-detected or under-represented extreme fires. The study compares multiple scaling scenarios (1×, 2×, and DM-scaling), evaluates results against satellite and ground-based AOD and CO observations, and analyzes differences in instantaneous radiative forcing (IRF).

The topic is timely and relevant, and the modelling framework is well structured. However, several key aspects of the methodology, interpretation, and physical justification remain underexplained. Some assumptions are insufficiently supported, and the generality of the approach beyond the California case is not convincingly demonstrated.

Moreover, while the paper contains extensive quantitative descriptions (AOD values, correlation coefficients, forcing magnitudes), many results are presented descriptively without adequate diagnostic evidence, mechanistic interpretation, or uncertainty assessment. Consequently, the manuscript currently reads more as a compilation of results than as an evidence-driven analysis, making it difficult to assess the robustness of the findings.

I recommend that the authors

(a) emphasize the underlying physical insights rather than reiterating numerical values,

(b) include uncertainty or statistical significance estimates in key comparisons (e.g., standard deviations, RMSE, or confidence intervals), and

(c) strengthen the mechanistic reasoning linking the observed differences to relevant processes (emission strength, plume injection, transport, cloud masking).

I therefore recommend **substantial modification** before the paper can be considered for publication. Strengthening the physical reasoning behind the emission-scaling scheme, clarifying the interpretation of radiative forcing, and reducing purely descriptive content will greatly enhance the clarity, robustness, and broader scientific relevance of the work.

Thank you for these carefully considered comments and suggestions. We appreciate there was a need to reorganise and substantially edit several sections of text (particularly in section 3) to better guide the reader through the results of the study. We have made numerous changes throughout to outline why particular comparisons were made, what they showed and what physical or pragmatic assumptions the methodology has been based on.

This has included providing additional results in a supplementary section demonstrating some of the mechanistic sensitivities that underpin the injection of emissions in the model (i.e. DM-scaling thresholds and injection height assumptions). We also bolster the statistical analysis, evaluating RMS errors, biases and corelation concurrence metrics for each scaling approach (newly added Table 4) to assess how significantly the scaling alters performance in capturing observed VIIRs AOD observations across wildfire regions and how much uncertainty or spread remains.

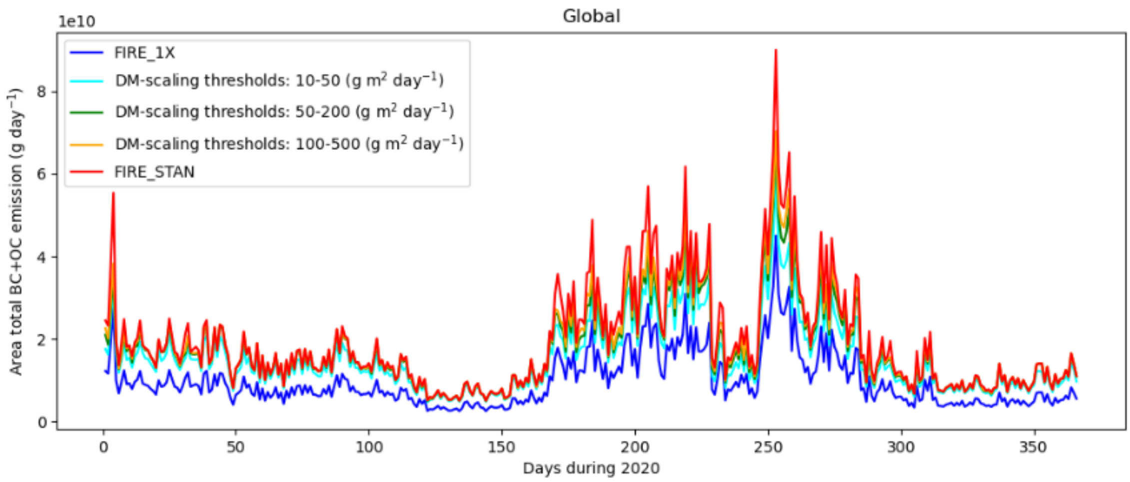
Major comments:

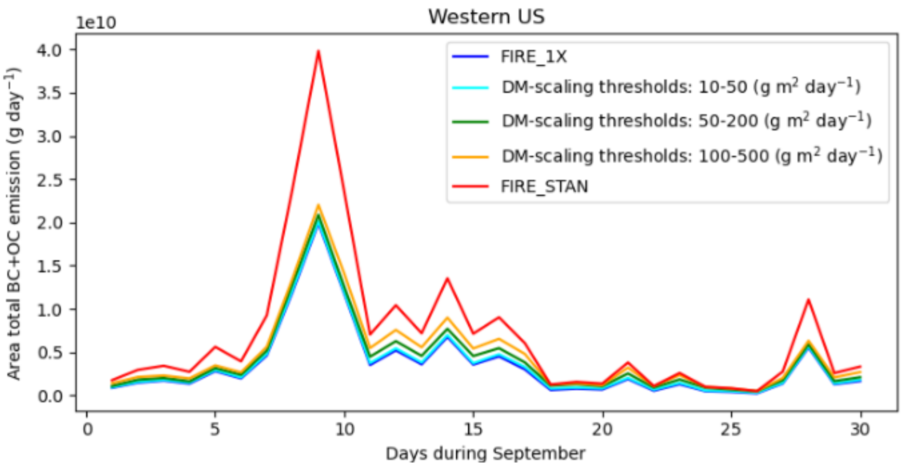
1. The 2020 California wildfires produced an unusually thick smoke layer, under which MODIS active-fire detection (used in GFED4.1s) can fail to identify ongoing fires. Please clarify whether GFED includes any correction or retrospective burned-area adjustment to account for such obscured fires. If not, this limitation should be acknowledged, as it could partly explain the need for emission-scaling and influence the interpretation of the results.

   Added to line 504: 'In addition, the unusually thick smoke layer could have obscured the detection of active fires by MODIS, leading to underestimating or missing some emissions.''

2. The threshold values for DM (< 50 g m^{-2} d^{-1} scaled by 2×; > 200 g m^{-2} d^{-1} no scaling; linear ramp in between) appear somewhat arbitrary. The authors mention testing other bounds (10–100 and 100–500), but the reasons for selecting 50 and 200 remain vague. I suggest including a sensitivity analysis or at least more transparent rationale (with reference to fire size, detectability, satellite burn-scar miss-rate literature) to support threshold choice.

   We don't have space in the main text to unpack sensitivity tests but we have included results from testing different DM thresholds in a supplementary material section. The selected 50-200 bounds were a pragmatic choice that minimized scaling emissions from the September Western US event and other extreme events (keeping the scaling close to 1, whilst nearly doubling the emissions elsewhere, particularly maintaining a scaling close to 2 for regions such as Southern Africa dominated by Savanah where this is necessary to keep the AOD in line with observations. Table S1 shows the 2020 GFED4.1s emissions with different dry matter thresholds. Table S2 shows the emissions for the month of September 2020, where the Western US emissions are 1.19 x the unscaled emissions, and Central Africa emissions are 1.963 x the unscaled emissions. This is also illustrated in figures S4 and S5, where the DM scaling with the 50-200 thresholds gives very similar global emissions to the 'standard' 2x scaling, until the northern hemisphere summer when global emissions become more intense, dominated by the extreme fires in Northeast Siberia and Western US and the difference between DM thresholds is more clear to see. In particular, in the Western US region (Fig. S2), this difference can be observed.





Added to text in line 310: 'Other lower and upper limits were tested (10–100 and 100–500) (see figures S4 and S5 in the supplement), but we found 50–200 to give the best compromise keeping the September 2020 western US fires emissions close to no scaling, and globally close to doubling the emissions.'

3. The study should discuss whether and how the DM-scaling scheme applies to different fire environments (e.g., savannah, small-patch, or boreal fires) where the relationship between dry-matter burnt and satellite detectability may differ. Clarifying this would help define the scope and limitations of the proposed approach.

   It is true that the relationship between detectability and DM could depend on region / fire environment. The philosophy to avoid ad hoc scaling based on region-specific information is to draw on a metric that correlates with the rate of fuel combustion and therefore energy release. Whilst the method will have limitations, it offers an improvement over the baseline starting point of a globally uniform scaling lacking any accommodation for how detectability varies with the nature of the fires. Different fire environments are examined (e.g. northeast Siberia (boreal), central Africa (savannah) etc.) to test the applicability of this scheme globally.

4. While the California 2020 case is compelling, especially in the title, the manuscript proposes broader applicability ("global" fire extremes) yet the evaluation across other regions is

relatively limited (five regions listed). The performance metrics (e.g., r^2 ~0.48 for western US) suggest substantial unexplained variance. I recommend adding a table summarizing performance metrics (bias, MAE, correlation) for all regions, and discussing where the method succeeds versus where it struggles. This will help readers assess transferability.

We thank the reviewer for the suggestion. Mean bias (MBE) for individual regions has been added to figure 5, highlighting that FIRE_DM is successful in regions where scale-up is needed (where standard 2x worked better than 1x), and regions where scaling is not so necessary (where 1x performed better than 2x). For the overall 2020 comparisons, a table of the key performance metrics has been added (MBE, MAE, CCC, RMSE). We have also adjusted the title of the paper to better reflect the content.

5. The purpose and main conclusion of Section 3.4 are unclear. What is the key message of this section? Are the authors aiming to demonstrate a net cooling or warming effect associated with different emission-scaling schemes? If so, this needs to be explicitly stated and quantified. Currently, the section reads largely descriptive; a more conclusive statement about whether the DM-scaling leads to systematically stronger or weaker shortwave (and potentially total) forcing compared to the 1× or 2× runs would clarify its scientific contribution.

   This now refers to section 3.5, which discusses the importance of accurately scaling emissions, as the resulting radiative impacts are diverse. The FIRE_DM has a weaker shortwave forcing than the standard simulation, which we have clarified in the text. We have also reordered the paragraph to emphasise the purpose of the section.

   'Accurately modelling the AOD is essential to quantify the climate impact of wildfires, as the emissions scaling factor makes a large difference to the global annual mean radiation budget. We estimate that BBA emissions in the current 'standard' simulation (FIRE_STAN) lead to a -0.338 W m^{-2} change in global mean clear-sky shortwave radiation at the top-of-atmosphere (TOA), i.e. increased reflection, whereas the FIRE_DM simulation results in a weaker shortwave effect of -0.251 W m^{-2}..'

6. Throughout Sections 3.2–3.5, the manuscript presents numerous quantitative values (e.g., AOD differences, correlation coefficients, radiative forcing magnitudes) in a largely descriptive manner, without providing sufficient diagnostic or statistical evidence to support the inferred conclusions. For instance, several paragraphs simply report value ranges ("the AOD increases by x%," "the forcing reaches –2 W m^{-2}," etc.) without explaining the underlying processes, assessing statistical significance, or connecting the numerical differences to physical interpretation.

   The text has been reworked in these sections to place a greater emphasis on the physical interpretations, and superfluous quantitative values which do not aid the conclusions have been removed.

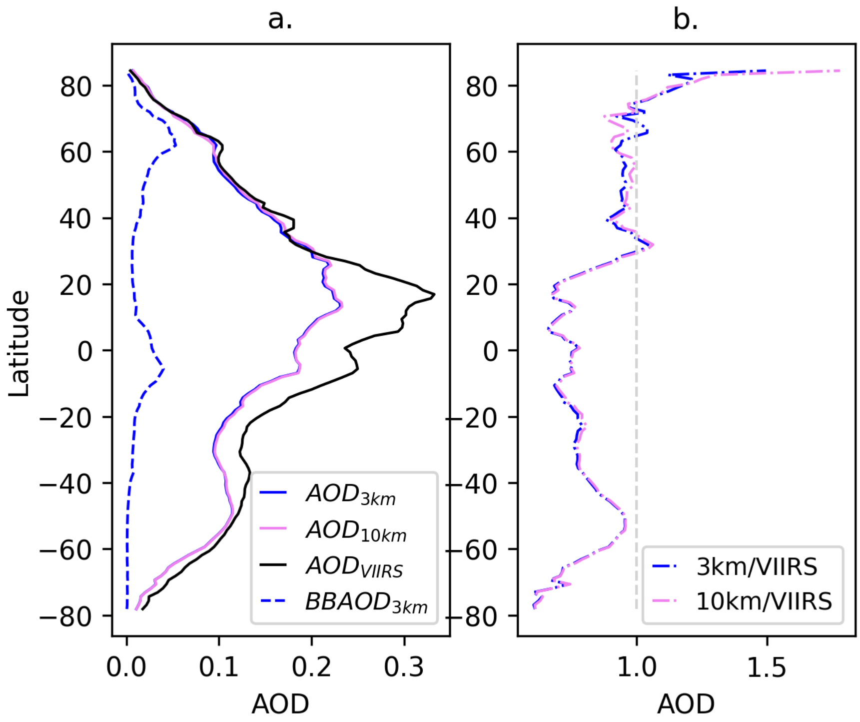
7. The model uses prescribed vertical injection rather than explicit plume-rise modeling. For extreme fires, injection height critically impacts aerosol dispersion, lifetime, cloud interactions, and radiative forcing. While the manuscript does not discuss this impact, especially the relative contribution compared to emission scaling. The manuscript would

benefit from a sensitivity discussion or diagnostic showing how injection height uncertainties might affect AOD, transport and forcing in the extreme-fire context. At minimum, this limitation needs to be emphasized in the conclusions with implications for interpretation.

We have explored how the vertical placement of emissions could affect our results by completing an additional model run (FIRE_10km) where unscaled GFED4.1s emissions from boreal and temperate forest and deforestation fires were injected uniformly through a much deeper layer extending from 0 and 10 km. This represents a somewhat upper bound to the likely plume rise and vertical mixing based on comparing model runs with lidar observations of the smoke plume in the Western US during 2020. The standard set up (here labelled FIRE_3km) injects smoke from forest fires to between 0 and 3 km, based on a simplification of AeroCom recommendations. In both cases emission from non-forest sources remained as surface level emission. These comparisons are in the supplementary figures (Fig. S1 – S3). These simulations were completed in HadGEM3-GA8, a configuration with less complex atmospheric chemistry than UKESM1.1 and therefore results in FIRE_3km are not identical to the equivalent 1x simulations elsewhere in the paper.

We found that raising the maximum injection height (from forest fires) to 10km made little difference to the AOD distribution globally (Fig. S1) and worsened the relationship between model and observations over the western us region (Fig S2 and S3). We have added this to the methods on line 168.

'An additional test simulation where unscaled GFED4.1s emissions were injected uniformly between 0 and 10 km was run, based on lidar observations of the western US smoke plume (Winker et al., 2009). However, due to tropical Savannah emissions making up a large part of the global BBAOD this had little effect on the AOD distribution globally, and worsened the relationship between model and observations over the western US region, compared to a 3 km injection height. (See figures Fig S1 to S3 in the supplement).'



8. The atmosphere-only modeling framework (fixed SST/sea-ice) is a limitation for assessing climate-feedbacks; this should be clearly noted and discussed in the implications.

   An additional caveat has been added to the end of section 3.6: 'To fully assess the climate-feedbacks of these extreme fire events, a full version of UKESM1.1 should be used, with interactive ocean and terrestrial biosphere components.'

Specific comments:

1. **Abstract:** Quantify phrases such as "little overall bias" in AOD to provide a clearer, evidence-based summary.
   We have added values of mean bias error to the abstract to quantify these statements: 'Running with daily emissions from Global Fire Emission Database v.4.1s (GFED4.1s) enables a realistic simulation of the thick smoke plumes from the Californian fires and large boreal fires more generally, with little overall bias (-0.08) in aerosol optical depths (AODs) between UKESM and collocated observations (AERONET, VIIRS) compared to the standard '2x' simulation (0.93). Modelled AODs were biased low across other regions (e.g. savannah, mean bias = -0.48) dominated by fires with lower fuel consumption, unless emissions were scaled up by a factor of 2 (mean bias = -0.15).'

2. **Section 1.3:** Define "extreme" or "megafire" explicitly (e.g., > 10 000 ha or > 100 000 ha) and discuss how definitions vary across ecosystems.

   This was defined in section 1.3 (line 80) as > 10 000 ha

3. **Figures 5 and 7 (AERONET vs VIIRS):** At the Catalina site, all three experiments miss high daily AOD (> 1), yet the monthly bias map shows near-zero bias. This discrepancy could be a color-scaling artifact (~0.4 intervals). Including VIIRS AOD in the Figure 5 time series and adding RMSE/mean-bias metrics to the scatter plots would help clarify consistency between datasets.

We have edited figure 7 (now figure 10) to have finer AOD intervals, and added the VIIRS AOD and RMSE to the timeseries in figure 5 (now figure 8).

4. **Figure 5:** Explicitly discuss sites with poor model–observation agreement (e.g., negative r^2) rather than highlighting only successes.

   The figure has now been changed to show all available sites, including those with a negative r^2, suggesting a poor model-observation agreement. In addition, we have extended the contour range of this figure to show the full range of gradients. In figure 8, the RMS error has been added to the plots to show the range of successes in the AERONET case studies.

5. **Figure 6:** Caption text for panel (c) appears incorrect and should be checked.

   Done

6. **Figure 7:** Consider adding an AOD_1× – AOD_VIIRS difference panel for consistency with Figure 6.

   Done

7. **Figure 8:** Clarify why the FIRE_1× run performs best and what physical explanation underlies this result.
   We are arguing that the mega fires are detectable and so 1x does best, although DM is very close to reproducing 1x for the peak AOD. Added to the text: 'The physical explanation for this may be that the larger fires contributing to the peaks in AOD are more detectable from MODIS retrievals, and therefore do not need to be scaled by 2x as in FIRE_STAN. '

8. **Figure 13:** The description of surface versus atmospheric components is inaccurate and should be revised.

   We have provided a fuller description to the figure caption: surface forcing '(reduction of net SW clear-sky radiation at the surface)', and atmospheric forcing '(absorption of SW radiation in the atmosphere under clear sky conditions)'. We also make some clarifications in the text (now section 3.5) to better clarify how the IRF was calculated.

9. **Section 3.5 and Figure 14:** The terms "all-sky" and "clear-sky" are used without definition. Please clarify that "all-sky" IRF includes the effects of clouds (both cloudy and clear regions), while "clear-sky" IRF represents forcing under cloud-free conditions only. This clarification will help readers interpret why clear-sky forcing shows stronger cooling.

   Added to figure caption 14 and text in section 3.5

10. **Units and consistency:** Ensure units are consistent (e.g., g m^{-2} d^{-1} for dry-matter burnt).

    g m^{-2} d^{-1} used for dry-matter burnt

11. **Grammar and formatting:** Review for minor typographical issues (e.g., "smoke injections" rather than "smoke ejections") and maintain consistent reference formatting.

    'ejections' swapped for 'injections'

Citation: https://doi.org/10.5194/egusphere-2025-3936-RC2