We greatly appreciate your constructive suggestions and comments. Our point-by-point responses are as follows.

Major comments:

1) The Authors provide a new dataset of nine fire model estimates of carbon and 33 other gas and aerosol emissions. They provide a present day analysis of the data and show that LULCC is the largest source of uncertainty when simulating historical fire emissions. The collection of this dataset is a useful step forward in synthesizing fire modelling and one which should be of great use to the climate and Earth system science community. The Authors are to be commended on such a large effort and the manuscript will be suitable for Atmospheric Chemistry and Physics once some improvements are made to the manuscript. While the content is of great interest I find myself agreeing with the previous reviewer that the grammar is not yet at a level suitable for final publication. Unfortunately, many parts of the manuscript (mainly in the first half) were hard to follow due to this. I therefore also propose an extensive review of the text. I have included some suggestions below, but it is not an extensive list.

Reply: Thanks for your suggestions. We have done an extensive review of the text and edited the language.

2) while the methodology and presentation of results is suitable for publication, the manuscript will benefit from further analysis in three ways. The manuscript's main objective is in presenting a dataset for use by the community, and these additions are all ways to make the manuscript more useful for that potential user: Extending the multi-model SD/zonal average plot in Figure 3 for other time slices across the dataset. A small discussion on which models are outliers for different regions/times would be insightful too. As the Authors do not know what regions will be of interest to the potential user in their studies I feel that Figure 9 should be for all regions, not just the three with the most variance, even if trends are small. Furthermore, as it is likely that the potential user will first want to compare to CMIP6, the GFED regions in Fig 8 should follow the CMIP6 version in van Marle (i.e., further segregate the Americas). Similar plots for other emissions species would also be useful and can be place in the SI.

Reply: 1) We have added Figs. S1b-c to show the multi-model SD/zonal average for two additional time slices, 1700–1850 and 1900–2000, and a discussion in Sec. 4.1 accordingly as "Spatial patterns of inter-model spread of fire emissions for 1700–1850 and 1900–2000 (Figs. S1b-c) are similar to the present-day patterns as shown in Fig. 3.".

It may be unsuitable to compare the spatial patterns of SD/zonal average among different time periods in detail because 7 models are used for 1700-1850 and 9 models for 1900-2000 and 2003-2008. MC2 and CTEM do not provide simulations for 1700-1850 (Table 1), and generate lower (MC2) and higher (CTEM) historical global fire emissions than most FireMIP models for the 20th century, respectively (Fig. 6).

2) Fig. 9 has been revised and included all regions.
Also, we have briefly described them, including outliers in these regions, in Sec. 4.3 as
"In other regions, the difference in long-term changes among models is smaller (Fig. 8b). Emissions of most models and CMIP5 estimates exhibit a significant decline in temperate North America (TENA) from ~1850 to ~1970, while historical changes of CMIP6 estimates are comparatively small (Fig. 9b). LPJ-GUESS-SIMFIRE-BLAZE has a more obvious long-term change than the other FireMIP models and CMIPs in boreal North America (BONA) and northern South America (NHSA) (Figs. 9a and d). MC2 and LPJ-GUESS-GlobFIRM emissions increase after ~1900 in Europe (EURO), while emissions of other models and CMIPs are overall constant (Fig. 9f). In boreal Asia (BOAS), emissions of most models and CMIP6 are relatively constant, while LPJ-GUESS-GlobFIRM and CMIP5 emissions decline form 1850 to the 1950s and from 1900 to the 1970s, respectively, and then rise (Fig. 9j). JULES, LPJ-GUESS-SIMFIRE-BLAZE, CLM4.5, CTEM, and CMIP6 emissions significantly decline since the 1950s in Southeast Asia (SEAS), while CMIP5 emissions increase (Fig. 9l). In equatorial Asia (EQAS), CMIPs emissions increase after ~1950, but in FireMIP only CLM4.5 partly reproduces it (Fig. 9m). "

3) We used the GFED regions because they represent key fire regions across the world and are the most widely used one by the community. In addition, Figs. 10-11 in van Marle et al. (2017, paper for CMIP6 fire emissions) already compared each of FireMIP models and their medians with historical charcoal-based reconstructions (i.e. CMIP6 estimates) in four sub-regions of North America, so we did not want to repeat the same analyses here.

4) As suggested, we have added Figs. S3-5 for regional fire BC, OC, and CH_4 emissions in the supplementary material, and the words
"As shown in Figs. S3-5, long-term changes of regional fire emissions for other species are similar to those of fire CO emissions." in Sec. 4.3.

3) The present-day evaluation is of a suitable level for publication as is; however, further historical evaluation can be undertaken. In particular, the contribution of crop burning and how the fire models compare against historical fire proxies (not just the CMIP5/6 reconstructions). As crop fires are only accounted for in CLM, please discuss what this means in terms of missing estimates of historical emissions across FireMIP, a figure of % contribution to total emissions over time for example would be insightful. Included should be a discussion of current knowledge of crop fires in the present day, their uncertainties in emissions back in time, and what this means for

CMIP/FireMIP as LULCC has been shown to be the largest uncertainty here. This then links to an overall evaluation of historical emissions with proxies. The inclusion of an updated Figure similar to the one from van der Werf's 2013 paper for example? I leave it to the Authors to decide on how best to do this, but it should be included to once again help guide the potential user; perhaps in section 4.3.

Reply: 1) We have compared the historical changes of the FireMIP simulations with other widely used reconstructions in global-scale fire studies and added

"..., but in disagreement with earlier reconstructions based on charcoal records (Marlon et al., 2008; Marlon et al., 2016), ice-core CO records (Wang et al., 2010), and ice-core $\delta^{13}CH_4$ records (Ferretti et al., 2005), which exhibit a rapid increase from 1700 to roughly the 1850s."

and a new paragraph

"Earlier reconstructions based on fire proxies also show a big difference in long-term changes after the 1850s. The reconstruction based on the Global Charcoal Database version 3 (GCDv3, Marlon et al., 2016) exhibits a decline from the late 19th century to the 1920s, and then an upward trend until ~1970, followed by a drop. The reconstructions based on the GCDv1 (Marlon et al., 2008) and ice-core CO records (Wang et al., 2010) show a sharp drop since roughly the 1850s, while a steady rise is exhibited in the reconstruction based on ice-core $\delta^{13}CH_4$ records (Ferretti et al., 2005). The simulated historical changes of FireMIP models (Fig. 6) fall into this fairly broad range of long-term trends in these reconstructions." in Sec. 4.1.

We will perform a detailed regional comparison with reconstructions based on various fire proxies (including but not limited to charcoal records, and considering that recently more paleofire records are being compiled, e.g., the number of sites with charcoal records in China will be increased from 15 in GCDv3 to113) and driver analyses in the near future in cooperation with scientists who work on fire proxies.

2) We have added Fig. S2 to show the historical change of crop fire emissions in the CLM and % contribution to total emissions, and have added discussion in Sec. 5 as:

"Fire has been widely used in agricultural management during the harvesting, post-harvesting, or pre-planting periods (Korontzi et al., 2006; Magi et al., 2012). Crop fire emissions are an important source of greenhouse gases and air pollutants (Tian et al., 2016; Wu et al., 2017; Andreae, 2019). GFED4s reported that fires in croplands can contribute 5% of burned area and 6% of fire carbon emissions globally in the present day (Randerson et al., 2012; van der Werf et al., 2017). In FireMIP, only CLM4.5 simulates crop fires, whereas the other models assume no fire in croplands or treat croplands as natural grasslands. In CLM4.5, crop fires contribute 5% of global burned area in 2001-2010, similar to GFED4s estimates. However, CLM4.5 estimates a total of 260 Tg C yr^{-1} carbon emissions (contribution rate:13%), which is higher than GFED4s estimate (138 Tg C yr^{-1}) because CLM4.5 simulates higher fuel loads in croplands than the CASA model used by GFED4s. In CLM4.5, both the carbon emissions from crop fires and the contribution of crop fire emissions to the total fire emissions increase throughout the 20th century (Fig. S2), consistent with earlier estimates based on a different crop fire scheme (Ward et al., 2018). In JULES-INFERNO, an increase in cropland area also leads to an increase in burned area and fire carbon emissions because this model treats croplands as natural grasslands. Grasses dry out faster than woody vegetation and are easier to burn in model setups, so increasing cropland area leads to increasing burned area and fire carbon emissions. On the other hand, for FireMIP models that exclude croplands from burning, expansion of croplands leads to a decrease in burned area and fire carbon emissions. Therefore, different treatment of crop fires can contribute to the uncertainty in simulated fire emissions. Because four out of six FireMIP models used for generating CMIP6 estimates exclude croplands from burning (van Marle et al., 2017b), CMIP6 estimates may underestimate the impact of historical changes of crop fire emissions in some regions (e.g., China, Russia, India). Given the small extent of crop fires, high resolution remote sensing may help improve the detection of crop fires (Randerson et al., 2012; Zhang et al., 2018), which can benefit the driver analyses and modeling of historical crop fires and their emissions in DGVMs.".

Minor comments:

1) Lines 61-62. The statement 'consistent with multi source merged historical reconstructions' is in reference to CMIP5/6; however, a multi-source merged historical reconstruction of the proxy data (ice cores, charcoal, tree scars etc.) would not result in the same conclusion. Please either rephrase in terms of CMIP, add that this disagrees with proxies, or remove.

Reply: We have added **"**as input data for CMIP5 and CMIP6"

2) Line 77: Species emitted from fires

Reply: Done

3) Lines 81-89: I think this sentence needs to be clearer, both in grammar and content. Are all the items in the list symptoms of the atmospheric composition changing in response to fires? For example, changes to the 'terrestrial nutrient and carbon cycles' are more a symptom of changes to the magnitude of deposition and alteration to the land vegetation itself and the human health impacts are linked to the air quality changes (as R1 has also mentioned). Perhaps writing as a numbered list would help?

Reply: The sentence has been changed to "Second, by changing the atmospheric composition, fire emissions affect the global and regional radiation balance and climate (Ward et al., 2012; Tosca et al. 2013; Jiang et al., 2016; Grandey et al., 2016; McKendry et al., 2018; Hamilton et al., 2018; Thornhill et al., 2018).

Third, fire emissions change the terrestrial nutrient and carbon cycles through altering the deposition of nutrients (e.g., nitrogen, phosphorus), surface ozone concentration, and meteorological conditions (e.g., diffuse radiation, temperature, precipitation) (Mahowald et al., 2008; Chen et al., 2010; McKendry et al., 2018; Yue and Unger, 2018). In addition, they degrade the air quality (Val Martin et al., 2015; Knorr et al., 2017), which poses a significant risk to human health... ”

4) Line 90: There have been observation campaigns, such as SAMMBA, which have attempted to observe aerosol from fires at the regional scale using a combination of ground based and aircraft measurements. While they are only snap shots, due to the inherent time limitations of campaigns (as compared to say satellites), for completeness I would ask the Authors to list some of these as attempts to bridge that gap.

Reply: As suggested, we have added “some attempts have been made to bridge the gap between local observations and regional estimates using combinations of aircraft and ground based measurements from field campaigns (e.g., SAMBBA, ARCTAS), satellite-based inventories, and chemical transport and aerosol models (e.g., Fisher et al., 2010; Reddington et al., 2019; Konovalov et al., 2018).”

5) Line 99: Define ‘present day period’, i.e. list years data available.

Reply: We have added “i.e., since 1997 for GFED and shorter periods for others”.

6) Line 100: Suggest altering to say something like ‘gases such as. . .’ as they way it is currently presented appears to be a definitive list but is not. For example, vanillic acid has also been used as a unique tracer of fires. Please also make it clear that is the C3 methane carbon isotope which is the tracer, as this species has many sources.

Reply: In the revised version, we have rephrased the sentence as “Historical change of fire emissions has been inferred from a variety of proxies, such as ice-core records of CH_4 (isotope $\delta^{13}CH_4$ from pyrogenic or biomass burning source), black carbon, levoglucosan, vallic acid, ammonium, and CO (Ferretti et al., 2005; McCornnell et al., 2007; Conedera et al., 2009; Wang et al., 2012; Zennaro et al., 2014), site-level sedimentary charcoal records (Marlon et al., 2008, 2016), visibility records (van Marle et al., 2017a), and fire-scar records (Falk et al. 2011).”

7) Line 104: Can the authors add a few words to describe aerosol indices, it is perhaps not as common as the others and would aid in reader comprehension.

Reply: The aerosol index represents the amount of absorbing aerosols. We have removed it, and changed to “fire-scar records” which is more commonly used.

8) Lines 104-109: Suggest that the Authors add something positive here about

proxies for balance. While it is true that no proxy can accurately define the past, it currently reads a bit as if you are suggesting all this work is not of any worth.

Reply: We have added "Fire proxies can be used to reconstruct the changes of fire emissions on a local to global scale and for time periods of decades to millennia and beyond"

9) Lines 117:119: Suggest: 'Fire emissions of trace gases and aerosols are derived from the product of the simulated DGVM carbon emission and a species emission factor (Li et al., 2012; Knorr et al., 2016).'

Reply: Done

10) Line 185: 'their estimates of' rather than 'the simulations of'

Reply: Done

11) Line 186: remove comma

Reply: Done

12) Line 190-195: Much of this is not grammatically correct, please rephrase.

Reply: Changed to "CLM4.5 models fires in croplands, human deforestation and degradation fires in tropical closed forests, and human ignition and suppression for both occurrence and spread of fires outside of tropical closed forests and croplands."

13) Lines 227-235: The information in this paragraph could come before the protocol in the paragraph before. Such that when reading the protocol, it is clear where the data is from already.

Reply: Reordered as suggested.

14) Line 255: See Andrea (2019) for details; as this paper is only in prep I would suggest not explicitly directing the reader to it for more details.

Reply: The manuscript was published recently. We have updated the reference.

15) Line 255-256: Suggest: 'All FireMIP model simulations used the same EFs from Table 2.'

Reply: Done

16) Line 261: Incorrect placing of semi-colon (should be a comma), it could however be placed before 'similar' if wanted. Also suggest adding 'are classified as' for each of the three PFT instances not just the first.

Reply: The sentence has been divided into two, so the semi-colon is now a period. Also, the words "are classified as" have been added.

17) Line 287: please define 'them'

Reply: "them" has been replaced by "satellite-based estimates of present-day fire emissions”.

18) Line 316: The definition of discrepancy is ‘a difference between two figures, results, etc. that are expected to be the same’. I do not think these results should be expected to be the same as the underlying factors have uncertainties in their representations, as the Authors mention?

Reply: “discrepancy” has been changed to “difference”.

19) Line 317: Emissions are ‘from’ the land, not ‘over’ them which is the concentration. Suggest to double check for occurrences elsewhere.

Reply: all “over the land” have been changed to “from the land”

20) Lines 347-350: More details here please. . .. Why? Which models are driving this variability? Do satellites suggest this is a variable region too? etc.

Reply: We have added “This is mainly driven by the MC2, CTEM, JSBACH-SPITFIRE, and ORCHIDEE-SPITFIRE simulations (Fig. 2).” and “The differences among the satellite-based estimates have a similar spatial pattern, but higher than the inter-model spread in savannas over southern Africa and lower in the temperate arid and semi-arid regions and north of 60ºN over Eurasia (Fig. S1a).” in Sec. 3.1.

Furthermore, we have added Fig. S1a in Supplementary Material which is similar to Fig. 3 but for satellite-based estimates of fire emissions.

21) Lines 402-403: But in disagreement with the ice-core/tree scar/charcoal proxies? These show variability in emissions from 1700-1900, with a peak ~ 1850?

Reply: Yes. We have added “but in disagreement with earlier reconstructions based on charcoal records (Marlon et al., 2008; Marlon et al., 2016), ice-core CO records (Wang et al., 2010), and ice-core $\delta^{13}CH_4$ records (Ferretti et al., 2005), which exhibit a rapid increase from 1700 to roughly the 1850s. ”

22) Lines 531-535 and 547-550: If most models do not capture these trends does it not therefore suggests that historical emissions are likely underestimated in most fire models (and hence also CMIP6)?

Reply: Yes, it does.

We also note that besides human suppression on fire spread and the decrease in fuel continuity from expanding croplands and pastures (Lines 531–535 and 547–550 in old version), human deforestation and degradation fires and crop fires are not modeled by most FireMIP models which can also affect the simulations of historical fire emissions. At this stage, we are unclear about the net effect of these factors. We think this is an important point to address and have added a discussion in Sec. 5 (see response to your next comment below).

23) Line 551: The conclusions appears to stop a bit abruptly, could the authors finish

the conclusions on an outlook or implication etc. to tie it all together a bit more. One example, global CMIP6 emissions are basically flat w.r.t. time, and so using model emissions which are much more variable will result in a different simulated climate/Earth system response.

Reply: Thank you for this suggestion. We have added a paragraph in Conclusions as "As discussed above, most FireMIP models do not consider the human suppression of fire spread, decreased fuel continuity from expanding croplands and pastures, human deforestation and degradation fires, and crop fires. Therefore, these models, and hence the CMIP6 estimates that are mainly based on them, may have some uncertainties in estimating historical fire emissions and long-term trends. This may further affect the estimates of the radiative forcing of fire emissions and the historical response of trace gas and aerosol concentrations, temperature, precipitation, and energy, water, and biogeochemical cycles to fire emissions based on Earth/climate system models which include these fire models or are driven by such fire emissions. It may also influence future projections of climate and Earth system responses to various population density and land use scenarios."

24) Figure 2: The lat/lon co-ordinates are too small to read. Remove as they are not actually necessary.

Reply: In the revised version, only lat labels at the rightmost and lon labels at the bottom are retained but with a bigger font size as some readers may want to have this information, and all other lat/lon labels have been removed.

25) Figure 7: suggest moving d and e to the a and b positions then decreasing the axis limits for the other three so the differences can be seen.

Reply: We decided to use the same y axis for Figs. 7a-e so readers can easily compare the magnitude of the simulated response of fire emissions to different drivers. The main objective of Fig. 7 is to highlight the importance of simulated responses to LULCC and population density change in the inter-model disagreement of historical fire emission changes, so the same y axis seems better.