Reply to reviewers second revision

Report #1

This paper corresponds to the revised version of the paper "Attribution of the Australian bushfire risk to anthropogenic climate Change" previously submitted to Natural Hazards and earth System Sciences. The authors have made a great effort to answer the main issues raised in my previous evaluation. In particular, they revised the introduction, results and discussion section, shortening the manuscript significantly.

I am satisfied with the new version of the manuscript itself. However, the solution adopted by the authors to shortening the manuscript was to send the entire section 3. to supplementary information. In fact, the supplementary information has more similarities to a second paper than to a supplement material. I strongly suggest reducing information in the supplement and reorganize it in order to fit the usual format of this kind of material.

Therefore, I am glad to give my approval on the publication of the improved version of the paper after the necessary changes on supplementary information.

We agree that the supplementary information is long and contains introductory (SI) paragraphs which were taken from the original main text. We do believe that the majority of the SI is indeed needed to corroborate and support the results in the main text. However, we have revised the SI by starting with a summary and justification for the necessity of this supplementary analysis and have shortened some other parts of the SI to hopefully avoid the character of a stand-alone paper.

Report #2

I appreciate the revisions that have been made to this paper. This version is much improved over the original paper – and indeed, is closer to the paper that should have been submitted in the first instance.

I disagree with the implied suggestion that allowances should be made due to time pressure associated with the goal of rapid attribution – indeed, if the rapid analysis of events has now evolved into a quasi-operational activity, then I would argue that results should be published elsewhere than in the peer reviewed scientific literature unless specific cases have something to teach us that was not previously known about the methods and tools that are used or the mechanisms that produced the event in question.

We agree completely with the reviewer for operational attribution analyses. However, this study is definitely not yet in this category. Only heat and cold events, and large-scale precipitation events are in our opinion ready to be operationalised because there are enough studies published, so that following a protocol will give the correct answers. This is the first wildfire analysis to be published from this team. There have not been many by other authors and some of those have in our view serious shortcomings, like analysing only a single climate model or only the reanalysis period, both of which can give spurious trends. We want to give an example of a more thorough analysis in the peer-reviewed literature that future operational studies can refer to and refine.

While the paper is improved, I do have some additional comments that are listed below. Many of these comments call on the authors to think a bit more carefully about how they are communicating with readers since they do not necessarily share the vocabulary that the authors use to communicate amongst themselves.

Thank you, it is always useful to get these types of comments from reviewers outside the field for a paper that probably will attract many readers from outside the attribution community.

51-52: The manuscript mentions in a few places that "ignition sources and type of vegetation[, which are] factors largely independent of meteorology[,] play an important role", but surely this is not true. Lightning is certainly a major ignition source, and the type of vegetation in a given location is certainly dependent on the climatology of that region. Better wording would be appropriate.

We agree and have reworded the text to 'In addition, ignition sources and type of vegetation play an important role. The types of vegetation depend on the long-term climatology, but do not vary on the annual and shorter time scales we consider, and the dry thunderstorms providing a large fraction of the ignition sources are too small to analyse with climate models. In this analysis we therefore only consider ... '

128: Do you mean similarity in the correlation coefficients (rather than confidence intervals)? The discussion goes on to mention explained variance, which implies that it refers to the square of the correlation coefficient.

Indeed, this was an oversight. We have revised this to read: 'Given the similarity in the correlation coefficients (r) within their confidence intervals, the log-linear relationship appears to explain equal variability (r^2) to that of the ranks.'

169-170: Exactly what is an "annual mean low precipitation" value? This seems a confusing combination of words.

'low annual mean precipitation'

185: I think what is meant is that the iterative maximization of the likelihood function does not converge. It is not the "fit", per se, that doesn't converge.

I am afraid I do not see the difference but I am not an expert on the terminology here so we revised according to the suggestion of the reviewer.

193: What is meant by the statement that precipitation is "positive-definite"? Most readers would think about a matrix when they see this term, and wonder whether they missed something in the description of the methods that involved the creation of a matrix of some kind. A few might think about other notions of positive-definiteness as defined by mathematicians, and I think they would also wonder what is meant. Maybe you just mean that precipitation observations are always non-negative?

Changed to non-negative (twice)

196-198: I have two comments on this new sentence.

First, ENSO (and other modes of internal variability that affect the region) are certainly relevant since it would be hard to think that the probability of extreme temperature, precipitation and FWI are all insensitive to the phases of these modes of variation.

We define a covariate that describes the trend as well as possible. This should exclude ENSO, as ENSO has no trend, hence we use a low-pass filter that excludes the variability due to ENSO that should not be included in the trend. Added 'not relevant for the trend'.

Second, what kind of externally forced variability are you talking about here? If the interest is in anthropogenic forcing, which has a largely monotonic response over time, wouldn't a longer time average that better isolates the signal by removing more of the high frequency variation (mostly internal, but perhaps also due to episodic volcanic forcing) be better?

It is not really possible to include volcanic forcing and exclude the internal ENSO variability by temporal filtering as they have the same timescale (Lehner et al., 2016, *Geophysical Res. Letters*). The length of the filter is a trade-off. In principle a longer time scale would be better, but as we mention the event that is being attributed is usually at the end of the unfiltered time series or just beyond it, so that a long filter becomes increasingly ill-defined for that part of the time series. A 4-yr running mean requires an estimate of the next two years, which can be assumed to be similar to the current year (e.g., van den Dool and Barnston, 1996), but a 21-yr running mean or LOESS filter requires estimates for the next 10 years, over which GMST will continue to rise.

201: Here and throughout, make it clear that all probabilities are ESTIMATED, and thus subject to uncertainty.

Certainly, added this three times.

204-212: There is something here that I seem to be missing. For example, if the GEV location parameter is a linear function of T', then how can it also be an exponential function of T', as in equation (3)?

It is a linear function for temperature and an exponential one for the non-negative parameters precipitation, FWI and MSR. Clarified the text.

251: The responses to my previous comments promised that this shorthand (χ2/dof) would be clarified – but that appears not to have been done (I could not find a definition of the statistic that is referred to here in either the main text nor the supplement.

We have added the definition and a clearer explanation how it is used.

271: How did you determine that the factor was "at least two"?

This is the lower bound of the model synthesis, which is unfortunately not shown in Figs. S6. We have added this bar to the figures S6 and S7 and mentioned it in the text. We rounded the lower value of the PR in the model world from nominally 1.84 to two in order not to suggest too much accuracy.

Note also that there is an important nuance of the communication of probability ratios that the paper seems to be sloppy about. When PR=2, p1 = 2p0, and thus the event is estimated to be 2 times AS likely in the current climate as in the counterfactual climate, or equivalently, 1 times MORE likely. If you say "at least two times more likely", then my interpretation would have to be that PR ≥ 3. This vagueness of interpretation seems to crop up in several places.

We assume the reader knows that the PR is declared REAL and not INTEGER.

330-332: This convoluted sentence will be very hard for readers to understand. I suggest breaking this up into two or three sentences that explain in bite-size chunks the differences between the analyses of the observations and that of the climate model output (rather than trying to make the readers swallow the entire sandwich at one go).

Thanks, done.

Also, here and elsewhere, be careful with the word "models". For example, on line 330, the text reads "the models use as covariate the model GMST". Evidently the first use of "model(s)" refers to the statistical models used in this study, whereas the second use of "model", only 5 words later, refers to climate models. That implicit distinction will be clear to some readers, but many other readers will be puzzled.

We attempted not to use the word 'model' to refer to statistical models but only to climate models. In this sentence that was the intention also, reformulated to make this clear.

Figure 3 caption: Describe the grey shaded band that is shown in the two right-hand panels.

Done: ''.. and the grey bands the 95\% uncertainty ranges.'

Figure 5 caption: Describe the shaded bands that are shown!

Done: 'The bars denote the 95% uncertainty ranges'.

343-344: The statement that "all model dispersion and shape parameters lie within the large observational uncertainties" seems a bit of a stretch. There is some overlap in the spread of the dispersion parameters from the climate models with that from the observations for 3 of the 4 climate models, but that overlap doesn't necessarily mean that for those climate models, one would accept the null hypothesis that the dispersion parameter from the climate model is the same as the dispersion parameter from the observations.

We agree that the correct way to do this would be to formulate the null hypothesis that the values are compatible and do the straightforward statistical analysis from there. However, due to a miscommunication within the group the procedure has been simplified in our published protocol (Philip et al, 2020) to whether the ranges overlap, which gives a slightly more lenient test. We made the choice to follow the published protocol here and update it in the near future.

353: Results from ERA5 are confounded with the impact of low-frequency internal variability during the single 41-year realization of the observed climate that ERA5 attempts to reconstruct…

This is exactly the point we try to make and why we also did the long-term drought analysis using a much larger ensemble (cf Goss et al, ERL, 2020 on the Californian fires). We added a sentence re-emphasising that here as well: ''Note that the ERA5 value is probably biased high as the positive contribution of trend towards a drier climate over 1979--2019 is not present over 1900--2019, see sections S2 and 5.6.'

355-356: Here is another example of ambiguity in the interpretation of the probability ratio. Judging from the figures, I think you should be saying seven rather than eight times MORE likely, and for the lower bound, perhaps only 2.5 times MORE likely.

The unrounded numbers are 7.96 and 4.31.

367: This might not be as much of a climate model deficiency as implied here (which is what "underestimation" implies). As very briefly discussed in Section 6, and further elaborated in Section S3, apparently the IOD and SAM did play a role.

Using our class-based event definition, the variability of the SAM and IOD is included in all these results, so does not explain the discrepancies. It could be that the discrepancies come from problems representing the IOD and SAM, but we show that this is unlikely to be the case in the supplementary material due to the small contribution of these modes to the historical variability.