EGUsphere, author comment AC1
https://doi.org/10.5194/egusphere-2022-345-AC1, 2022

Reply on RC1

Michael A. Storey and Owen F. Price

Author comment on "Statistical modelling of air quality impacts from individual forest fires in New South Wales, Australia" by Michael A. Storey and Owen F. Price, EGUsphere, https://doi.org/10.5194/egusphere-2022-345-AC1, 2022

Thanks you to the reviewer for taking the time to read our manuscript and provide insightful comments, it is much appreciated. I have provided a response below to each of the individual comments (my comments underlined and italics):

Reviewer Comments:

- **Introduction:** The intro starts citing a reference which is 10 years old. Is this true now? Please comment on or change it to reflect more recent numbers.

We will add some other more recent references here

- **Introduction: Minor comment, Line 94,** "We need to better tools (…)"

We will correct this

- **Introduction: Line 96-99,** Add some references to justify on why these attributes are important.

We will add supporting references

- **Introduction:** Another alternative to the methods cited is using satellite information.
- Please add some information on that, e.g., Gupta et al., 2006 https://www.tandfonline.com/doi/full/10.1080/01431160701241738

We will mention this and provide a reference

- **Introduction:** The option on using random forest models is not justified in any way in the introduction. Some information should be added on this type of approach. Another kind of modelling approaches could be used and should be mentioned, e.g., neural networks, regression models, GAM, among others. Why are random forest models better than other options? Please justify.

We can add a couple of sentences about random forest modelling and its advantages, which we believe would be best placed in the methods.

- **Methods: Line 158,** Why use a threshold of 125? Please justify

This is actually a scaled score and translates to a cover fraction of 0.25. We will update the manuscript to say 0.25 instead as it makes more sense. Our study focused on forest fires. This threshold ensured only dense woodlands, open forest and closed forest was included in our study.

- **Methods: Line 159-160**, Why use a buffer of 2.5 km? Please justify.

We can provide some further information here, and acknowledge that using a different buffer may have produced slightly different results

- **Methods: Line 164-166,** The results for the 3-days window could have been shown in the sup. material.

Would believe our mention of this is sufficient here and that added a second set of plots in the appendix would not be all that informative. Given this was not raised by the other reviewer, we would prefer not to make any changes here.

- **Methods: Fig.1 and Line 203,** How do you justify using monitoring stations with data records with only 3 or less years of records? In what way the exclusion of stations with shorter lengths would influence the results of the aggregation process?

For a new row in our data tables, for each active fire we only required at least one day (afternoon, night & morning periods) at an air quality station to be captured. We did not make the sampling dependent on the length a station is active, we only searched for occasions where any active fire was near any available station. Excluding short term records may alter results (probably not change the overall conclusions much), but much of our data does come from more recent years due to many fires in this period and more stations becoming active.

- **Methods: Line 211**: "We also calculated the sum of the hotspot day and night fire area as a predictor". The sum of what? Please clarify.

Note that we had to change our method of area calculation (which produced very similar final results), so this section will be updated in a revised manuscript.

- **Methods: Section 2.4,** Please add some more information on random forest models, namely, way to apply, advantages and caveats.

We will add a few sentences on random forests here, as also suggested in a prior comment.

- **Methods:** Why do you opt by doing a simple validation (75% training and 25% test) in detriment of a cross-validation or other validation method. Please justify.

We were aware of several approaches to cross-validation that could possibly be used. The random forest out-of-bag r-squared is an in-built independent validation method which we have reported on for our training steps. In the end, and given that we had sufficient data, we chose to use a simple training-test split which is a regularly used approach. Adding results from a completely independent set has the advantage of giving a practical example of accuracy of predictions to a new test set and confirming the OOB accuracy. The prediction observation on the training (OOB) sets are consistent with the correlation on the test sets, which suggests changing the validation approach wouldn't change the results substantially.

- Do you compare models' outputs to observations in an independent sample, not used to create the models? An independent sample was mentioned in the results section but information on this should be added on the methods.

Yes, the comparison with the test set is included in Table 1 and Figure 7. We refer to the table and figure in the results and state the accuracy on the independent test set too.

- **Results:** Did you analyse if the under/over predictions were due to models not being able to capture the intensity of the events or because they are able to capture that but with a delay? This is important to try to understand and correct models' performance.

We did not look at this detail. We could have a further look. It would be helpful if a couple of references could be provided that look at this issue in other random forest models.

- **Results:** Table 1: Please verify the legend.

We have updated this to "Table 2", as it was mis-labelled. The rest if the caption looks accurate.

- **Results:** Do you think it might be possible that night results are worse as afternoon presents higher values and thus the models have a harder time to reproduce night values? If this is true, how can you correct it?

The distribution of $PM_{2.5}$ for night and afternoon are fairly similar: Vast majority are < 20 ugm^{-3} for morning and night, with a small proportion > 20 (Figure 3). The 20 highest values for each (for both WF and HRB) are also within a similar range, being mostly 50 and 150 ugm^{-3}. There is also only a small difference in accuracy on the training and tests sets in terms of prediction and observation correlation (~0.8). Given this, we don't believe differences between the distributions of $PM_{2.5}$ between morning and night would have greatly affected the model results.

- **Discussion:** One of the caveats of the approach is that only considers part of the fires. Therefore, the first goal is only partially achieved as a great part of the fires and corresponding weather conditions are not analysed.

We have now ensured that in the methods, results and discussion that fire area is referred to as "daily active fire area", so that it clear we are not referring to the total final fire area. We believe this is the clarification needed, but if there is a further issue needing attention, please provide some more details and we are happy to look at it.

- **Discussion:** The authors do not account nor mention the effects of recirculation potential on these events, nor even the link to PBLH. There several papers connecting poor air quality events and recirculation and would be nice to refer them in the discussion. https://www.sciencedirect.com/science/article/pii/S1352231002009263
- https://www.sciencedirect.com/science/article/pii/S1309104221003305
- https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007JD009529

Thanks for the references. We will add recirculation into an existing part in the discussion about broader weather patterns, and include these references. We have briefly discussed PBLH in the discussion and believe this is sufficient, particularly given the discussion is already quite long.