Comment on egusphere-2022-345
Anonymous Referee #1

Referee 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-RC1, 2022

This paper models PM2.5 emitted from 1500 historical individual fires in NSW Australia, as a function of fire and weather variables. The PM2.5 values were measured at air quality stations in NSW Australia. The study combines data from satellite, reanalyses of meteorological data and monitoring station’s data (for air quality). VIIRS satellite hotspots were used to identified days where one fire was burning within 150 km of one of 48 air quality station. To model and predict air quality events, the authors used random forest models for afternoon, night and morning PM2.5.

The authors identified the most important variables and drivers which are a mix between fire characteristics and meteorological conditions, namely fire area, boundary layer height, temperature, wind speed and relative humidity. The models underestimate the observed pollution values.

The subject discussed in the present article is of high importance, and the paper is well structured and easy to follow. The introduction is very clear and includes interesting and recent references. The objectives are very well presented. The limitations of the study are in a great measure identified. This is an important point, as one of the caveats relates to the fact that not all fires are accounted and some of the fires not accounted (because of the co-occurrence of fires) can exclude fires which contributed greatly to large pollution events. The discussion is quite interesting. Still, there are points that need further details and work. I believe that this document should be considered for publication after minor/moderate changes.
Below I point some comments and suggestions, which hopefully can help the authors to enhance the manuscript.

**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.

- **Introduction:** Minor comment, Line 94, “We need to better tools (...)”

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

- **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

- **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.

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

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

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

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.

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

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.

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.

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

- **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?

- **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.

- **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.

