

Nat. Hazards Earth Syst. Sci. Discuss., author comment AC1  
<https://doi.org/10.5194/nhess-2022-52-AC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Reply on RC1

Annalie Dorph et al.

---

Author comment on "Modelling ignition probability for human- and lightning-caused wildfires in Victoria, Australia" by Annalie Dorph et al., Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2022-52-AC1>, 2022

---

We thank the reviewer for their comments. We have addressed each of their comments and proposed solutions or clarifications were necessary. Below we outline the changes that can be made in response to each comment (reviewer comments are italicized, response in plain text).

*I would like to congratulate to the authors for their work.*

*In this work authors train 6 ML algorithm (random forest) using variables of topography, fire, infrastructure, climate, weather, dryness and fuel moisture, to estimate the probability of fire ignition. They create a training dataset for 4 different combinations human/lightning in native/matrix vegetation. The fire ignition data is obtained from Victorian Country Fire Authority, which uses 20 classes for causes of ignition. Then the authors reclassify the ignitions between human or lightning to create the training dataset for the 6 models. In the lightning case, there is the need to create add random fires to create absence data. Then the authors analyse the classification error and the R2. Following, the analyse the importance of the variables for the different models using the decrease of the gini index.*

Notes:

- 1. It is hard to follow the number of presences and absences for the 4 datasets created for the models. But would good to visualise the presence/absences in Fig 1 for the 4 datasets. Also, the number of presences and absences of each of the 4 datasets.*
- 2. Since part the data is not public the previous point 1 would help a bit. Please, specify which part of dataset is not public. Some efforts to show partially that data would help the reader.*
- 3. It may be impossible to reproduce the work because the data used is not public.*
- 4. L85 "Each random point[...]. These random points" I consider there is in assumption about completeness of CFA data, which it may be fine in this case. But the reader has to accept the assumption without being able to see and check the data.*

Comments 1-4 refer predominantly to the datasets not being publicly available and impacting the reproducibility of the work. The data can be acquired by other researchers;

however, it requires gaining permission from the fire agencies who manage the fire ignitions data and purchasing of the lightning data from GPATS.

As suggested by the reviewer we can change Figure 1 to a four-panel figure plotting the points where presence/ absence data were located for both the human and lightning datasets across the landscape. However, for potential lightning ignitions a density map of lightning strike locations may represent the number of strikes more clearly. Alternatively, we can change Figure 1 to a table summarising the number of presences/ absences for lightning and human-mediated ignitions within the two different vegetation types and move large maps of the spatial distributions of these points to the Supplementary material.

*5. When merging all these datasets would be good to have quick look to the temporal and spatial resolution (in table 1?).*

We can add rows to the end of Table 1 with the response variables, a description of the data and its spatial resolution and the temporal scale covered as suggested.

*6. It is not so trivial for the reader to do a diagram of the use of the data flow for the lightning models. This point is more a doubt than nothing else. So, data from GPATS is used to create a training dataset computing the probability of ignition, and then, applying a two stage approach (classification and after regression) to avoid 0 probability of ignition. But, when the fire ignitions (caused by lightning) from CFA are used? GPATS give you lightning info occurrences, you use that to model the probability of fire ignition (Larjavaara et al 2005) as a "ground truth", and after evaluate that with the two stage approach with the variables from table 1. What about the presences of ignition by lightning from CFA? How the model performs with these cases? So, here you are comparing two models (Larjavaara vs two stage random forest) using considering GPATS for Larjavaara and the data from table 1 for the two stage random forest. However, the 8781 cases from CFA seems not to be used at all. I assume the contribution here is the comparison of two completely different models that use different input data. Would be possible to see just the performance for that 8781 cases for matrix and native.*

There a couple of points here where we think the reviewer has misunderstood the process illustrated in the flow chart. The diagram attempts to highlight that once GPATS and CFA data was acquired, the equations from Larjavaara were used to link the two datasets by:

- calculating the probability of lightning strokes (GPATS) being linked in space and time to the lightning strike ignitions (CFA) and;
- group the lightning strokes (i.e. single cloud to ground transition of electricity) into lightning flash events (i.e. what is seen by the eye when lightning strikes)

When GPATS lightning strikes were within 5 days and 10 km of a CFA ignition then they were counted as ignition presences in the final dataset. Any lightning strokes which were not linked to a CFA ignition were counted as absences. Thus, it is not a method of ground-truthing as the reviewer has suggested, but a method to integrate the two datasets to give presence-absence data. Consequently, the CFA data is not compared separately, because it is integrated into the presence/ absence calculations for the lightning strike data. We considered this approach to provide more accurate information than following the same approach used in the human ignitions model in which the 8781 recorded CFA ignitions would be recorded as presences, and absences would be generated using random points. The process of linking lightning strokes to fires and grouping them together is explained in text. We can add further clarification in text to make the outcomes of this process clearer, for example by adding to the end of the section regarding GPATS lightning data after the lines 153 – 156 which reads:

"The inclusion-exclusion principle was used rather than the sum of probabilities as the latter could be greater than one when a lightning stroke was linked to more than one fire. More information on the inclusion-exclusion principle is included in Appendix 1."

We could add the following:

"By completing these calculations, all lightning strokes that were linked to a CFA fire ignition using the proximity index were assigned an ignition probability greater than zero. Any lightning strokes that were not linked to a CFA ignition by the proximity index were assigned an ignition probability of zero and were treated as absence data in the analysis."

*7. I would specify the criteria for the reclassified ignition causes from the CFA classes.*

The criteria were specified in the text (see Lines 80 - 83 which read): "For this study, ignition causes were reclassified broadly into human-caused (n = 59,146; e.g. from arson or accidental sources) and lightning-ignited fires (n = 8,781) as previous work found consistent patterns in the drivers of the different types of human ignitions in the study area (Clarke et al., 2019)". However, to improve clarity further we could add all of the CFA classes to each of the classifications rather than providing an example.

*8. One doubt in the partial dependence plot. When the rainfall increases, also increases the probability of ignition. The rainfall of these plots is an annual mean as mentions in table 1. So, each pixel has the average from 1970 to 2020 of the mean annual rainfall? So, each pixel has a single value with the average of the mean of the annual rainfall. It would mostly provide spatial information and is not related at all with seasonality moisture. It may be related with the fuel production. I suggest try to define the parameters of rainfall (time) in the figures...*

In this paper, average annual rainfall was not being used as a measure of seasonal moisture variation. Instead, we represent moisture variability using dead and live fuel moisture estimates. Instead average annual rainfall is being used to represent the number of storms different parts of the landscape are subject to. In this landscape, rainfall and elevation are highly correlated. Therefore, we used rainfall here as a surrogate for topographic variation as areas of higher annual rainfall correspond to higher altitude and more topographically complex landscapes. As these areas are prone to a higher number of storms, average annual rainfall was considered an appropriate variable to represent this variation. An additional explanation of this can be added to the manuscript around lines 101-102.

*9. L265 "In matrix vegetation, Rainfall was the second most important variable (Fig. 3c). In the manuscript I have rainfall in on top for Fig. 3c., second is dead fuel moisture.*

The reviewer is correct, this is mistake in the writing of the manuscript. The sentence should read "In matrix vegetation, Rainfall was the most important variable (Fig. 3c)."

*Please cite the datasets used wherever is possible.*

All predictor variable datasets are listed in Table 1. As mentioned above, we are able to add the response variables to the tables to re-iterate the spatial and temporal information as well as the data's source.

*The results and conclusions of the work rely on the data. Please describe more the dataset used, versions, resolutions and for the ones that are not public may be analysis and map. See for instance Clark et al. 2019, similar concept, but different methods and inputs. Despite this, the description on the data used is more clear and easy to follow.*

Comments are provided on this above.

*Again, congratulations for the work. It is really interesting and potentially useful.*