Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/nhess-2022-52-RC1, 2022

Comment on nhess-2022-52

Tomàs Artés Vivancos (Referee)

Referee 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-RC1, 2022

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, infraestruture, 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 completness of CFA data, which it may be fine in this case. But the reader has to accept the assumtion without being able to see and check the data.
5. When merging all this datasets would be good to have quick look to the temporal and spatial resolution (in table 1?).
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 regresion) to avoid 0 probability of ignition. But, when the fire ignitions (caused by lightning) from CFA are used? GPATS give you lightning info ocurrences, you use that to model the probability of fire ingition (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 perfomarce for that 8781 cases for matrix and native.
7. I would specify the criteria for the reclassified ingition causes from the CFA classes.
8. One doubt in the partial dependece 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.

(https://www.worldclim.org/data/worldclim21.html and
https://www.worldclim.org/data/monthlywth.html) with that possible citations:

Monthly:
CRU-TS 4.03 (Harris et al., 2014) downscaled with WorldClim 2.1 (Fick and Hijmans, 2017).

Fick, S.E. and R.J. Hijmans, 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37 (12): 4302-4315.

Harris, I., P.D. Jones, T.J. Osborn, and D.H. Lister (2014), Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. International Journal of Climatology 34, 623-642. doi:10.1002/joc.3711

Climate:
Fick, S.E. and R.J. Hijmans, 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37 (12): 4302-4315.

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.

Please cite the datasets used wherever is possible.

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.

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