

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/nhess-2021-245-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on nhess-2021-245

Anonymous Referee #2

Referee comment on "ProbFire: a probabilistic fire early warning system for Indonesia" by Tadas Nikonovas et al., Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2021-245-RC2>, 2021

The manuscript presents a probabilistic early fire warning system for Indonesia which is based on both climate forecasts and non-climate datasets. Overall, I congratulate the authors on a well written manuscript, both in terms of presentation quality and scientific quality.

I only have a couple of minor comments:

- MODIS imagery is used as a basis for fire inventory. I am wondering whether this choice might introduce a bias towards large fire events, consequently leading to an underestimation of small fires? This is mentioned around line 118, but might be worth a short comment in the discussion section.
- l. 144: how is the threshold motivated? Is this just because 10 is a nice number, or is there evidence that counts > 10 are particularly dangerous or may indicate particularly dangerous conditions?
- Section 2.3: There are some indicators available through copernicus Global Land Service (<https://land.copernicus.eu/global/themes/vegetation>) which might be of interest for future work. Most of them are based on Sentinel-3/OLCI or PROBA-V and are available from 2014 onwards.
- Section 3.1: It is unclear to me whether the hyperparameters were tuned or simply set to the values reported in the manuscript? The text does not provide any indication of hyperparameter tuning (e.g. via cross validation), a quick glance into the code implies that values were simply determined beforehand. This should be stated more clearly.
- Section 3.1: Minor technical nitpicking: I think the split described here refers to 'train' and 'test' sets, respectively. The 'validation' set is usually a subset of the training dataset used for hyperparameter tuning, before testing the tuned and validated model on the unseen 'test' data.
- Section 3.6 / Formulas (3) and (4): I think it is more common to use the terms "true positive" (TP) instead of "hits"; "false negative" (FN) instead of "misses" and "false positive" (FP) instead of "falsealarms". "pod" could be "sensitivity" or "recall".
- I assume that the data set is somewhat imbalanced - i.e., there are more non-event pixels than fire pixels. Was this accounted for (in terms of model formulation or in

terms of performance metrics)?

- Finally, I would like to acknowledge the provision of the ProbFire source code via GitHub and the comprehensive data availability statement.