



Comment on **essd-2021-113**

Anonymous Referee #3

Referee comment on "Refined burned-area mapping protocol using Sentinel-2 data increases estimate of 2019 Indonesian burning" by David L. A. Gaveau et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-113-RC3>, 2021

The study presents a new burned area product for the year 2019 in Indonesia based on high spatial resolution Sentinel 2 imagery and machine learning classification algorithm. Given the significant impacts that severe burning episodes in Indonesia have on global carbon cycle and population health across the wider region, the product presented in this study represents an important development and will be of interest to the scientific community. The approach and the dataset, nonetheless, have several limitations which I believe should be better articulated in the revised manuscript. In addition, I don't think that the comparison of fire patch size distributions between different products adds much to the discussion here due to (i) large differences in spatial and temporal resolutions (or both) between the datasets and (ii) lack of definition what does fire patch represents here.

While the validation methodology does seem robust and the authors do demonstrate that total burned area estimates of the study are more accurate when compared to the alternative sources (MCD64A1 and the Official ba product), it has to be articulated that the algorithm of this study was optimised for the specific region and fire season and for a specific commission/omission error ratio. As a result, it is not clear how the burned area estimate for 2019 would change if the algorithm was optimized to fit training data from different years and regions by different users. In addition, extending temporal coverage of the dataset is not that straightforward as this would require substantial further work (somewhat arbitrary and time-intensive selection of training data). Please see the bellow specific comments for further detail.

Specific comments

Lines 56-57: Given the uncertainty in burned area estimates (line 56), the Huijnen et al., 2016 estimate of CO₂ emissions quoted in line 57 seems too certain. Do Huijnen et al give uncertainty estimate? Also, would be good to give another estimate for the event, given by GFED or Lohberger et al., (2018) or some other study etc. Large uncertainties in emission estimates is yet another reason why we need better burned area products, hence it would be good to point this out here.

Lines 160: Please explain what "Every two days" means here.

Line 161: Was data from the central day of the window included in prior or after median values (or neither)?

Line 163: This relates to the previous two points regarding temporal precision. Here and elsewhere the authors use "The day of the year". How day of burn was determined if temporal resolution of Sentinel 2 is ~5 days as stated earlier? This suggests considerable uncertainty in day of burn estimate?

Lines 173-176: It is not clear here what was the total number of features used for classification? Please state in this paragraph.

Lines 177-178: Was the training sample fully independent from the validation sample? This is important to state clearly as it underpins the validity of the study's findings.

Lines 260-265: What was done with classification of the sites which had either direct fire evidence (flame or smoke) but not indirect evidence (reddening) observed and vice versa? Where these samples (if any) discarded from the analysis?

Line 278: The final validation sample number N=1298 is the same as given in line 269 (all reviewed sites) and line 227 where it is termed as "initial sample". Please clarify this.

Line 306: This statement needs a reference and an explanation. MODIS burned area pixel size is ~21ha. While sub-pixel burning can be detected, the actual minimum burned scar size will depend on environment/vegetation where burning is occurring. Is this estimate of 6.25ha is specific to Indonesia/tropical regions? In addition, I am puzzled by how MCD64A1 fire size histogram shown in Fig. S2 was computed; i.e. how counts for bins for fires < 21ha were derived given that MODIS pixel size is ~21ha?

Lines 366-367: The sentence is too "wordy" and complex. "greater detection of the realm of fire activity characterized by small-scale..." could be replaced with "greater detection of small fires" to the same effect.

Line 371: "lesser estimation" – perhaps change to underestimation?

Lines 380-389: The paragraph is too wordy. The first two sentences say nearly everything that needs to be said. Sentinel 2 sensor can indeed detect smaller fires enabling the detection of small scale agricultural burning. Perhaps cut shorter or even merge into previous paragraphs.

Lines 407-416: This paragraph is very speculative and not well supported. Not sure I agree with such interpretation of fire size frequency distributions. Any differences in distribution shape may arise from huge differences in sensor spatial (MODIS) and temporal (Official map) resolutions and also from the clustering (patch agglomeration) method. For example, the algorithm of the official product may have merged diagonally adjacent pixels as well and that would result in shift towards larger sizes. In addition, the provided references do not show that power-law approximates fire event sizes in Indonesia. While I'm not aware of fire size studies in Indonesia, log normal fire size distributions are common in some ecosystems (see Lehsten et al., 2014). As a result, it is perhaps better to avoid saying that fire sizes should follow power-law relationship and that this itself is a desirable property.

Lines 423-433: While advantages of the approach are discussed across several paragraphs in the Discussion, this is the only paragraph considering the limitations. Please add discussion on implications of changes in training dataset (due to different selection

criteria, addition of data from different year etc.) on burned area estimates for 2019 and beyond in future application of the algorithm.

Lines 434-435: Not clear who are those "commentators" and "us" in "our ability" the authors refer to in the sentence. Please be more specific.

Lines 448-449: I do not understand why "large discrepancy for peatland burning" between the datasets would make the dataset of this study a "gold-standard"? Please explain this bold statement. Also, please consider replacing "gold-standard" with something less flashy as only time will tell how the dataset fares among users.

References

Lehsten, V., de Groot, W.J., Flannigan, M., George, C., Harmand, P. and Balzter, H., 2014. Wildfires in boreal ecoregions: Evaluating the power law assumption and intra-annual and interannual variations. *Journal of Geophysical Research: Biogeosciences*, 119(1), pp.14-23.

Lohberger, S., Stängel, M., Atwood, E.C. and Siegert, F., 2018. Spatial evaluation of Indonesia's 2015 fire-affected area and estimated carbon emissions using Sentinel-1. *Global change biology*, 24(2), pp.644-654.