Referee 1

Overall, this is an interesting topic and highlights the various limitations in burned area products. I applaud the effort that went into creating the validation dataset, however, as highlighted below, there are some crucial details lacking in the training and validation descriptions that are necessary for a full review. Please see below for specific comments.

We are grateful for the astute comments raised by Referee #1 as they have helped us clarify and improve the manuscript. Here, we answer against each point below.

Specific/Major Comments

Section 2.1: I assume the November and December 2018 images were only used for the pre-fire window analysis? How did the authors account for any burns that occurred during those 2 months?

The images taken in November and December 2018 were only used for the pre-fire compositing. Fires that occurred during these two months were not detected.

Section 2.2: What happens if a pixel burns more than once in a year? This might occur in the agricultural regions?

Our detection method does not record multiple burning events. We selected the burning event (the day of burning) when the NBR difference (dNBR) had the maximum value of all the differences calculated every two days throughout the year. The day of the year when dNBR reached a maximum value corresponded to the moment NBR dropped most markedly during the year. We flagged this day a disturbance to the pixel's vegetation, potentially caused by fire.

Line 162: Are the authors not calculating dNBR here? Also, does the moving window continue beyond the first instance of detecting a potential burn? i.e. if the drop occurred on February 1st, does the moving window continue to Dec 31 to see if it burned again?

The moving window calculated dNBR values every two days throughout the year. During this first step, we only search for the date with the highest dNBR. The detection of fires fires is done afterwards by classifying the pre- and post-composites for the date with the highest dNBR.

For example, if a fire occurred on 01 February, and the resulting dNBR recorded on this day was the maximum difference all the differences calculated (every two days) throughout the year, this day was retained as the day of the year when a disturbance to the pixel's vegetation was potentially caused by fire. If the resulting dNBR was not the maximum difference, this day (01 February) was not retained. That would be the case if another more severe fire occurred at the same pixel later in the year, which would result in a higher dNBR.

Line 177: Can the authors please provide a map of the locations of the 988 training pixels and their associated land cover types in the supplementary? Is 988 training pixels enough? How was that number decided upon?

We have included Supplementary Figure S1 to show the location of the 988 training points used to train our supervised classification algorithm (Random Forest).

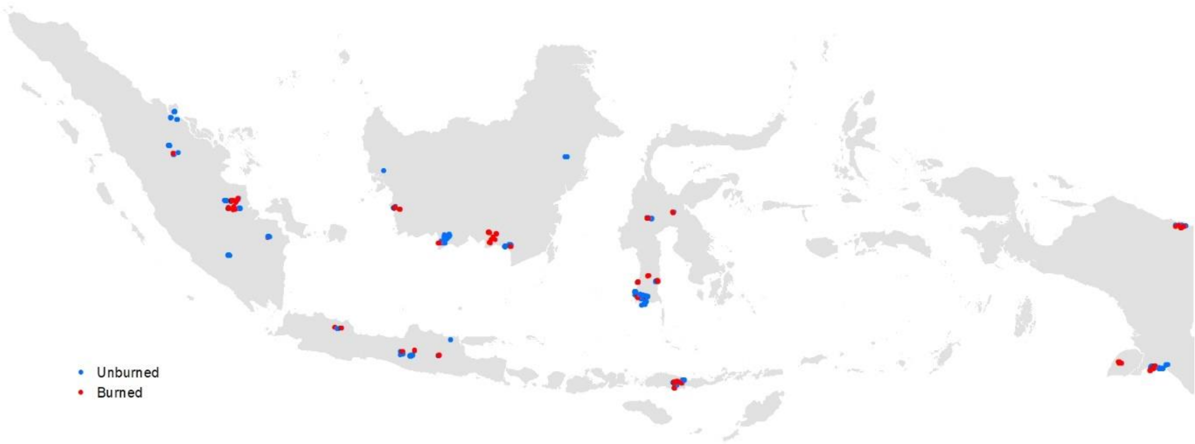


Figure S1. Location of 988 training pixels (317 'burned' and 671 'unburned') used to train our supervised classification model (Random Forest) across Indonesia (grey area).

The required number of points used to train our supervised classification model (here a Random Forest) depends on the spectral separability of the classes (in our case two classes: "class burn" and "class not-burn"). The pixels that show the burn scar present a singular spectral signature and, for this reason, it is necessary to collect a particularly high amount of training points. We collected training points until we were satisfied with the results of the classification by visual inspection. Please note that the training points differ from the validation points. They do not overlap. An adequate validation set is important as the number of validation points limits how narrow the confidence intervals are.

We added supplementary to show that the training pixels were collected in a variety of landcover types.

Table S3. Landcover types associated with the training sites one year before fire (2018) based on the ESA CCI global land cover maps described here (http://maps.elie.ucl.ac.be/CCI/viewer/index.php). The training sites were associated with 15 landcover types.

CCI 2018 Land Cover	Unburned	Burned
Cropland, rainfed	54	33
Herbaceous cover	58	44
Tree or shrub cover	29	6
Cropland, irrigated or post-flooding	6	0
Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover)	93	52
Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland	40	28
Tree cover, broadleaved, evergreen, closed to open (>15%)	132	47
Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	0	5
Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	2	12
Shrubland	9	0
Sparse vegetation (tree, shrub, herbaceous cover) (<15%)	31	39
Tree cover, flooded, fresh or brakish water	57	33
Tree cover, flooded, saline water	18	0
Urban areas	136	18
Water bodies	6	0
Total number of training sites	671	317

Line 183: dNBR already shows burn severity (here is an article with more information: https://un-spider.org/advisory-support/recommended-practices/recommended-practiceburn-severity/in-detail/normalized-burn-ratio). Did the authors quantify these values over their training pixels or simply rely on the color? I suggest the authors quantify these values to ensure the training pixels are in fact medium-to-high severity especially since the authors are prioritizing mapping high burn severity fires to reduce false positives.

Thank you for providing the link. We assessed burn severity during training based on visual interpretation. RGB composites with bands 11 (SWIR wavelength = 1.610 µm), 8 (NIR wavelength=0.842 µm) and 4 (RED wavelength = 0.665 µm) provide information about the severity of the fire; burn scars with high severity present a dark (low albedo) red/brown color. We understand that visual interpretation can be subjective. We included the histogram of dNBR ($NBR_{postfire}$ - $NBR_{prefire}$) for the 317 training points labelled 'burned' in Supplementary Figure S2 to corroborate that the training samples were selected in areas with medium to high severity fires.

81% (256) of 'burned' training points (317) had dNBR values ($NBR_{postfire}$-$NBR_{prefire}$) < - 0.44, which represents the threshold for medium to high severity burns according to the proposed classification table of the United States Geological Survey (USGS).

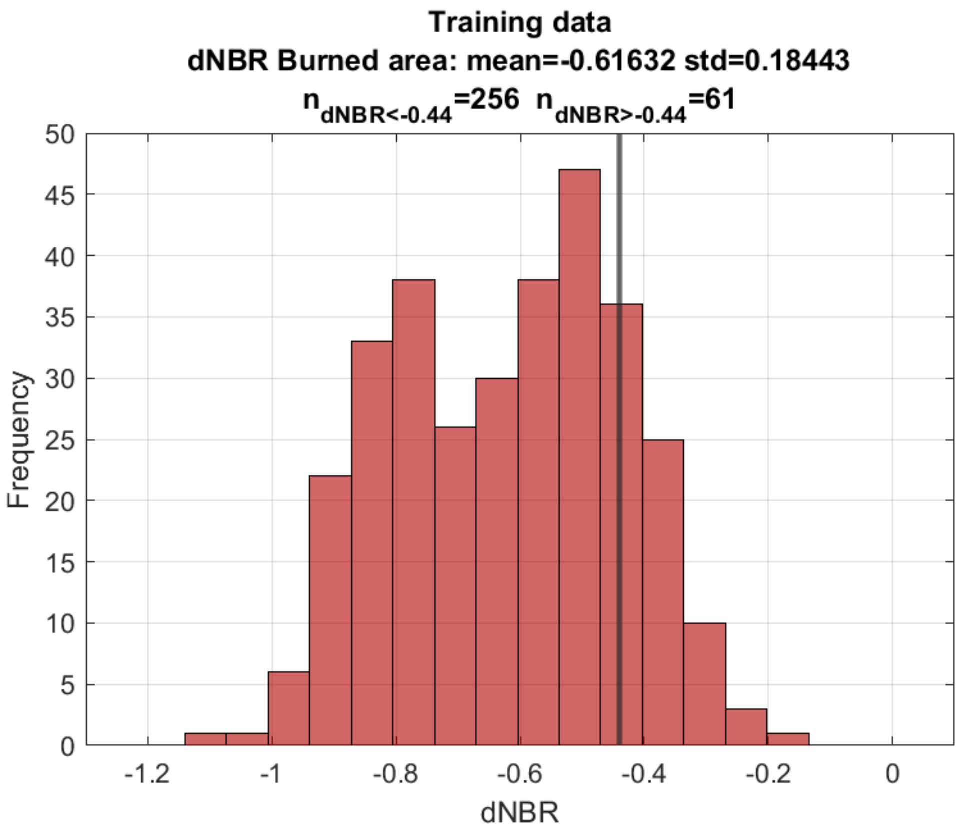


Figure S2. Histogram of dNBR ($NBR_{postfire}$ - $NBR_{prefire}$) for the 317 training points labelled 'burned'.

Section 2.4.1: As with the training samples, what land cover types were associated with the validation samples?

We added a Supplementary Table S4 to show the landcover types associated with the reference sites one year before fire (2018) based on the land cover maps described here (http://maps.elie.ucl.ac.be/CCI/viewer/index.php). The Reference sites were associated with 14 landcover types. Here the 'burned' and 'unburned' classes are the 'truth' labels deemed 'burned' by visual inspections.

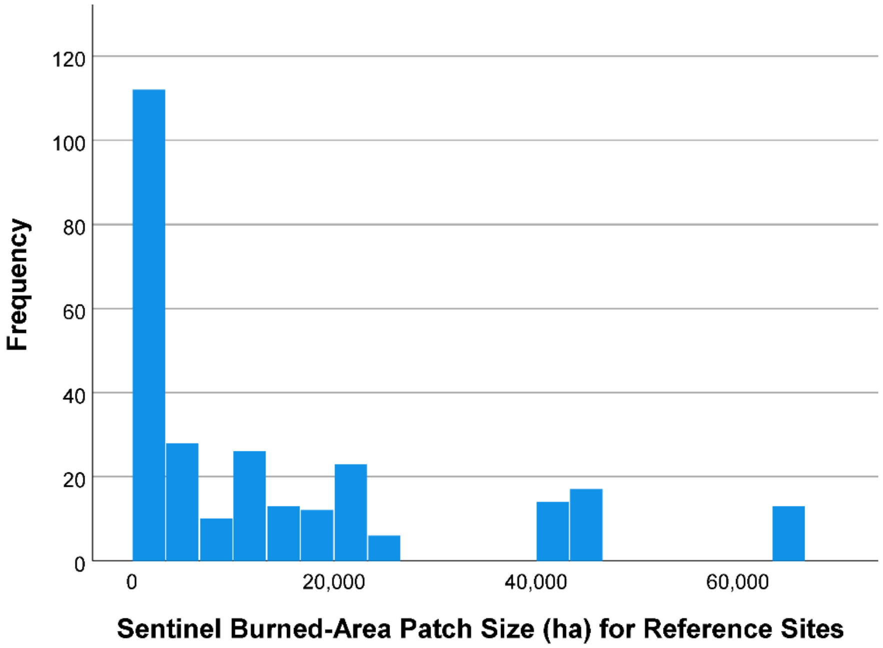
Table S4. Landcover types associated with the reference sites one year before fire (2018) based on the ESA CCI global land cover maps described here (http://maps.elie.ucl.ac.be/CCI/viewer/index.php).

CCI 2018 Land Cover	Unburned	Burned
Cropland, rainfed	16	13
Herbaceous cover	16	3
Tree or shrub cover	23	6
Cropland, irrigated or post-flooding	1	0
Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover)	127	101
Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland	150	73
Tree cover, broadleaved, evergreen, closed to open (>15%)	467	63
Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	5	6
Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	7	20
Sparse vegetation (tree, shrub, herbaceous cover) (<15%)	16	33
Tree cover, flooded, fresh or brakish water	94	31
Tree cover, flooded, saline water	19	4
Urban areas	1	0
Water bodies	3	0
Total number of reference sites	945	353

Secondly, what size burn scars were these validation pixels associated with? For example, if all validation pixels were associated with very large burn scars then the validation results will be biased because large burns are easy to detect. Also, I assume the training and validation samples did not overlap?

Our reference sites (i.e., validation pixels')'were associated with a wide range of burn scar sizes. The uppermost histogram in the Figure below (added as Supplementary Figure S3) shows the frequency distribution of Sentinel-2 burn scar sizes for scars coincident with a subset of our 1298 reference sites used to validate our Sentinel burned-area map and deemed 'burned' by visual inspections. As seen, the patches coincident with these reference sites range from very small (a few hectares) to very large (over 60,000 ha). This diversity of patch size is to be expected considering that reference-site sampling was realised randomly across the entirety of burned areas, without regard to patch size, with the partial exception that patches <6.25 ha were excluded from consideration for reasons noted in the main text. Correspondingly, the positive skew of the reference-site histogram is in keeping with the positively skewed frequency distribution of all Sentinel burned-area patches for Indonesia, shown in the lower histogram of the Figure below.

Notwithstanding the points above, the frequency distributions of the upper and lower histograms in Figure are ultimately statistically different from one another, insofar as the distribution for the reference sites is *comparatively* biased towards larger patches. The most likely reason for the statistical difference in question is that the hyperabundance of very small patches in our Sentinel-2 burned-area map would require an exceptionally large sample of reference sites to fully represent such small patches alongside a proportional diversity of intermediate and larger patch sizes.



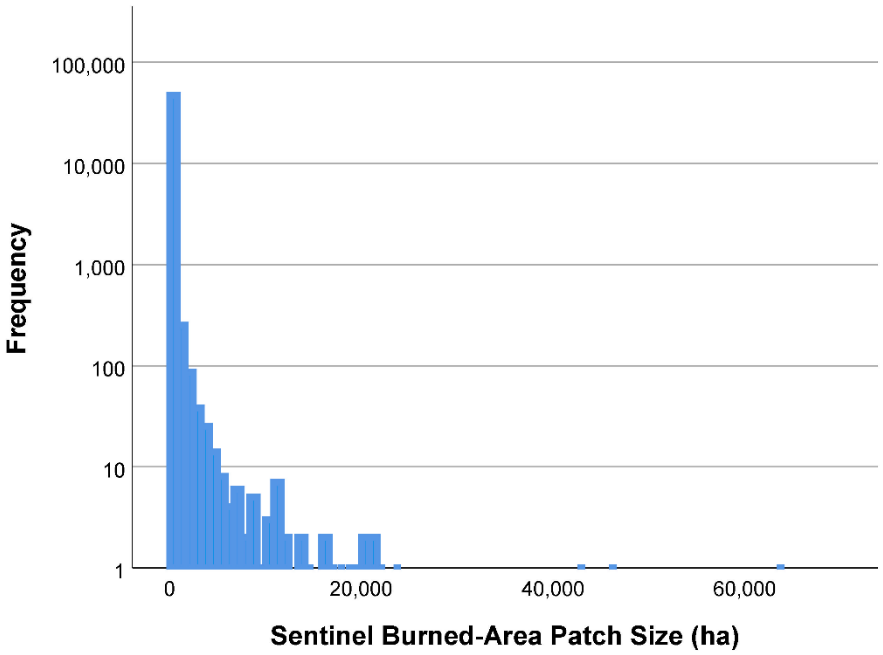


Figure S3. Frequency distributions of patch sizes of the Sentinel burned-area map, for select spatially coincident reference sites used to validate the Sentinel burned-area map (top), and for all of Indonesia (bottom). Note: Bin widths are not consistent between upper and lower panels. In the lower panel, note the logarithmic scale of the y-axis and the presence of rare patches above 40,000 ha. Patches <6.25 ha are excluded. Reference sites are those 274 sites deemed 'burned' by visual inspections (labelled as 'truth') and coincident with Sentinel-2 burn scars.

Finally, we affirm that our training and validation samples were independent and did not overlap.

Line 310-311: Please explain why the authors only chose the cardinal directions?

The choice to define burned-area patches in terms of pixel contiguity in the cardinal directions but not on diagonals was intended to render the resultant burned-area map conservative with respect to patch size. Given that the Sentinel data has a relatively fine spatial resolution (10 m), minor sub-pixel burning within a single 'burned' pixel could conceivably link two much larger burned areas into a single discrete patch if diagonal contiguity were recognized. Such an outcome would, in our view, potentially inflate overall patch-size estimates, perhaps especially for smaller-scale patches for which burning often adopts 'patchy' spatial patterns. Of course, the same issue also applies to a single pixel contiguous in a cardinal direction. However, our preliminary inspections of the geography of burning in our Sentinel burned-area map suggested that 'undo' contiguity along individual diagonal pixels was more common and/or potentially problematic. We revised the text to justify our choice accordingly.

Line 389: While doing some reading into power laws and fire size, I came across this paper from the US Forest Service with the following quote: *"Newman (2005) specifically excludes fire size distributions, while admitting that they might follow power laws over portions of their ranges. Current opinion is divided among those who would globally assign power laws to fire-size distributions (Minnich 1983; Bak et al. 1990; Malamud et al. 1998, 2005; Turcotte et al. 2002; Ricotta 2003) and those who would attribute them only to portions of distributions or rule them out altogether in favor of alternatives (Cumming 2001; Reed and McKelvey 2002; Clauset et al. 2007; Moritz et al., Chap. 3)"* - https://www.fs.fed.us/rm/pubs_other/rmrs_2011_mckenzie_d001.pdf

Please can the authors go through the literature and ensure their power-law assumption is correct and justify it in the paper.

We are not making any theoretical claims that fire size distribution follows a power law. We just note that this pattern has been observed by other studies, and is observed over a greater range of scales in our refined burned area analysis as we would expect if these methods are better able to detect burns (which is our main point here, and why this emphasis is helpful). Indeed, as we note in the discussion the comparisons also highlight how the detection of these patterns depends on the nature of the methods used to detect them, which is something that is not appreciated in the published literature around this theory. We revised our text accordingly.

Line 399: The current analysis does not support this finding regarding agricultural burning. Based on Figure S3, the small patches are likely associated with the small burn patches surrounding the larger burn scars. Agricultural burning is a very difficult fire type to map and although the current methodology is likely to map more agricultural burning than MODIS (due to the finer resolution) the mapping methods and validation assessment was not adequately designed for agricultural burning. The authors can mention that the S2 mapping is better suited for identifying "small fires". Furthermore, it was noted on lines 427 – 433 that the approach omitted hard–to–detect fires (e.g. savanna grasslands) which are much easier to detect than agricultural fires therefore that statement is not supported.

We rephrased accordingly: "Our estimate is the most reliable and accurate and therefore captures more of the 2019 total burned area, 399 confirming that 20-m Sentinel-2 imagery is better suited to widespread small-scale ~~agricultural~~ burning in Indonesia

Minor Comments

Line 23: change to "..which occur on.."

We rephrased accordingly.

Line 27: Should the size of the intermediate fires read (100ha – 1000ha) similar to what you have in the main body of the paper?

Thank you for noting this error. We have corrected it.

Line 88: change “excepting” to “except”

Thanks, changed

Introduction: When are the peak burning months? It seems based on the GWIS country profiles (https://gwis.jrc.ec.europa.eu/apps/country.profile/charts) that the peak occurs August – October and since the authors are also referring to agricultural fires then please also include the cropland burning months.

We clarified this point in the text in the introduction. We wrote: “Many fires are lit by farmers and companies when conditions permit to wood burn debris after deforestation, and enrich the soils before planting, or to maintain existing agricultural land (paddy farm fallow). Burning occurs throughout the year, but generally more often during July to October (dry months) especially in parts of Indonesia where rainfall is unimodal. “

Line 137: Remove “and finally” after “Fourth. The authors go on to a final step on line 139.

Fixed

Figure 3: There is no h panel (line 679)

Thank-you for spotting this mistake. We fixed it

Line 237: Should that be referencing Figure 3? Also, there is no panel h

Thank-you for spotting this mistake. It should read Figures 3e,f,g

Line 696: change to “minimum mapping unit”

Thanks. Fixed

Figure 5: Please add to the caption that the light grey represents countries outside of Indonesia. This was confusing at first before I looked at a map.

We removed the grey areas outside of Indonesia to improve the clarity of Figure 5. We also removed the outline of ‘burned’ polygons during the display of the burned area shapefile used to generate the map because the outline display exaggerated the burned area extent visually. The new Figure 5 is below.

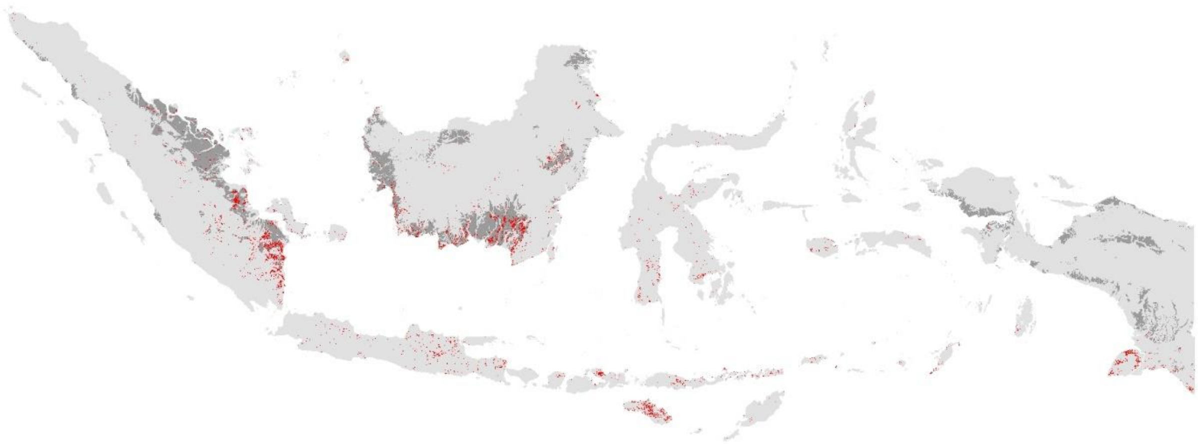


Figure 5. 2019 burned areas (red) for Indonesia (grey) derived using a time-series of the atmospherically corrected surface reflectance multispectral images (level 2A product) taken by the Sentinel-2 A and B satellites. The spatial resolution of this map is 20 m x 20 m, and minimum mapping unit is 6.25 ha. The officially recognized peatlands extent is shown with the darkest shade of grey. A provincial breakdown of burned areas according to our map estimates and those of the Official and the MCD64A1 product are given in Figure S1.

Table S3: please add the meanings of Am and Wh to the caption

We added meanings of Am and Wh in Figure caption

Table S3. Confusion matrix. Am = Area mapped (the area classified as class i by the Random Forest; the sum of this column is equal to the total area of study). Wh = Proportion of area mapped (the proportion of area classified as class i; the sum of this column equals to 1)

Line 370: add a comma between "Figure 6 Figure S2"

Fixed

Line 702: Change to "MCD64A1". There were a few other instances where the A was lowercase (i.e. MCD64a1)

Fixed

Line 377: It would be interesting for the authors to create a 3-panel figure showing this scar from each of the 3 products to show the omissions made in MCD64 and the Official dataset.

We added a 6-panel Figure in the main text (Figure 7) to illustrate this case.

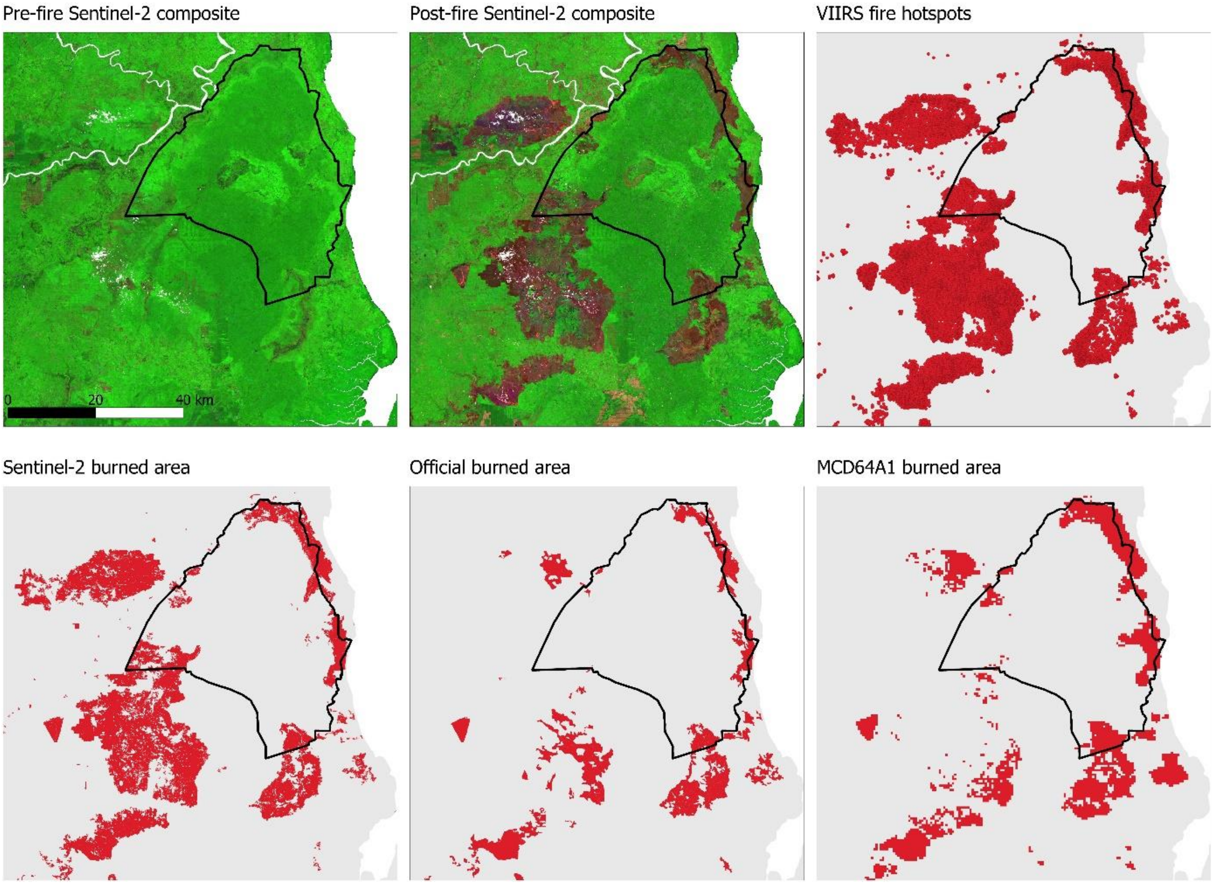


Figure 7. The pair of cloud-free pre-and post-fire Sentinel-2 composites over Berback National Park (black line) and surrounding areas in Jambi Province (see also Inset A, Figure 1), revealing large, burned areas around Berbak National Park (areas that have transitioned from 'green' to dark 'brown/red' tones). These large burn scars have been detected by VIIRS hotspots and by the Sentinel-2 burned area map, but some have been missed by the Official and MCD64A1 datasets.

Line 443: change to "addressing"

fixed