Refined burned-area mapping protocol using Sentinel-2 data increases estimate of 2019 Indonesian burning

David L.A. Gaveau[1] Adrià Descals[2], Mohammad A. Salim[1], Douglas Sheil[3], Sean Sloan[4]

[1] TheTreeMap Bagadou Bas 46600 Martel, France

[2] CREAF, Centre de Recerca Ecològica i Aplicacions Forestals, E08193 Bellaterra (Cerdanyola de Vallès), Catalonia, Spain

[3] Forest Ecology and Forest Management Group, Wageningen University & Research, PO Box 47, 6700 AA, Wageningen, The Netherlands

[4] Department of Geography, Vancouver Island University, Nanaimo, BC, Canada &

Correspondence to: David Gaveau (d.gaveau@thetreemap.com)

Abstract

Many nations are challenged by landscape fires. A confident knowledge of the area and distribution of burning is crucial to monitor these fires and to assess how they might best be reduced. Given the differences that arise using different detection approaches, and the uncertainties surrounding burned-area estimates, their relative merits require evaluation. Here we propose, illustrate, and examine one promising approach for Indonesia.

Drawing on Sentinel-2 satellite time-series analysis, we present and validate new 2019 burned-area estimates for Indonesia. The corresponding burned-area map is available at: https://doi.org/10.5281/zenodo.4551243 (Gaveau et al., 2021). We show that >3.11 million hectares (Mha) burned in 2019. This burned-area extent is double the Landsat-derived Official estimate of 1.64 Mha from the Indonesian Ministry of Environment and Forestry, and 50% more that the MODIS MCD64A1 burned-area estimate of 2.03 Mha. Though we observed proportionally less peatland burning (31% versus 39% and 40% for the Official and MCD64A1 products, respectively), in absolute terms we still observed a greater area of peatland affected (0.96 Mha) than the official estimate (0.64 Mha). This new burned-area dataset has greater reliability as these alternatives, attaining a user's accuracy of 97.9% (CI: 97.1%-98.8%) compared to 95.1% (CI: 93.5%-96.7%) and 76% (CI: 73.3%-78.7%), respectively. It omits fewer burned areas, particularly smaller- (<100 ha) to intermediate-sized (100 ha -1000 ha) burns, attaining a producer's accuracy of 75.6% (CI: 68.3%-83.0%) compared to 49.5% (CI: 42.5%-56.6%) and 53.1% (CI: 45.8%-60.5%), respectively. The frequency–area distribution of the Sentinel-2 burns follows the apparent fractal-like power-law or "pareto" pattern often reported in other fire studies, suggesting good detection over several magnitudes of scale. Our relatively accurate estimates have important implications for carbon-emission calculations from forest and peatland fires in Indonesia.

1. Introduction

Accurate burned area maps are key to characterizing landscape fires, clarifying emissions and identifying the probable causes. Such information is needed to target interventions, to assess policies and practices intended to reduce or control fires, such as law enforcement and restoration of fire-prone degraded lands, and to measure progress to international climate commitments (Sloan et al., 2021). Here, we focus on Indonesia where recurring forest and peatland fires have become an international crisis (Tacconi, 2016). These concerns arise from the large carbon emissions associated with these fires, and the impact of associated aerosol emissions for human health in the wider region (Van der Werf et al., 2008;Marlier et al., 2013). Although fires have occurred locally in Southeast Asia for millennia, they are increasingly frequent in Indonesia's disturbed forests and deforested peatlands (Field et al., 2009;Gaveau et al., 2014). The causes and motivations of fire use can be complex (Dennis et al., 2005), but many are lit to create or maintain agricultural land (Gaveau et al., 2014; Adrianto et al., 2020). Most fires occur during drier months (July to October) and the threats are greatly heightened during years of anomalously low rainfall (Sloan et al., 2017;Field et al., 2016). During 2015, a strong El Niño-induced drought year, fires burned an estimated 2.6 million hectares according to official estimates (Sipongi, 2020). Although 2015 burning was approximately half as extensive as 1997, the most severe El Niño and fire season on record (Fanin and Werf, 2017), about 50% more peatlands burned (Fanin and Werf, 2017). The 2015 fires emitted between 0.89 and 1.5 billion tons of CO_2 equivalent (Huijnen et al., 2016; Lohberger et al., 2018; Van Der Werf et al., 2017), representing about half of Indonesia's greenhouse gas emissions for that year (Gütschow et al., 2019). In Palangkaraya, the capital city of Central Kalimantan province, daily average particulate matter (PM_{10}) concentrations often reached 1000 to 3000 μg m^{-3} amongst the worst sustained air quality ever recorded worldwide

(Wooster et al., 2018). Over half a million people suffered respiratory problems in the aftermath, and between 12,000 and 100,000 premature deaths were estimated (Koplitz et al., 2016;Crippa et al., 2016). Other impacts include loss and degradation of habitats with high conservation values, and the associated consequences for impacted wildlife (Harrison et al., 2016).

In response to the catastrophic 2015 fires, the Indonesian government instituted several ambitious schemes including fire bans enforced by dedicated command posts (Sloan et al., 2021) and an ambitious national program of peatland restoration (Carmenta et al., 2020). Despite the investment in these approaches and measures, and initial success, severe burning struck Indonesia again in late 2019. While Sloan et al. (2021) suggest that 2019 fire activity was lower than expected given the severe drought conditions, the total number of MODIS active-fire detections in late 2019 on peatlands was still amongst the greatest recorded since 2001 (Sloan et al., 2021). However, counts of active-fire detections don't provide estimates of area burned (Tansey et al., 2008) and for 2019 such estimates remain uncertain.

Those wishing to assess and monitor burned areas have various approaches to consider. Several global burned area products generated using coarse-resolution satellites (>250 m) can be applied over Indonesia. These include the FireCCI41 product derived from Envisat-MERIS (Alonso-Canas and Chuvieco, 2015), the FireCCI51 and MCD64A1 products derived from TERRA&AQUA-MODIS (Giglio et al., 2018; Lizundia-Loiola et al., 2020), the FireCCILT10 product derived from AVHRR (Otón et al., 2019) and the C3SBA10 product derived from Sentinel-3 (Lizundia-Loiola et al., 2021). Currently, the MCD64A1 (collection 6), based on MODIS 500 m bands, is considered one of the most accurate global product (Chuvieco et al., 2019), with omission and commission errors of 40% and 22% globally for the 'burned' class (Giglio et al., 2018). This validation is based on independent globally distributed visually interpreted reference satellite data, however none over Indonesia. These coarse-resolution datasets generally omit small-scale fires and, thus, the reported burned area is underestimated (Ramo et al., 2021). This has motivated research in the use of medium-resolution satellites (10 to 30 meters) such Sentinel-1 (Lohberger et al., 2018 in Indonesia), Sentinel-2 (Chuvieco et al. 2018 in Sub-Saharan Africa), and the Landsat constellation (Hawbaker et al., 2017 in North America) to produce more detailed burned area maps. Lohberger et al. (2018) reported 4.6 Mha burned in 2015 in Indonesia, nearly double the estimate of 2.6 Mha from the Indonesian Ministry of Environment and Forestry (MOEF), using visual interpretations of time-series Landsat-8 imagery (Sipongi, 2020).

For year 2019, The MOEF (hereafter 'Official estimate') estimated that 1.64 Mha burned in 2019 (Sipongi, 2020), while the MCD64A1 (collection 6) indicated 2.03 Mha burned in 2019. The coarse 500-m spatial resolution MCD64A1 data omit smaller fires and thus likely overlook many localized events. The Landsat imagery underlying the Official estimates are, while finer scale, observed every 16 days at best (typically much less due to cloud and smoke), meaning that many burns may remain undetected. Also, smaller-scale and/or dispersed fire activity may be underestimated, considering the challenges of their visual interpretation and delineation. Visual interpretation entails a manual delineation of burns perimeters, which yields accurate results for large burn mapping at local scales, but is very time consuming at large spatial scales, particularly when mapping small fires. A thorough accuracy assessment is also not available for the official burned-area products. Given the unknown errors around burned-area estimates, and the differences between them, the accuracy, and merits of different mapping approaches over Indonesia require formal examination.

Here, we present new and validated 2019 burned-area estimates for Indonesia using a time-series of the atmospherically corrected surface reflectance multispectral images (level 2A product) taken by the Sentinel-2 A and B satellites. With higher spatial resolution (20-m) and more frequent observations (5-day revisit time), the Sentinel-2A and B satellites offer relatively comprehensive and accurate burned-area mapping (Huang et al., 2016; Ramo et al., 2021). We used the Google Earth Engine (Gorelick et al., 2017), thus permitting wide application. We also developed an independent reference dataset to compare the accuracy of our estimate against the Official and MCD64A1 burned-area maps. Given the lack of objectively distributed ground truthing, we sought ways to extract reference sites by visually detecting a smoke plume, burn, or heat source (flaming front, or hotspot) from the archive of original Sentinel-2 images. Finally, we examined differences in terms of burn-size frequency distributions among these three burned-area estimates to examine spatial patterns.

2. Methods

2.1. Summary of methods

A burned area is identified by alteration of vegetation cover and structure along with deposits of char and ash. We mapped such areas using a change-detection approach, i.e. by comparing Sentinel-2 infrared signals recorded before and after a burning event (Liu et al., 2020). We analyzed a time-series of the Normalized Burned Area Ratio (see section 2.2) to assemble two national composite images depicting the spectral condition of vegetation shortly before and shortly after a disturbance (Figure 1). These composites represent a convenient way to capture the entire burned landscape stored in just two image files. Although we refer to these images as "pre- and post-fire composites", they also capture damage due to other causes, for example a cutting event (e.g. mechanical conversion to agriculture, to timber plantation, to roads, population centers, mining or natural timber harvesting), a disease, strong winds, floods, or landslides (Gaveau et al., 2021). After the production of the pre- and post-fire composites, we used a "Random Forest" classification model (see section 2.3) trained on visually identified pairs of pre- and post-fire pixels to confirm if the spectral changes indicating vegetation damage corresponded to a burning event. Third, three independent interpreters assembled a reference dataset by visually identifying burns in the original time-series Sentinel-2 images. Fourth, we assessed our burned-area map, as well as the Official and MCD64A1 burned-area maps, against our reference dataset to gauge the reliability and accuracy of the three burned-areas products. Finally, we tested whether, and how, the three burned-area estimates differed in their tendencies to incorporate burns of different sizes.

2.2. Pre- and post-fire Sentinel-2 national composite images of 2019

Here, we describe our automated procedure to create a national pair of pre- and post-fire composites from 47,220 original Sentinel-2 images acquired between 01 November 2018 and 31 December 2019. Prior to creating the composites, we removed non-valid pixels using the Sentinel-2 imagery quality flag (this flag provides information about clouds, cloud shadows, and other non-valid observations) produced by the ATCOR algorithm and included in the atmospherically-corrected surface reflectance multispectral images of the Sentinel-2 A and B satellites Surface Reflectance products (Level 2A product) (Fletcher, 2012).

A time series of the Normalized Burned Ratio (NBR), given as (NIR-SWIR) / (NIR+SWIR), represents a convenient index to detect the approximate day when the vegetation was damaged. Before damage, vegetated pixels register high NBR values close to 1 because reflectance in near-infrared spectrum (NIR; wavelength=0.842 μm; Band 8) is high due to the chlorophyll content of the vegetation (open circles before a disturbance, in this case a fire, in Figure 2). The NBR of damaged vegetation typically declines abruptly towards 0 (or ≤ 0 for severe damage) because the NIR reflectance declines due to chlorophyl and leaf destruction, while the reflectance of short-wave-infrared spectrum (SWIR; wavelength = 1.610 μm or 2.190 μm; Band 11 or Band 12) increases due to dead or charred material and exposed ground cover. NBR values ≤ 0 are often apparent for several weeks after severe burning or clear-cutting. We analyzed NBR time series for approximately 94.5 billion 400 m^2 pixels (Indonesia's landmass =198 Mha). We describe the procedure to detect drops in the NBR time series in the following paragraph.

We detected drops in NBR time series with a moving-window approach. A moving window scanned NBR values three months prior and one month after the central day of the window. The output value of the moving window (blue dots in Figure 2) is the difference between average NBR values observed before and after the central day. The NBR average after the central day included the value at the central day. The difference between the average NBR values was estimated every 2 days in the time series, skipping the day of year that was an odd number (day of year equal to 2, 4, 6, 8...). Although Sentinel-2 has a temporal resolution of 5 days, the overlap between satellite passes may increase the temporal resolution regionally up to 2 days at the equator. Thus, we estimated the NBR difference (dNBR) every 2 days instead of 5 days. Taking this into consideration, our 'disturbance' date estimate has a maximum temporal precision of 2 days in specific regions, but generally 5 days when satellite passes do not overlap. The day of the year when dNBR reached a maximum corresponded to the moment NBR dropped most markedly in each pixel, flagging a disturbance to the pixel's vegetation potentially caused by fire. At this date, we created a pair of pre- and post-fire pixels by selecting the median Red, NIR and SWIR spectral values acquired three months before and one month after the disturbance. We selected a one-month window rather than a three-month window to compute the post-fire image to maximize our chances to detect recent burns, given that burned areas on degraded lands and savanna tend to re-green rapidly. We repeated this procedure for approximately 94.5 billion pixels to assemble two national composite images depicting the spectral condition of vegetation shortly before and shortly after a disturbance (Figure 1).

2.3. Supervised burned/unburned classification.

We used the Random Forest supervised classification algorithm (Breiman, 2001), available via the Google Earth Engine to determine whether the spectral changes detected by the pre- and post-fore composites corresponded to a burning event, and subsequently classify burned areas. Supervised classifiers require 'training data', that is, exemplary spectral signatures of 'burned' and 'unburned' lands in the present case, to guide the algorithm to reliably classify the target classes. The spectral signatures (i.e., the reflectance values in the pre- and post-fire composite images) are the predictive variables of the classification model. The features used in the Random Forest are the bands of Sentinel-2 in the pre- and post-fire composites plus their respective NBR index. We excluded the bands at 60-meter spatial resolution (bands B1, B9, and B10) since these bands present a low spatial resolution for the aim of the study. Therefore, we used a total of 22 features; the NBR and bands B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12 of the pre and post-composites.

We used a 10-fold cross-validation to assess the accuracy obtained with a set of different parameters in the Random Forest. The splitting 'train-test' in the cross-validation was done only with the training dataset, since the reference dataset used for the final validation must be completely independent of the training and model parametrization. The two parameters that we tuned were the *number of trees* and the *minimum leaf size*. Random

Forest is an ensemble classifier composed of several Decision Trees; the parameter *number of trees* represents the number of Decision Trees in the Random Forest. The *minimum leaf size* represents the minimum number of samples that result from a splitting node at the Decision Tree. We found that a *minimum leaf size* equal to 1 performed the best on average and, thus, we used this value. We selected a conservative *number of trees*, 50, to ensure the good performance of the Random Forest. We did not set any limit to the maximum nodes in each tree and the variable to split in the random forest was set to the square root of the number of variables, which is the common practice among machine learning practitioners and the default configuration in Google Earth Engine.

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: "burned" and "unburned"). The pixels that show the burn present a singular spectral signature and, for this reason, it is necessary to collect a large amount of training points. We collected training points until we were satisfied with the results of the classification by visually comparing the resulting burned area map against the pre- and post-fire composites. We trained the Random Forest algorithm using 988 independent training pixels (Supplementary Figure S1 for locations), being point coordinates labelled as either 'burned' (317 points) or as 'unburned' (671 points). These pixels were selected by visual interpretation of the pre- and post- fire image composites. Burned areas show a distinctive dark (low albedo) brown/red color in the SWIR-NIR-Red composite image when displayed as Red-Green-Blue channels (Figure 1). The training pixels were collected across landcover types (Supplementary Table S1 for landcover types) to ensure the representativeness of the training dataset and the satisfactory generalization of the classification model across Indonesia. We selected training pixels focused explicitly on medium-to-high burn severity, i.e. areas where the distinctive red color in the SWIR-NIR-Red composite image looked the darkest, indicating that all or most of the vegetation/soil burned. This aspect of the method minimized "false positives" but may exclude areas with implied low-burn severity or low-visibility impacts, such as understorey fires (below an intact forest canopy, see e.g., van Nieuwstadt and Sheil, 2005. By prioritizing confident identification of fires over absolute burned-area coverage, as well as by duly validating our estimates, this approach avoids the problems caused by frequent false positives (Rochmyaningsih, 2020).

We assessed burn severity during algorithm 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 with high severity present a dark (low albedo) red/brown color (Figure 1). 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 'burned' training samples were selected in areas with medium to high severity fires. Eighty one percent (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).

2.4. Burned-area map validation.

The Gold standard is to validate the map against a sufficiently large reference dataset developed based on ground visits to 'burned' and 'unburned' sites sampled objectively and randomly across the region of interest (Olofsson et al. 2014). We sought alternative ways to generate the reference dataset because the sample of GPS locations of 'burned' locations collected by Indonesian government were not available. Given the laborious scale of this validation exercise, we validated our burned-area estimates for only the seven provinces prioritized by the

Indonesian Government for restoration of fire-prone degraded lands (Kalimantan Barat, Kalimantan Tengah and Kalimantan Selatan, Papua, Jambi, Riau, and Sumatra Selatan). These provinces are also those that typically burn most extensively. We used visual interpretations of the original time-series Sentinel-2 imagery acquired every 5 days over 2019 at 1298 randomly selected sites (one site = one pixel of 20 m x 20 m) to detect flaming fronts (fire hotspots) and other signs of burning (smoke and charred vegetation). We used these reference data to calculate the overall accuracy (OA), producer's accuracy (PA), and user's accuracy (UA) with a 95% confidence interval, of all three burned area maps (i.e., our Sentinel-derived burned-area classification, the official Landsat-based burned-area map, and the MCD64A1 product) following "good practices" for estimating area and assessing accuracy reported by Olofsson et al. 2014. We use the term '*mapped burned-area'* for the area classified as burned by each burned-area map. We employ the term '*corrected burned-area'* for the estimation of the burned area based on the validation of a given burned-area map against the reference dataset, following the practices in Olofsson et al. 2014. For instance, a high omission rate in the 'burned' class of a given burned-area estimate would potentially lead to a lower *mapped area* than a *corrected area* for that estimate, while a high commission rate would potentially lead to a higher *mapped area* than the *corrected area*. The *corrected area* represents an estimation of the actual burned area for year 2019 computed for each of the three datasets separately. The accuracy of the burned area map, and the sample size of the reference dataset, play a role in the confidence interval of *corrected area* estimate. Lower map accuracy and smaller sample size mean wider confidence intervals.

2.4.1. Reference site sampling design

Good practices for estimating area and assessing accuracy, as reported in Olofsson et al. (2014), assumes a simple random sampling or a stratified random sampling in the generation of the reference dataset. In our case study, we employed a stratified-random sampling approach to ensure an acceptable sample of 'burned' reference sites. Our stratified approach was necessary given that the 'burned' class was rare over the study area: the area of seven provinces of interest is 87.6 Mha and the combined area detected as burned by all three datasets represented only 3.1% of this area.

For the generation of the 1298 reference sites (see Supplementary Table S4 for associated landcover types one year before fire), we randomly sampled (i) 419 sites across from the areas classified 'burned' by the three datasets (red area in Figure 3a; Supplementary Table S2), and (ii) 879 sites in areas classified as 'unburned' by all three datasets hereafter denoted U (grey area in Figure 3a). This sample size is deemed sufficient and comparable to other map assessments at larger scale (Stehman et al., 2003;Olofsson et al., 2014).

This initial sample of 1298 total sites present a shortcoming for direct pair-wise comparisons of between the reference dataset and each of the three burned-area maps individually. Specifically, sampling densities in the reference dataset were far greater in areas classified 'burned' by the three datasets (red area in Figure 3a) compared to the area deemed 'unburned' by all three datasets, hereafter denoted U (grey area in Figure 3a). Consequently, for the validation of a given burned-area dataset, its total number of 'unburned' reference sites would be over-sampled upon defining 'unburned' reference sites with reference to U as well as areas classified as burned uniquely by one of the other two maps (cyan areas in Figure 3b, c, d, hereafter denoted as U'). Such over-sampling of reference sites in the realm of U' would violate the stratified-sampling approach described in Olofsson et al. (2014) and would lead to an erroneous accuracy assessment. To achieve a balanced stratified sampling of reference sites across 'burned' and 'unburned' areas of each dataset, we generated three subsamples from the initial 1298 reference sites (red areas in Figures 3e,f,g) and used these subsamples to validate each dataset. These three subsamples were generated by randomly excluding reference sites from the realm of U' in Figure 3b, c and d, respectively, until the density of reference sites in U' equaled the density of the larger unburned area U. For instance, for the validation of the Official burned-area map, the density of reference sites in U was 10.36 sites/Mha, and the extent of U' was 1.551 Mha, such that the number of reference sites to retain in U' for this validation was given as 1.551 Mha x 10.36 sites/Mha =16 sites. The calculations of the number of sites removed from each subsample are illustrated in Supplementary Table S3. The final, adjusted, stratified subsamples of reference sites used for validation is given in Table 1.

2.4.2. Interpretation of the burned-area reference dataset

We developed a series of scripts in the Google Earth Engine to streamline the visual interpretation of the reference sites. Specifically, we adapted a script written by (Olofsson et al. 2014) to rapidly scan the time-series of original Sentinel-2 images in visible and infrared bands and thus visually detect either a smoke plume, a burn, or a heat source (flaming front), and determine whether and when in 2019 a reference site burned. The script enabled the interpreter to interactively track the evolution of NBR values and patterns over the 2019 time series of 5-day images. Reference sites were investigated for burning wherever a marked drop in the NBR time series was detected, indicating a disturbance in the vegetation. For reference sites where a disturbed area was observed, we subsequently reviewed the last few images before the drop in NBR and the first few images after the drop. Interpreters looked for three distinct signs of burning in these images to confirm them as burned: (i) smoke plumes; (ii) flaming fronts – that is, a line a moving fire where the combustion is primarily flaming; and (iii) rapid changes in color from 'green' to 'dark red', characteristic of a transition to charred vegetation (Figure 4). If rapid changes in color were observed over the reference site, with at least one direct feature (smoke or flame) in its vicinity, this indicated a fresh burn, and the reference site was declared 'burned'. If rapid changes in color from 'green' to 'dark red' were observed without smoke or flame, the reference site was also declared 'burned'. If no change in color was observed, with at least one direct feature (smoke or flame) in its vicinity, the reference site was declared 'unburned'. If none of these three features were observed, the reference site was declared 'unburned'.

Three interpreters independently reviewed the time-series of original Sentinel-2 images and associated NBR trends for all reference sites (N=1298) (see Supplementary Figure S3 for a frequency distribution of burn sizes of the Sentinel-2 burned-area map, for select spatially coincident 'burned' reference sites). To reduce uncertainties associated with the interpretation of the imagery, the results of the three interpreters were compared to each other. If all three interpreters recorded the same interpretation and timing of a burning event for a given reference site, their interpretations were retained. If one or more interpreters disagreed, all interpreters reviewed the data and resolved discrepancies by consensus. In some cases, it was difficult to reconcile disagreements because of poor image quality or because of uncertain spectral patterns. Therefore, if possible, interpreters also explored other satellite images (e.g. Landsat) to detect the presence of fire and resolve disagreements for a given reference site. The sites in which the three interpreters disagreed were ultimately excluded (70 sites) from the reference dataset. For these excluded sites, disagreement typically resulted from uncertainties over the boundary of burned or unburned areas, or because the imagery was not clear enough. The sample size of reference points explored here, N=1298, excludes the discarded points of disagreement in question.

We created a second script to generate snapshot images (see examples in Figure 4) depicting infrared spectral conditions, shortly before and shortly after a fire, as well as the corresponding image dates. Interpreters recorded and geotagged a snapshot of before and after fire condition at every reference site (for which a burned area was detected) to enable third-party reviewers to check the consistency and validity of interpretations on site-by-site basis (See Data Availability).

2.4.3. Burn size comparisons.

We tested whether, and how, the three burned-area estimates differed in their tendencies to incorporate burns of larger or smaller sizes. Specifically, we compared the frequency distributions of burn areas (or "scars") amongst the three estimates to test for similarity and qualify any distinguishing differences on the part of our Sentinel-based estimate. Differences amongst burn size frequency distributions implies that a given burned-area estimate is inclusive of burn of a given size, regardless of absolute differences to total burned area between the estimates. Inter-estimate comparisons of burn-scar size frequency is analogous to tests of whether the 'samples' of burns defined by each estimate describe the same, ultimately partially-observed universe of fire activity. Significant inter-estimate differences imply greater or lesser inclusion of a given realm of fire activity – e.g., small-scale agricultural burning, plantation fires, extreme wildfires – thus indicating bias (or lack thereof) without defining such realms explicitly.

For all three estimates, we employed the Kruskal-Wallis H test of differences with respect to the 'location' of frequency distributions along a continuum of burn sizes. Given significant inter-estimate differences according to this three-way test, we tested for two-way differences in the shape and location of the burn-size frequency distribution (Kolmogorov-Smirnov test), as well as two-way differences in medians (Mann-Whitney U test), between our Sentinel estimate and either the Official or MODIS estimate individually. Testing for similarity over increasingly large scar-size cohorts clarified the degree to which significant inter-estimate differences were attributable to the inclusion or omission of a given cohort.

We excluded burns <6.25 ha because this is the minimum observable burn-size of the Landsat-8 Official estimates due to the challenging nature of visual interpretations at such scales. We note that the minimum size of the MODIS data is 25 ha, hence for comparison with MCD64A1 product we used a 25-ha threshold. In relation to Sentinel and MODIS estimates, for which burned areas were originally mapped as arrays of pixels, we defined a burn to be any array of pixels contiguous across cardinal directions but not diagonals to render the resultant burned-area map conservative with respect to patch size (Figure S4). For the Official estimate, burns are as manually delineated via visual interpretation by interpreters from the Government of Indonesia. All burns are spatially and temporally discrete, such that burns of a given estimate that overlap spatially but not temporally are considered separate.

3. Results

3.2. Increased Burned-Area Estimates

Our Indonesia-wide burned-area estimate, based on the classification of the pair of pre- and post-fire Sentinel-2 composites, are larger than the Official estimates as well as the MODIS MCD64A1 to a lesser degree. We estimate 3.11 million hectares (Mha) burned in 2019 across Indonesia, of which 31% were on peat (Figure 5). The extent of peatlands were defined using a national dataset from the Ministry of Agriculture (Ritung et al., 2011). In contrast, Official burned-area estimates, based on visual interpretation of Landsat-8 imagery, report only about half as much burned area, at 1.64 Mha, of which 39% was on peat. Our estimates too are greater than the MODIS MCD64A1 product, which reports 2.04 Mha burned in 2019, or two-thirds of our estimate, with 40% on peat. The greater burning extent and proportionally lesser extent of peatland burning according to our estimates suggest that our estimates are particularly more inclusive of burning across mineral soils.

In the seven provinces for which we assessed accuracy, our Sentinel-2 estimates, and the Official Landsat-8 estimates both report excellent user's accuracies (UA) for the 'burned' class, at 97.9% (CI: 97.1%-98.8%) and 95.1% (CI: 93.5%-96.7%) respectively, indicating a mere 2.9%-4.9% commission-error rate (Table 2, Supplementary Table S5). The producer's accuracies (PA) are comparatively lower for both datasets, but notably less so for our estimates, at 75.6% (CI: 68.3%-83.0%) and 49.5% (CI: 42.5%-56.6%) for our estimate and the Official dataset, respectively. In other words, for any burned area in our reference dataset, there is a 75.6% chance that it will be correctly mapped as burned by our estimate, compared to only a 49.5% for the official estimate. This is in keeping with the greater tendency of the Sentinel-2 estimate to capture more smaller and intermediate-size burns. The MCD64A1 data had a much lower UA for the burned class, at 76.0% (CI: 73.3%-78.7%), as well as a much lower and a PA for the burned class, at 53.1% (CI: 45.8%-60.5%), qualifying it as the least reliable and accurate of the three estimates notwithstanding comparable high overall accuracy (Table 2).

All three burned-area maps underestimate the true burned area extent, as per their respective PA figures, but our Sentinel-based map has the smallest shortfall and also maintains user accuracy. The corrected burned area of the seven provinces is higher than the mapped area for all the three burned area maps. Again, however, our map area most closely approximates its corresponding corrected burned area (Table 2). Whereas our Sentinel-based mapped burned area indicates that 1.84 Mha burned in the seven provinces (or 59% of our total national estimated burned area), the corrected burned area is 2.38 Mha (CI: 2.14 Mha-2.61 Mha) (Table 2), for a discrepancy of 0.54 Mha. In contrast, the official estimate indicates 1.19 Mha burned in the seven provinces (73% of its corresponding total), and a corrected burned area of 2.29 Mha (CI: 1.96 Mha-2.63 Mha), for a 1.1 Mha discrepancy. Likewise, the MCD64A1 dataset mapped 1.58 Mha burned in the seven provinces and has a corrected burned area of 2.27 Mha (CI: 1.94 Mha-2.59 Mha), for a 0.69 Mha discrepancy. Although, we cannot extrapolate a corrected burned area across Indonesia, we are confident that more than 3.11 Mha burned in 2019.

3.1. Burn size comparison.

The Sentinel, Official and MCD64A1 estimates captured significantly distinct realms of fire activity, as represented by relative burn size frequencies (Figure S6). The three estimates differ from one another most notably for small burns, however, they are statistically indistinguishable for burns > 5000 ha indicative of extreme fire activity (Table 3). In other words, all three estimates capture very large burns (>5000 ha) equally well, and distinctions amongst the estimates concentrate amongst small (<100 ha), intermediate (100-1000 ha) and larger burns (1000-5000 ha), in decreasing order of degree as indicated by the magnitude of the test statistics in Table 3.

Inclusivity of smaller and intermediate burned areas is the primary source of difference among estimates. Compared to Official or MCD64A1 estimates, the Sentinel estimate has a significantly greater relative frequency of small burned areas (< 100 ha), especially amongst the smallest of these (Table 4). This is indicative of a greater detection of small fires presumably characterized by small-scale agriculture fires and similar, small-scale controlled burning. The Sentinel estimate similarly has a greater relative frequency of intermediate sized burns (100-1000 ha), but less acutely so, with inter-estimate differences being more moderate for the Official estimate than the MCD64A1 estimate (Table 4, Figure 6, Figure S6). For burns >1000 ha, the Sentinel estimate differs only relative to the official estimate (Table 3), seemingly due to the latter's underestimation of large and very large scars (Figure 6). Note for instance the increasingly large divergence between the cumulative burned-area curves for the Sentinel-2 and the Official estimates in Figure 6 for burn areas > 1000 ha. For very large burns (> 5000 ha), two-way comparisons in Table 4 again report no significant statistical differences in burn-scar detection rates between the Sentinel and alternative estimates. However, given the small sample of patches > 5000 ha, it is noteworthy that the Sentinel estimate captures more very large scars compared to Official estimates (n=56 vs n=16) and avoids critical omissions made by both Official, or MCD64A1, estimates for extremely large burns (>15,000 ha) on peatlands around Berbak National Park in Jambi Province, Sumatra (Figure 7).

In summary, the greater overall burned-area estimate of our Sentinel data compared to the Official and MCD64A1 alternatives reflects differences in the inclusion of smaller and intermediately sized scars. The sum of all Sentinel burn areas that are individually <~860 ha equals the entirety of the official burned-area estimate (Figure 6). The Sentinel-2 data exhibit a size-frequency pattern that approximates a near scale-free power-law (Figure 6).

4. Discussion

We developed a method that generates two national composite Sentinel-2 images depicting vegetation condition before and after burning in 2019 (Figure 1), and then classified this pair to extract burned areas using a Random Forest supervised classification algorithm. We developed a comprehensive validation protocol to strictly assess the reliability and accuracy of our product based on visual interpretation of dense time-series Sentinel-2 original images, and also applied this validation to the widely used global MODIS burned-area product (MCD64A1, collection 6) (Giglio et al., 2018) and to the Official burned-area product of the Indonesian Ministry of Environment and Forestry (MOEF) (Sipongi, 2020).

Our estimate is the most reliable and accurate and therefore captures more of the 2019 total burned area, confirming that 20-m Sentinel-2 imagery is better suited to widespread small-scale burning in Indonesia (Huang et al., 2016), while it also captures large burn scars relatively thoroughly. The study finds similar omission and commission errors (47% and 24%) for the 'burned' class of MCD64A1 product as those presented globally (40% and 22%) (Giglio et al., 2018). The underestimation of total burned area according to the MCD64A1 product compared with our Sentinel-2 estimate is unsurprising, considering that the MODIS 500-m pixel resolution struggles to detect smaller fires (Giglio et al., 2018). Similar conclusions were reached by Ramo et al. (2021) when comparing the new 'Small Fire Dataset' derived using Sentinel-2 over Sub-Sahara Africa (Chuvieco et al., 2018) and the MCD64A1 product. More surprising is the near 2:1 ratio by which the Sentinel-2 estimates surpass the Landsat-8 Official estimate. Our examination shows that this difference reflects differential detection of small- (<100 ha) to intermediate-sized (<1000 ha) burn scars.

The Sentinel-2 data exhibit a size-frequency pattern that approximates closer to a near scale-free power-law, or pareto distribution (Karsai et al., 2020;Falk et al., 2007). These patterns are typical of large-scale fire studies (Malamud et al., 1998). Both other methods yield an S-shaped curve with less area at smaller and larger sizes than captured in the Sentinel-2, indicating likely bias by omission over the entire range of scales and are not determined by image resolution alone (Figure 6). These results, with different frequency patterns arising from burns from the same regions in the same period, also highlight the danger in interpreting apparent burned-area patterns without careful consideration of the limitations and biases that arise from the methods used to map them—an issue that may not have always been sufficiently recognized in past assessments or policy.

Although both Sentinel-2 and Landsat-8 both observe the infrared wavelengths required to detect charred vegetation and have similar spatial resolutions (20 m x 20 m and 30 m x 30 m, respectively), Sentinel-2 detects more burns of the greater frequency of its coverage (five- versus sixteen-day revisit time). Also, our method avails of the massive computational capabilities and automation of the Google Earth Engine, allowing us to analyze more images and thus map more and smaller burn scars and associated details than could even the most well-equipped team of visual interpreters.

Despite high reliability that every burn scar detected on the map was valid (2.9% commission error rate), our method suffered a 24.4% omission error rate (burned areas that remained undetected). These rates reflect necessary tradeoffs between commission and omission error in a context where conservative estimates are much preferred for environmental policy and monitoring. We prioritized a low commission error rate (i.e. high user's accuracy) over absolute burned-area coverage to address sensitivities (Rochmyaningsih, 2020). By hedging against commission errors, our approach omitted hard-to-detect events, including low-intensity burns, such as those that occur beneath the forest canopy on mineral soils (van Nieuwstadt and Sheil, 2005) or on savanna grasslands, which tend to re-green rapidly. While further work is required to clarify and refine the optimal levels of inclusivity and reliability, we emphasize that the production of before and after fire annual composite images is relatively straightforward for the user community, given the availability of both the necessary imagery and our Google Earth Scripts.

While the accuracy assessment proved that our training dataset is valid for the classification of Sentinel-2 composites for the year 2019 in Indonesia, this training dataset might not achieve equivalent accuracy for other years and regions. The pre- and post-fire composites might show different spectral changes under different conditions. For instance, high rainfall in 2020 influenced reflectance. Similarly, representative training points should be used in other regions. Those adapting these methods should ensure adequate local training data and validation.

Doubts may persist concerning confident estimates of burn areas without extensive and costly ground-checks. Modern high-resolution remote sensing makes such on-the-ground checks less essential than in the past as burned areas are readily identified with good accuracy in modern high-resolution imagery such as that we used for our validation. The protocol developed here to generate a reference dataset based on visual inspection of dense (5-day revisit time) satellite imagery is better suited than ground verifications of 'burned' and 'unburned' locations, because it allows the generation of extensive randomly distributed well characterised reference sites, a process too time-consuming and costly with field visits. The identification and quantification of less-readily-detected burned areas, such as those under a closed forest canopy, remain a challenge but will require dedicated and targeted research and would not be solved by ground-checks alone.

Accurate estimates of burned lands, in particular on peat, are central to addressing concerns about regional air quality, and to ambitious national climate-change atmospheric carbon reduction commitments heavily reliant on improved land/fire management (DGCC, 2019). Though we observed proportionally less peatland burning than the alternative burned-area estimates (31% versus 39% and 40% for the Official and MCD64A1 products, respectively), due to our more complete coverage, we observed more peatland burning absolutely (0.96 Mha) than the official estimate (0.64 Mha). Given this large discrepancy for peatland burning, we anticipate that our refined burned area product will enable others to better estimate carbon emissions from the 2019 fires in Indonesia. Combined with daily fire hotspots detected using thermal remote sensing, our detailed burned-area map can help identify ignition sites and estimate fire duration more precisely, and therefore contribute to forensic analyses of burning across landholdings (Gaveau et al., 2017) as well as assess policies and practices intended to reduce or control ignition events and the scale of fires (Watts et al., 2019).

The Indonesian government has shown some success in reducing fires (Sloan et al., 2021). Apparent reductions to fire activity would however ideally be qualified using our more inclusive and accurate burned-area estimates. Further, the Indonesian government must also develop improved protocols to quantify the resulting carbon emissions (DGCC, 2019). Our protocols for creating reliable pre- and post-fire composites are replicable. To further the adoption and reproduction of our approach, we have published all our protocols, scripts, applications, burned-area map, reference data, pre-fire and post-fire Sentinel-2 composite images, and various other outputs so that anyone may employ and revise them as they wish (see Data Availability).

5. Code availability

The code that generates the Sentinel-2 pre- and post-fire composites can be found at: https://github.com/thetreemap/IDN_annual_burned_area_detection

6. Data Availability

All the data including pre- and post-fire composites, all three burned area products, and reference points with screenshots can be visualized online at this application portal: https://thetreemap.users.earthengine.app/view/burn-area-validation-simplified

The Sentinel-based burned area map and reference dataset are freely available for download at: https://doi.org/10.5281/zenodo.4551243 (Gaveau et al., 2021).

The dataset *2019_burnedarea_indonesia.shp* contains the 2019 burned-area estimates that we developed for Indonesia using 20 m x 20 m time-series Sentinel-2 imagery. The reference dataset *Reference_dataset.shp* contains 1298 reference points that we assembled and used to validate all three burned area products described in this study. Each reference point includes attribute 'REFERENCE' to describe the values obtained by visual interpretation: either 'NO' unburned or 'YES' burned. Each reference point has three attributes: 'C_SENTINEL' 'C_OFFICIAL' and 'C_MCD64A1' to describe the values of the classification of each burned area product: either 'NO' unburned or 'YES' burned. Finally, each reference point has three additional attributes: 'SENTINEL', 'OFFICIAL', and MCD64A1' to describe which burned area product this reference point validates. The values are either 0: not validate or 1: validate.

The MODIS MCD64A1 dataset was obtained at: https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MCD64A1. The official burned area dataset from the Ministry of Environment and Forestry (MOEF) was obtained at: https://geoportal.menlhk.go.id/webgis/index.php/en/

The Sentinel-2 Level 2A used in this study are available at https://scihub.copernicus.eu/ and can be retrieved in Google Earth Engine. The Sentinel- 2 data are hosted and accessed in the Earth Engine data catalog (the links to the data are https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR). Data ingested and hosted in Google Earth Engine are always maintained in their original projection, resolution, and bit depth (Gorelick et al., 2017).

Financial support. Funding by the CGIAR Research Program on Forests, Trees and Agroforestry (CRP-FTA), with financial support from the donors to the CGIAR Fund, is recognized.

Author Contributions. D.L.A.G. designed the study. DL.A.G, M.A.S. and A.D designed the burn scar detection method. M.A.S. and A.D wrote the code in Google Earth Engine. D.L.A.G, M.A.S. and A.D. carried out the validation. S.S. carried out the burn scar size analysis. D.L.A.G., A.D. S.S. and D.S. interpreted the results and wrote the manuscript and produced the figures.

Competing interests. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper.

References

Adrianto, H. A., Spracklen, D. V., Arnold, S. R., Sitanggang, I. S., and Syaufina, L.: Forest and Land Fires Are Mainly Associated with Deforestation in Riau Province, Indonesia, Remote Sensing, 12, 3, 2020.

Alonso-Canas, I., and Chuvieco, E.: Global burned area mapping from ENVISAT-MERIS and MODIS active fire data, Remote Sensing of Environment, 163, 140-152, 2015.

Breiman, L.: Random forests, Machine learning, 45, 5-32, 2001.

Cai, W., Yang, K., Wu, L., Huang, G., Santoso, A., Ng, B., Wang, G., and Yamagata, T.: Opposite response of strong and moderate positive Indian Ocean Dipole to global warming, Nature Climate Change, 11, 27-32, 10.1038/s41558-020-00943-1, 2021.

Carmenta, R., Zabala, A., Trihadmojo, B., Gaveau, D., Salim, M. A., and Phelps, J.: Evaluating bundles of interventions to prevent peat-fires in Indonesia, Global Environmental Change, 102154, 2020.

Cochrane, M. A.: Fire science for rainforests, Nature, 421, 913-919, 2003.

Chuvieco, E., Mouillot, F., van der Werf, G. R., San Miguel, J., Tanase, M., Koutsias, N., García, M., Yebra, M., Padilla, M., and Gitas, I.: Historical background and current developments for mapping burned area from satellite Earth observation, Remote Sensing of Environment, 225, 45-64, 2019.

Chuvieco, E.; Pettinari, M.L.; Bastarrika, A.; Roteta, E.; Storm, T.; Padilla Parellada, M.: ESA Fire Climate Change Initiative (Fire_cci): Small Fire Dataset (SFD) Burned Area pixel product for Sub-Saharan Africa, version 1.1. Centre for Environmental Data Analysis, *12 October 2018, 2018*. doi:10.5285/065f6040ef08485db989cbd89d536167.

Crippa, P., Castruccio, S., Archer-Nicholls, S., Lebron, G., Kuwata, M., Thota, A., Sumin, S., Butt, E., Wiedinmyer, C., and Spracklen, D.: Population exposure to hazardous air quality due to the 2015 fires in Equatorial Asia, Scientific reports, 6, 1-9, 2016.

Dennis, R. A., Mayer, J., Applegate, G., Chokkalingam, U., Colfer, C. J. P., Kurniawan, I., Lachowski, H., Maus, P., Permana, R. P., and Ruchiat, Y.: Fire, people and pixels: linking social science and remote sensing to understand underlying causes and impacts of fires in Indonesia, Human Ecology, 33, 465-504, 2005.

DGCC: Emission Reduction Report, Directorate General of Climate Change, 2019.

Falk, D. A., Miller, C., McKenzie, D., and Black, A. E.: Cross-scale analysis of fire regimes, Ecosystems, 10, 809-823, 2007.

Fanin, T., and Werf, G. R.: Precipitation–fire linkages in Indonesia (1997–2015), Biogeosciences, 14, 3995-4008, 2017.

Field, R. D., van der Werf, G. R., and Shen, S. S.: Human amplification of drought-induced biomass burning in Indonesia since 1960, Nature Geoscience, 2, 185-188, 2009.

Field, R. D., Van Der Werf, G. R., Fanin, T., Fetzer, E. J., Fuller, R., Jethva, H., Levy, R., Livesey, N. J., Luo, M., and Torres, O.: Indonesian fire activity and smoke pollution in 2015 show persistent nonlinear sensitivity to El Niño-induced drought, Proceedings of the National Academy of Sciences, 113, 9204-9209, 2016.

Fletcher, K.: SENTINEL 2: ESA's Optical High-Resolution Mission for GMES Operational Services, European Space Agency, 2012.

Gaveau, D.L.A, Salim, M, Hergoualc'h, K, Locatelli, B, Sloan, S, Wooster, M, Marlier, M, Molidena, E, Yaem, H, Defries, R, Verchot, L, Murdiyarso, D, Nasi, R, Holmgren, P & Sheil, D.: Major atmospheric emissions from peat fires in Southeast Asia during non-drought years: evidence from the 2013 Sumatran fires. *Scientific Reports* 4:6112, 2014.

Gaveau, D. L.A, Pirard, R., Salim, M. A., Tonoto, P., Parks, S. A., and Carmenta, R.: Overlapping land claims limit the use of satellites to monitor No-Deforestation commitments and No-Burning compliance, Conservation Letters, 2017.

Gaveau, D. L.A, Descal, A, Salim, M.A, Sheil, D, & Sloan, S. 2019 burned area map for Indonesia using Sentinel-2 data [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4551243, 2021.

Gaveau, D. L.A, Santos, L., Locatelli, B., Salim, M. A., Husnayaen, H., Meijaard, E., Heatubun, C., and Sheil, D.: Forest loss in Indonesian New Guinea (2001–2019): Trends, drivers and outlook, Biological Conservation, 261, 109225, 2021.

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., and Justice, C. O.: The Collection 6 MODIS burned area mapping algorithm and product, Remote sensing of environment, 217, 72-85, 2018.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sensing of Environment, 202, 18-27, 2017.

Harrison, M. E., Ripoll Capilla, B., Thornton, S. A., Cattau, M. E., and Page, S. E.: Impacts of the 2015 fire season on peat-swamp forest biodiversity in Indonesian Borneo, Peatlands in harmony–Agriculture, industry & nature. Proceedings of the 15th international peat congress: Oral presentations, 2016, 713-717,

Hawbaker, T.J., Vanderhoof, M.K., Beal, Y.-J., Takacs, J.D., Schmidt, G.L., Falgout, J.T., Williams, B., Fairaux, N.M., Caldwell, M.K., Picotte, J.J., Howard, S.M., Stitt, S., Dwyer, J.L., 2017b. Mapping burned areas using dense time-series of Landsat data. Remote Sensing of Environment 198, 504–522.

Huang, H., Roy, D. P., Boschetti, L., Zhang, H. K., Yan, L., Kumar, S. S., Gomez-Dans, J., and Li, J.: Separability analysis of Sentinel-2A Multi-Spectral Instrument (MSI) data for burned area discrimination, Remote Sensing, 8, 873, 2016.

Huijnen, V., Wooster, M., Kaiser, J., Gaveau, D., Flemming, J., Parrington, M., Inness, A., Murdiyarso, D., Main, B., and Van Weele, M.: Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997, Scientific reports, 6, 26886, 2016.

Karsai, I., Schmickl, T., and Kampis, G.: Forest Fires: Fire Management and the Power Law, in: Resilience and Stability of Ecological and Social Systems, Springer, 63-77, 2020.

Koplitz, S. N., Mickley, L. J., Marlier, M. E., Buonocore, J. J., Kim, P. S., Liu, T., Sulprizio, M. P., DeFries, R. S., Jacob, D. J., and Schwartz, J.: Public health impacts of the severe haze in Equatorial Asia in September–October 2015: demonstration of a new framework for informing fire management strategies to reduce downwind smoke exposure, Environmental Research Letters, 11, 094023, 2016.

Liu, S., Zheng, Y., Dalponte, M., and Tong, X.: A novel fire index-based burned area change detection approach using Landsat-8 OLI data, European journal of remote sensing, 53, 104-112, 2020.

Lizundia-Loiola, J., Otón, G., Ramo, R., and Chuvieco, E.: A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data, Remote Sensing of Environment, 236, 111493, 2020.

Lizundia-Loiola, J., Franquesa, M., Boettcher, M., Kirches, G., Pettinari, M. L., and Chuvieco, E.: Operational implementation of the burned area component of the Copernicus Climate Change Service: from MODIS 250 m to OLCI 300 m data, Earth System Science Data Discussions, 1-37, 2021.

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

Malamud, B. D., Morein, G., and Turcotte, D. L.: Forest fires: an example of self-organized critical behavior, Science, 281, 1840-1842, 1998.

Marlier, M. E., DeFries, R. S., Voulgarakis, A., Kinney, P. L., Randerson, J. T., Shindell, D. T., Chen, Y., and Faluvegi, G.: El Niño and health risks from landscape fire emissions in southeast Asia, Nature climate change, 3, 131-136, 2013.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A.: Good practices for estimating area and assessing accuracy of land change, Remote Sensing of Environment, 148, 42-57, http://dx.doi.org/10.1016/j.rse.2014.02.015, 2014.

Otón, G., Ramo, R., Lizundia-Loiola, J., and Chuvieco, E.: Global detection of long-term (1982–2017) burned area with AVHRR-LTDR data, Remote Sensing, 11, 2079, 2019.

Ramo, R., Roteta, E., Bistinas, I., Van Wees, D., Bastarrika, A., Chuvieco, E., and Van der Werf, G. R.: African burned area and fire carbon emissions are strongly impacted by small fires undetected by coarse resolution satellite data, Proceedings of the National Academy of Sciences, 118, 2021.

Ritung, S., Wahyunto, Nugroho, K., Sukarman, Hikmatullah, Suparto, and C, T.: Peatland map of Indonesia, Department of Research and Development of Agricultural Land Resources, Ministry of Agriculture, 2011.

Rochmyaningsih, D.: Wildfire researcher deported amid growing rift between Indonesian government and scientists, Science, 367, 722-723, 2020.

Sipongi.: Recapitulation of Land and Forest Fires Area (Ha) per Province in Indonesia 2015-2020: http://sipongi.menlhk.go.id/hotspot/luas_kebakaran, 2020.

Sloan, S., Locatelli, B., Wooster, M. J., and Gaveau, D. L.: Fire activity in Borneo driven by industrial land conversion and drought during El Niño periods, 1982–2010, Global environmental change, 47, 95-109, 2017.

Sloan, S., Tacconi, L., and Cattau, M.: Fire prevention in managed landscapes: Recent success and challenges in Indonesia, Mitigation and Adaptation Strategies for Global Change, 26: Article 32, 2021.

Stehman, S. V., Wickham, J., Smith, J., and Yang, L.: Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: Statistical methodology and regional results, Remote Sensing of Environment, 86, 500-516, 2003.

Tansey, K., Beston, J., Hoscilo, A., Page, S., and Paredes Hernández, C.: Relationship between MODIS fire hot spot count and burned area in a degraded tropical peat swamp forest in Central Kalimantan, Indonesia, Journal of Geophysical Research: Atmospheres, 113, 2008.

Van der Werf, G. R., Dempewolf, J., Trigg, S. N., Randerson, J. T., Kasibhatla, P. S., Giglio, L., Murdiyarso, D., Peters, W., Morton, D., and Collatz, G.: Climate regulation of fire emissions and deforestation in equatorial Asia, Proceedings of the National Academy of Sciences, 105, 20350-20355, 2008.

Van Der Werf, G. R., Randerson, J. T., Giglio, L., Van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., Van Marle, M. J., Morton, D. C., and Collatz, G. J.: Global fire emissions estimates during 1997–2016, Earth System Science Data, 9, 697-720, 2017.

van Nieuwstadt, M. G. L., and Sheil, D.: Drought, fire and tree survival in a Borneo rain forest, East Kalimantan, Indonesia, Journal of Ecology, 93, 191-201, 2005.

Watts, J. D., Tacconi, L., Hapsari, N., Irawan, S., Sloan, S., and Widiastomo, T.: Incentivizing compliance: Evaluating the effectiveness of targeted village incentives for reducing burning in Indonesia, Forest Policy and Economics, 108, 101956, 2019.

Wooster, M., Gaveau, D., Salim, M., Zhang, T., Xu, W., Green, D., Huijnen, V., Murdiyarso, D., Gunawan, D., and Borchard, N.: New tropical peatland gas and particulate emissions factors indicate 2015 Indonesian fires released far more particulate matter (but less methane) than current inventories imply, Remote Sensing, 10, 495, 2018.

Figures

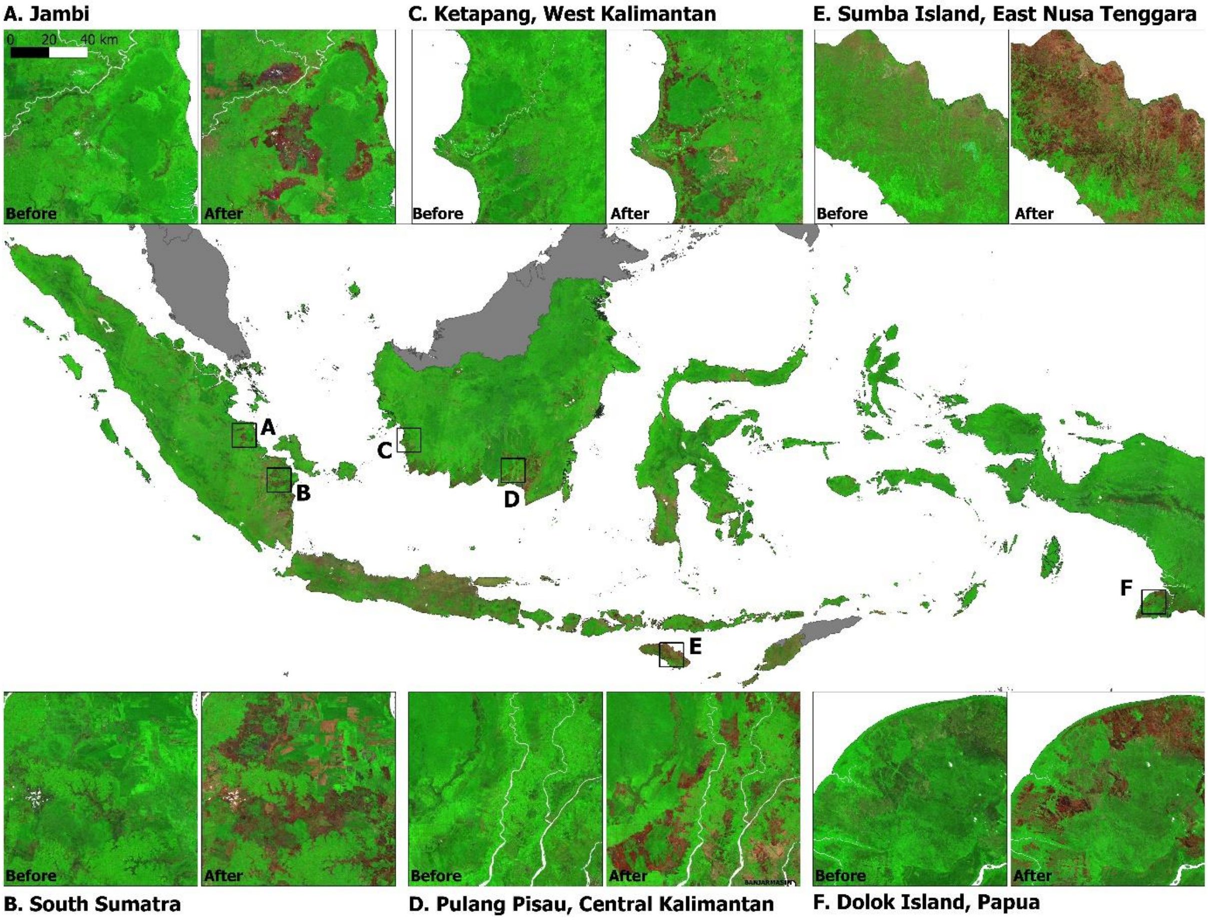

Figure 1. The pair of cloud-free pre-and post-fire Sentinel-2 composites shown over six locations in insets A, B, C, D, E, F (all insets have the same scale). The base Indonesia-wide imagery is the post-fire composite. Imagery displayed in false colors (RGB: short-wave infrared (band 11); Near infrared (band 8), Blue: red (band 4)). In this pair of composite images acquired shortly before and after fire a recently burned area will readily appear to have transitioned from ‘green’ to dark ‘brown/red’ tones. Areas cleared without burning appear bright pink. Areas covered with vegetation appear dark to bright green.

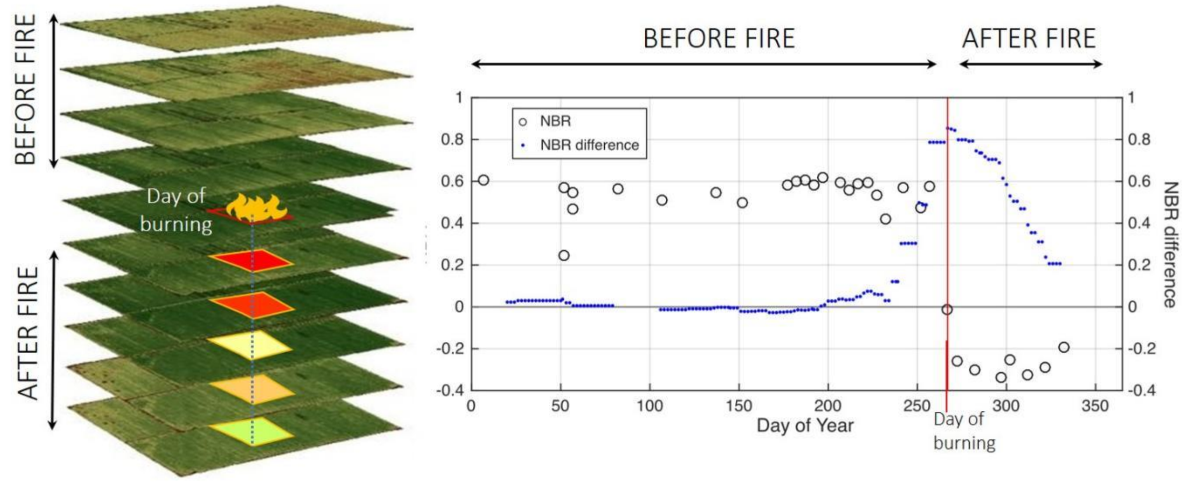

Figure 2. A schematic of Sentinel-2 time-series imagery, associated NBR values (open circles) and NBR differences between average NBR values observed before and after the central day of a 2-day moving window (blue dots). A burned pixel (20 m x 20 m) is represented by a red rectangle at left. Before fire, the vegetated pixel registers positive NBR values (open circles). The NBR rapidly drops during the fire and, for a few weeks, the satellite observations show a negative NBR. The day of the year when the NBR difference observed via the moving window reaches a maximum corresponds to the moment NBR dropped (red line). This day marks a decline in the pixel's vegetation, possibly reflecting a burning event. Over time, the vegetation regenerates (re-greening) and the spectral characteristic of charred vegetation fades. Regreening can happen within days in the case of savanna grasslands, or within months in the case of forest fires on peatlands.

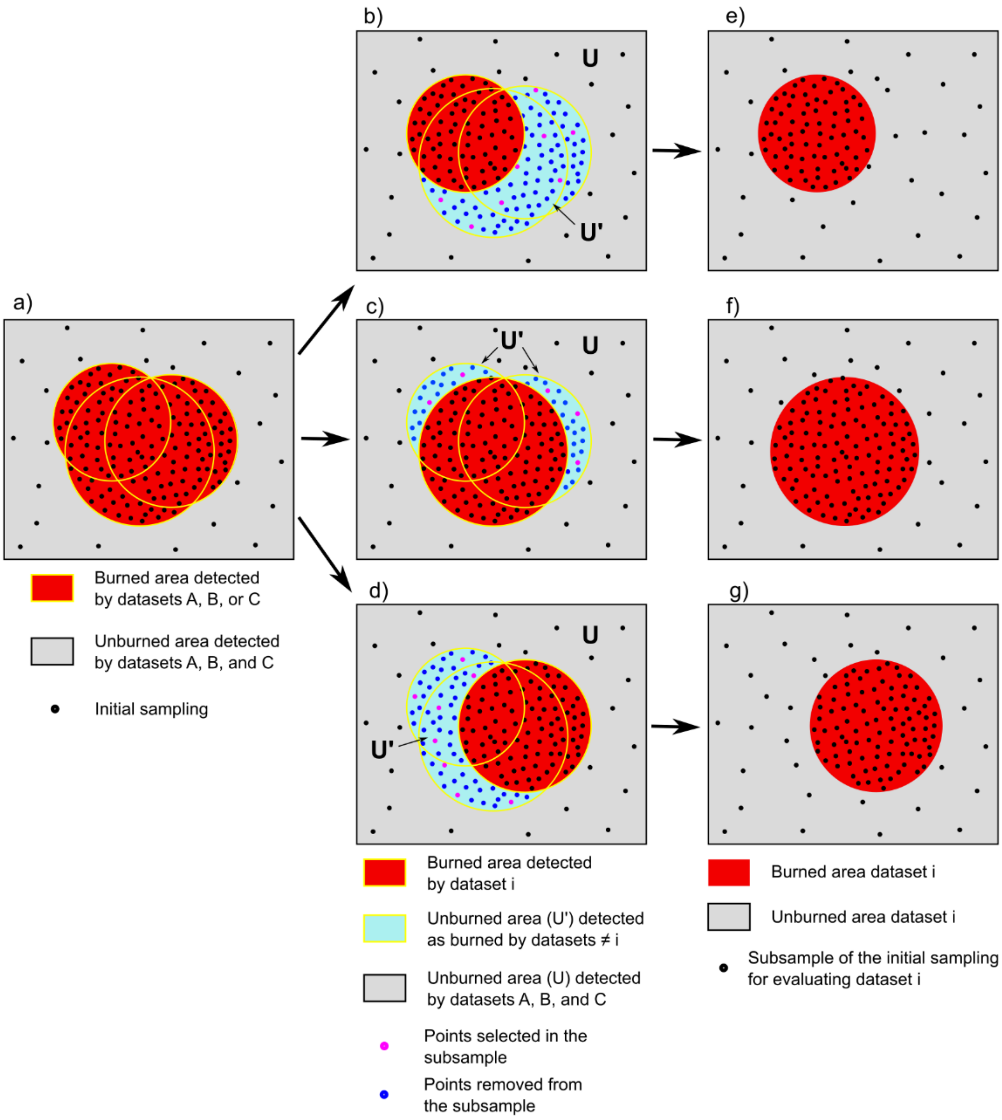

Figure 3. Representation of the adjusted, stratified-sampling design for the validation of three burned area datasets (A, B, and C) against reference sites (dots). Panel (a) shows the stratified random sampling of reference sites (black points) over the combined burned area. Note that the density of samples is higher in the combined burned area than the unburned area. Panels (b), (c), and (d) show, in cyan, the area U', being classified as unburned in a given dataset *i* but classified as burned in at least one other datasets ≠ *i*. For a given validation of A, B, and C, the sample points in the corresponding area U' (panels (b), (c), (d)) were randomly excluded until the sampling density in the area U' equaled that of the larger unburned area U (area in gray). Panels (e), (f) and (g) show the three final, adjusted, stratified subsamples of reference points derived from the initial sample of 1298 reference points. Note that the relative areas and number of sites per class in Figure 3 do not correspond to the actual datasets being evaluated.

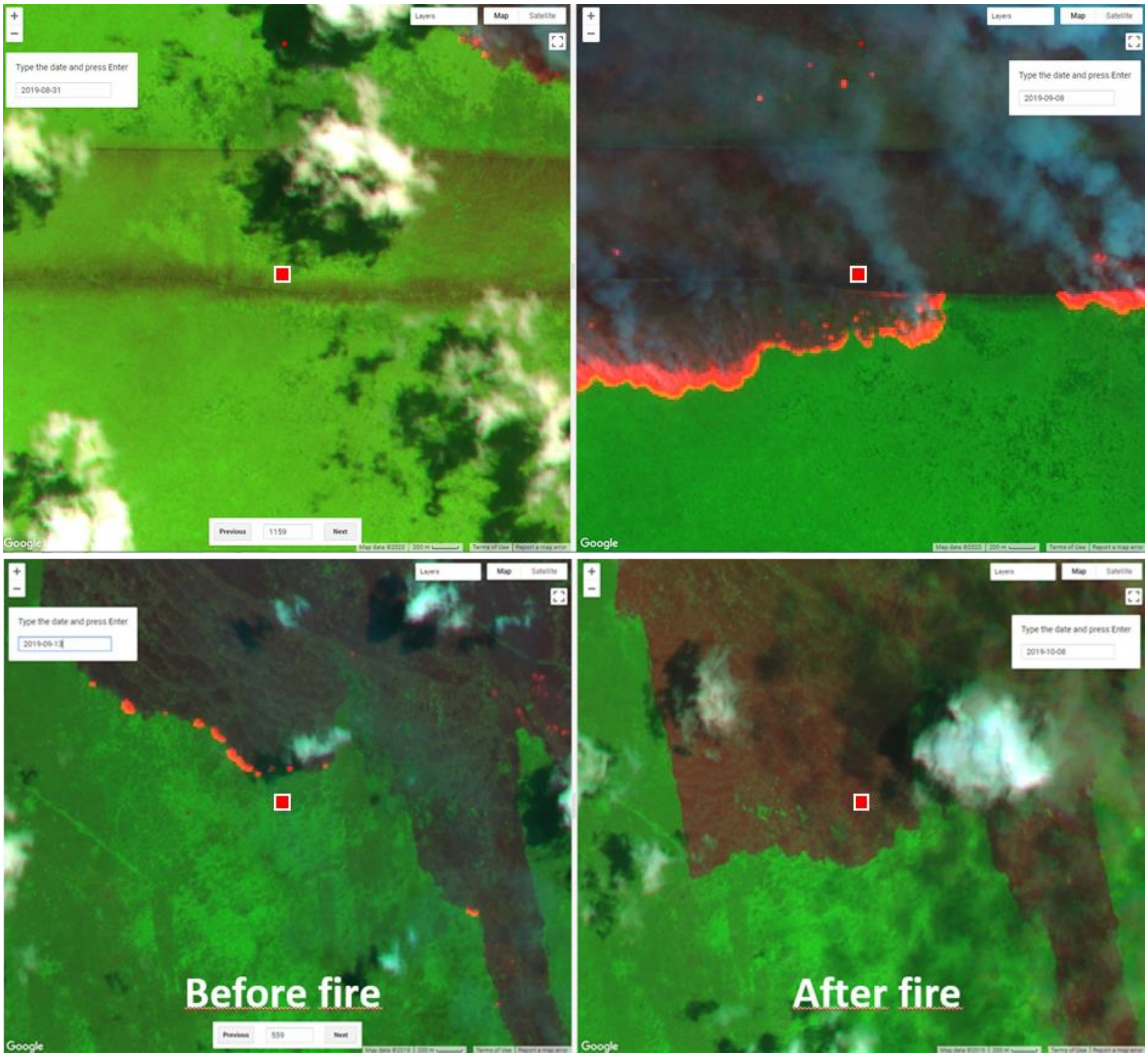

Figure 4. Two snapshots recording the pre-fire (left panel) and post-fire (right panel) original Sentinel-2 images acquired shortly before (13 September 2019) and shortly after (08 October 2019) fire for two reference site (red squares). Imagery displayed in RGB: SWIR, NIR, RED. Sentinel-2 provides two SWIR Bands. Band 12=2.190 µm is more suitable than Band 11=1.610 µm to detect the intense heat from flaming fronts. On these image pairs, one can see flaming fronts traveling towards the reference sites (red dot) from the north on the pre fire images (left), and sharp changes in color from ‘green’ to ‘dark red’ characteristic of charred remains with continuing flaming on the post-fire images (right). Layout built using © Google Earth Engine.

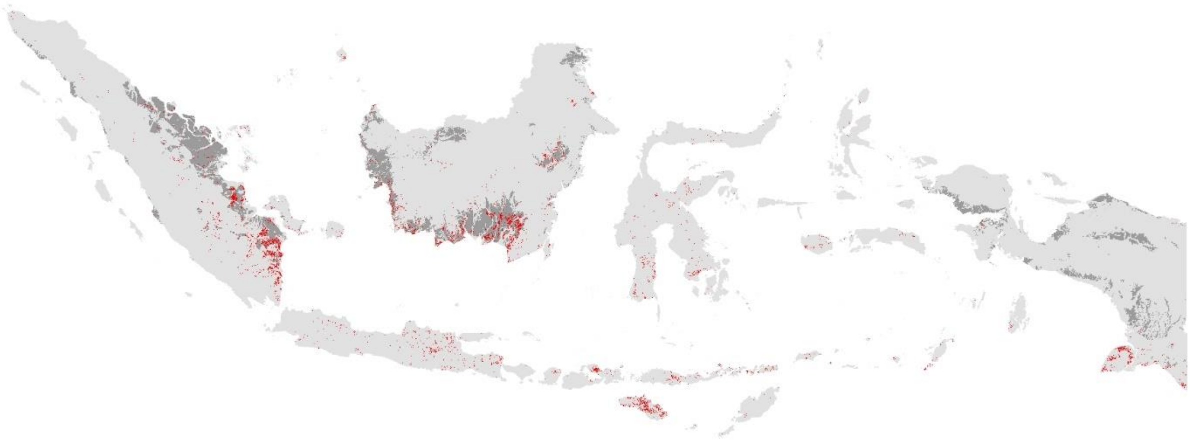

Figure 5. 2019 burned areas (red) for Indonesia (grey area) derived using a time-series of the atmospherically corrected surface reflectance multispectral images (level 2A product) taken by the Sentinel-2 A and B satellites. The spatial resolution of this map is 20 m x 20 m, and minimum mapping unit 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 S5.

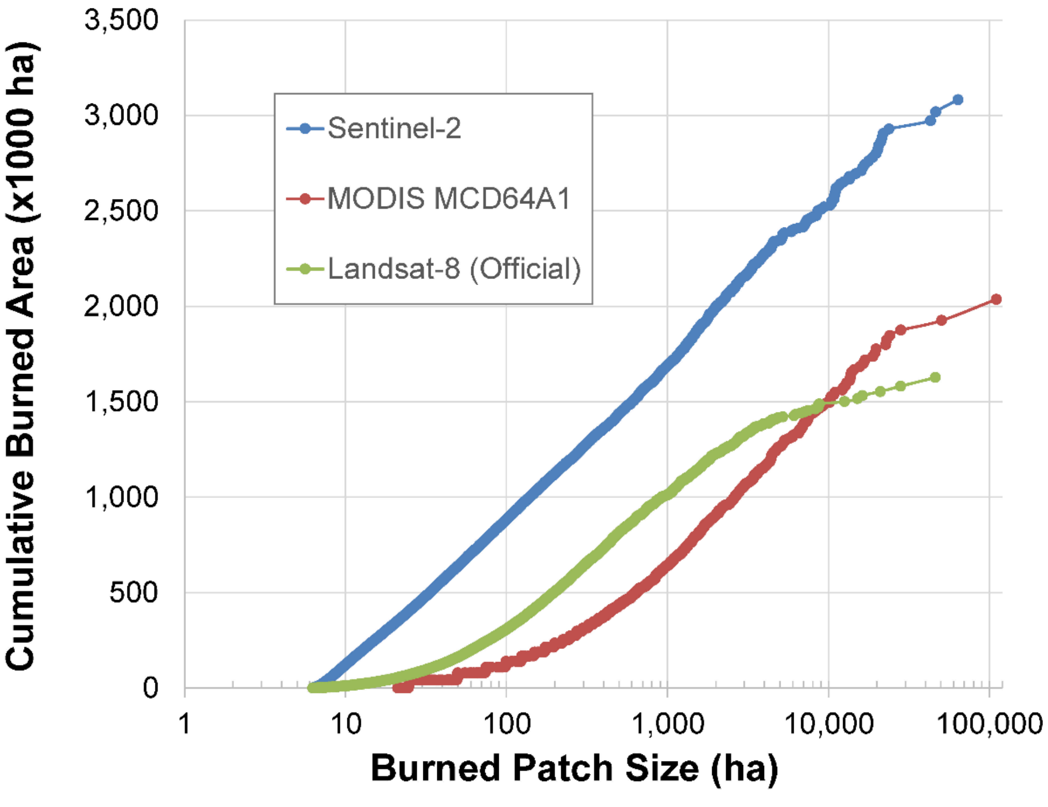

Figure 6. Cumulative national total burned area versus burned-scar area, for Sentinel-2, Landsat-8 (Official), and MODIS MCD64A1 burned-area estimates. Note the logarithmic axis. For a given segment of the x-axis between scar sizes X_1 and X_2, a difference in the slopes for any two estimates is indicative of inter-estimate differences in terms of inclusivity of scars between X_1 and X_2.

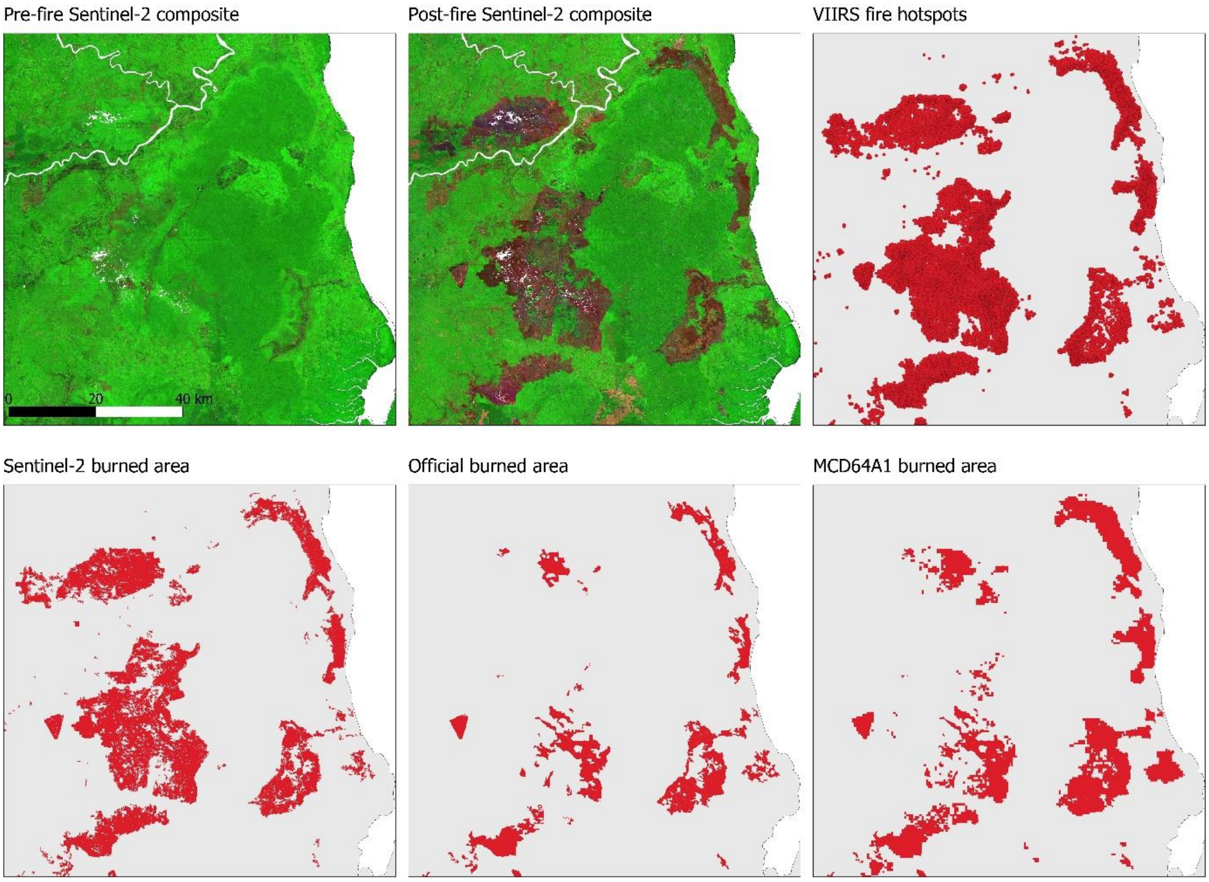

Figure 7. The pair of cloud-free pre-and post-fire Sentinel-2 composites over Berback National Park (black line) and surrounding areas in Jambi Province (see also Inset A, Figure 1), revealing large, burned areas around Berbak National Park (areas that have transitioned from 'green' to dark 'brown/red' 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.

Tables

Table 1. Adjusted, Stratified Subsamples of Reference Sites to Validate Burned-Area Estimates.

Burned-Area Estimate	**Reference Sites**		**Total Reference Sites**
	In Areas Classified as Burned	*In Areas Classified as Unburned (U & U')*	
Sentinel-2 (this study)	888	280	1168
MODIS MCD64A1	891	242	1133
Landsat-8 (Official)	895	182	1077

Table 2. Accuracy assessment of each of the three burned area maps performed in seven Indonesian provinces (87.60 Mha) targeted for peatland restoration. The accuracy metrics were estimated with an initial total of 1,298 points randomly distributed using stratified sampling. The reported metrics are: 1) the overall accuracy (OA), the user's accuracy (UA), and the producer's accuracy (PA) with their 95% confidence intervals, and 2) the mapped burned area and the corrected burned area with their 95% confidence intervals.

		SENTINEL	*OFFICIAL*	*MCD64A1*
OA (%)		99.3 (99.1, 99.6)	98.1 (97.8, 98.5)	98.4 (98.1, 98.8)
	Burned	97.9 (97.1, 98.8)	95.1 (93.5, 96.7)	76.0 (73.3, 78.7)
UA (%)	*Unburned*	99.3 (99.1, 99.6)	98.6 (98.2, 99.0)	98.8 (98.5, 99.2)
	Burned	75.6 (68.3, 83.0)	49.5 (42.5, 56.6)	53.1 (45.8, 60.5)
PA (%)	*Unburned*	99.9 (99.9, 99.9)	99.9 (99.9, 99.9)	99.6 (99.6, 99.7)
Mapped burned area (Mha)		1.84	1.19	1.58
Corrected burned area (Mha)		2.38 (2.14 , 2.61)	2.29 (1.96 , 2.63)	2.27 (1.94 , 2.59)
Difference (Mha)		0.54	1.1	0.69

Table 3. Tests statistics with respect to three-way differences in burned area scar-size frequency distributions for Sentinel, MODIS, and official estimates.

Scar Size (ha)	**Kruskal-Wallis H[a]**
> 25	998*
> 100	335*
> 1000	14*
> 5000[a]	0.61

Significance: * p<0.001

Notes: Scar-size thresholds in the table denote the set of scars included in a test. Tests pertain to whether frequency distributions have equivalent 'distribution location', that is, position along a continuum of scar sizes. Tests thus pertain to whether the estimates capture distinct realms of fire activity, assuming similarly shaped frequency distributions. Higher test statistic values indicate greater probability that the estimates differ with respect to distribution location. The tree-way comparisons of the estimates may flag differences where all three estimates differ or where only two of the three differ. Significance is not Bonferroni corrrected. (a) There are 56, 60 and 16 scars > 5000 ha for Sentinel, MCD64A1, Official estimates, respectively.

Table 4. Test statistics with respect to two-way differences in burned area scar-size frequency distributions, with respect to distribution shape and situation (Test I) or situation alone (Test II), for Sentinel estimates compared to either MCD64A1 or Official estimates.

Scar Size (ha)	**Sentinel vs. MCD64A1**		**Sentinel vs. Official**	
	I. Kolmogorov-Smirnov Z-score (Most Extreme Difference [positive/negative])[b]	*II. Mann-Whitney U Z-score*	*I. Kolmogorov-Smirnov Z-score (Most Extreme Difference [positive/negative])*[b]	*II. Mann-Whitney U Z-score*
> 6.25	N/A		31.8** (+0.32)	-70.6**
> 25	14.7** (+0.24/0.-15)	-20.1*	13.2** (+0.18)	-28.6*
> 100	7.9** (+0.23)	-16.6*	1.6† (+0.04/-0.04)	-0.57
> 1000	0.76 (+0.06/-0.03)	-0.79	1.5‡ (+0.01/-0.12)	-3.1•
> 5000[a]	0.72 (+0.14/-0.08)	-0.77	0.70 (+0.13/-0.20)	0.10

Significance: ** p<0.0001; * p<0.001; • p<0.01; † p=0.014; ‡ p<0.05

Notes: Scar-size thresholds denote the cohort of scars included in a test. Test I and Test II both pertain to whether the Sentinel estimates capture distinct realms (scar-size cohorts) of fire activity compared to the other two estimates. Test I pertains to whether the scar-size frequency distribution of the Sentinel estimate has the same shape and 'distribution location' as either the MODIS or official estimate. Test II is the same but with respect to distribution location only. Distribution location refers to the situation of a frequency distribution along a continuum of scar sizes. Higher test statistics indicate greater probability that the estimates differ significantly with respect to distribution shape and/or location. Reported statistical significance is without Bonferroni corrections. a) There are 56, 60 and 16 scars > 5000 ha for Sentinel, MODIS, official estimates, respectively. (b) Largest positive and negative differences in the cumulative probability functions of Sentinel vs. MODIS or official scar-size estimates. No difference was reported where it was <0.00 absolutely.