ESA Climate Change Initiative – Fire_cci
D4.1 Product Validation and Intercomparison Report (PVIR)

Project Name	ECV Fire Disturbance: Fire_cci
Contract Nº	4000126706/19/I-NB
Issue Date	07/06/2022
Version	2.1
Author	Daniela Stroppiana, Matteo Sali, Mirco Boschetti, Lorenzo Busetto, Luigi Ranghetti, Magín Franquesa, Joshua Lizundia-Loiola, M. Lucrecia Pettinari
Document Ref.	Fire_cci_D4.1_PVIR_v2.1
Document type	Public

To be cited as: Stroppiana D., Sali M., Boschetti M., Busetto L., Ranghetti L., Franquesa M., Lizundia-Loiola J., Pettinari M.L. (2022) ESA CCI ECV Fire Disturbance: D4.1 Product Validation and Intercomparison Report, version 2.1. Available at: https://climate.esa.int/en/projects/fire/key-documents/

Project Partners

Prime Contractor/ Scientific Lead & Project Management	UAH – University of Alcala (Spain)
Earth Observation Team	UAH – University of Alcala (Spain) UPM – Universidad Politécnica de Madrid (Spain) CNR-IREA - National Research Council of Italy – Institute for Electromagnetic Sensing of the Environment (Italy)
System Engineering	BC – Brockmann Consult (Germany)
Climate Modelling Group	MPIM – Max Planck Institute for Meteorology (Germany) CNRS - National Centre for Scientific (France)

Distribution

Affiliation	Name	Address	Copies
ESA	Clément Albergel (ESA)	clement.albergel@esa.int	electronic copy
Project Team	Emilio Chuvieco (UAH)	emilio.chuvieco@uah.es	electronic copy
	M. Lucrecia Pettinari (UAH)	mlucrecia.pettinari@uah.es	
	Joshua Lizundia-Loiola (UAH)	joshua.lizundia@uah.es	
	Gonzalo Otón (UAH)	gonzalo.oton@uah.es	
	Amin Khaïroun (UAH)	amin.khairoun@uah.es	
	Mihai Tanase (UAH)	mihai.tanase@uah.es	
	Consuelo Gonzalo (UPM)	consuelo.gonzalo@upm.es	
	Dionisio Rodríguez Esparragón (UPM)	dionisio.rodriguez@ulpgc.es	
	Daniela Stroppiana (IREA)	stroppiana.d@irea.cnr.it	
	Mirco Boschetti (IREA)	boschetti.m@irea.cnr.it	
	Thomas Storm (BC)	thomas.storm@brockmann-consult.de	
	Angelika Heil (MPIM)	angelika.heil@mpimet.mpg.de	
	Idir Bouarar (MPIM)	idir.bouarar@mpimet.mpg.de	
	Florent Mouillot (LSCE)	florent.mouillot@cefe.cnrs.fr	
	Philippe Ciais (LSCE)	philippe.ciais@lsce.ipsl.fr	

Summary

This Product Validation and Intercomparison Report (PVIR) describes the implementation of the validation methods and results derived for assessing the accuracy of global, regional and local BA products, and presents an intercomparison between the currently existing global burned area products.

	Affiliation/Function	Name	Date
Prepared	CNR UAH	Daniela Stroppiana Matteo Sali Mirco Boschetti Lorenzo Busetto Luigi Ranghetti Magín Franquesa Joshua Lizundia-Loiola M. Lucrecia Pettinari	06/06/2022
Reviewed	UAH	M. Lucrecia Pettinari Emilio Chuvieco	06/06/2022
Authorized	UAH	Emilio Chuvieco	07/06/2022
Accepted	ESA	Clément Albergel	07/06/2022

This document is not signed. It is provided as an electronic copy.

Document Status Sheet

Issue	Date	Details
1.0	31/07/2020	First release of the document
1.1	15/10/2020	Addressing comments of RID of 24/08/2020
2.0	10/05/2022	Second version of the document
2.1	07/06/2022	Addressing comments of the TO sent on 17/05/2022

Document Change Record

Issue	Date	Request	Location	Details
1.1	15/10/2020	ESA	Executive summary Sections 2.1, 2.3.3.1, 2.3.3.2 Section 2.2 Section 2.3 Section 2.3.1 Section 2.3.1.2 Section 2.3.2 Section 3.2	Minor grammar changes Sections updated Last paragraph updated Figure 2 updated Caption of figure 3 updated Minor changes in the text Minor changes in the text and Figure 11 updated Figure 39 updated
2.0	10/05/2022	CNR UAH	Sections 1, 2.2, 2.4.1, 2.4.1.1, 2.4.1.3, 2.4.3, 2.4.3.2, 2.5, 2.5.1.1, 2.5.1.2, 2.5.1.3, 2.5.2, 2.6, 2.6.1, 2.6.2, 3 and sub-sections Sections 2.1, 2.4, 2.4.1.2, 2.4.2, 2.4.3.1, Section 2.3, 2.5.3 and sub-sections Section 4	Text updated Small changes in the text Sections added New references added
2.1	07/06/2022	CNR	Section 2.2	Figure 2 updated.

Issue	Date	Request	Location	Details
		UAH	Sections 2.4.1, 2.4.1.1, 2.4.3, 2.5.2 Sections 2.4.3.2, 2.5 Section 2.5.1.2 Section 3.1	Small changes in the text. Text expanded. Figures 21, 22 and 23 updated. Text expanded, figure 38 updated.

Table of Contents

List of Tables

List of Figures

1 Executive Summary

The Product Validation and Intercomparison Report (PVIR) describes methods and results of tasks carried out for assessing the quality of burned area (BA) global and regional products derived by applying Fire_cci algorithms. In particular, the validation of global FireCCI51 and FireCCIS310, continental FireCCISFD20 and regional FireCCIS1S2AF10 BA products is the object of this document.

In order to assess the accuracy of the above-mentioned BA products, fire reference perimeters are derived from Earth Observation (EO) data at higher spatial resolution, when available. Validation of BA products requires systematic sampling of validation units in order to provide robust statistical accuracy; for this reason, the most suitable sources of data for deriving reference fire perimeters are commonly based on Landsat 8 (L8) and Sentinel-2 (S2) for global and regional/continental products, respectively.

A separate case is the validation of BA products over test sites in Africa where the SAR-Optical Algorithm has been developed and tested (Belenguer-Plomer and Tanase, 2020). In this case, Planetscope data at 3 m spatial resolution were used to assess algorithm performance.

2 Validation of Fire_cci products

2.1 Validation protocol

Validation is the assessment of the accuracy of BA products by comparison with reference data/fire perimeters that are assumed to be the best representation of the ground truth. At global, continental and regional scales, reference data suitable for validation can be extracted primarily from EO data systematically covering the Earth surface; indeed, systematic collection of reliable and representative ground/in situ fire data is hardly feasible to be achieved over large areas.

The major key requirements for reference datasets derived from EO data are i) to be highly accurate and ii) to be generated independently. Independence is a critical characteristic of any validation assessment, since it assures that unbiased accuracies are obtained among products. Independence implies that reference datasets devoted to validation should not be used during the calibration or “tuning” of BA algorithms. EO data with higher spatial resolution compared to the data used for deriving the BA product should be used as long as they provide spatial and temporal systematic coverage of the Earth surface.

Validation of BA products derived from moderate spatial resolution EO data (e.g. MODIS) can be carried out by exploiting data provided by decametric spatial resolution data such as those provided by the Landsat and Sentinel missions (Chuvieco et al., 2018).

In the case of BA products derived from EO data such as Landsat and/or Sentinel (e.g. Roteta et al., 2019, Roy et al., 2019), higher spatial resolution could be provided only by high/very high resolution (HR/VHR) remotely sensed data. Yet this data is characterized by limited temporal and geographical coverage thus not assuring systematic sampling in time and space for the estimation of statistically robust and unbiased accuracy metrics. Moreover, systematic acquisition of HR/VHR data over large areas might be not easily accessible and/or might have high costs of acquisition. In these cases, satellite images with comparable spatial resolution are commonly used (e.g. Roteta et al., 2019, Roy et al., 2019) and independence is assured by independent processing.

In this framework, and within the Fire_cci project, reference fire perimeters will be mainly built from EO data collected from Landsat and Sentinel missions, which can assure systematic and sufficiently frequent acquisitions over the globe.

For what concerns BA products over the three test sites in Africa, spot acquisitions of HR/VHR EO data could be considered given the limited size of the test areas. This data will complement the validation carried out with medium resolution imagery (mainly Sentinel-2).

Therefore, the protocol implemented for validation of global and regional African BA products is composed of the following steps:

- Identification of validation units: sampling units, which are designed as 100 km by 100 km regions are selected by stratified random sampling in each calendar year, taking into account the major Olson biomes (Olson et al. 2001) and regions with high and low fire activity as depicted by the FireCCI51 product (Lizundia-Loiola et al., 2020). Specific sampling schemes are adopted for extracting validation units at global and continental scales and based on L8 and S2 acquisition and archive systems, respectively.
- Generation of reference fire perimeters: the area covered by the validation units is subsampled using a 100 km by 100 km window located in the centre of the scene area (i.e. region). Reference fire perimeters over the region are generated from supervised classification of consecutive satellite images, i.e. temporal series of images (L8 for global and S2 for regional Africa). A maximum time step between consecutive images is set to 16/32 days to assure a clear discrimination of the burned surface signal. A 16-day time step is preferred to preserve a clear burned signal of recent burns in those ecosystems where the signal quickly disappears (e.g. tropical savanna, Padilla et al., 2014); however, in ecosystems where the persistence of the burned areas is longer and in presence of frequent cloud cover, the time step could be increased up to 32-days.
- Computation of accuracy metrics: the error matrix (Congalton and Green 1999; Latifovic and Olthof 2004) is extracted by comparing BA products and reference fire perimeters and the following accuracy metrics are computed: commission error ratio, omission error ratio, Dice Coefficient (DC) (Dice 1945), bias and relative bias.

2.2 Validation units: definition

A **validation unit** can be described as a spatio-temporal sampling of the L8/S2 archive used to build temporal series of consecutive scenes suitable for extracting fire reference perimeters (Franquesa et al., 2020).

Spatially the area to be validated (global and continental Africa) is divided into non-overlapping regions to assure equally distributed sampling probability between units. For validating global BA products, the spatial dimension of sampling units will be based on Thiessen Scene Areas (TSAs) constructed over the Landsat WRS-2 by Cohen et al. (2010) and Kennedy et al. (2010) and exploited in previous Fire_cci phases (Padilla et al. 2014; 2015, 2017; Lizundia-Loiola et al., 2020). For validating the regional-Africa BA product (FireCCISFD20), the non-overlapping regions are assumed coincident with S2 tiles (see section 2.5.1) to simplify data download and processing.

However, compared to previous Fire_cci phases, the region for extracting fire reference perimeters within each unit was increased to 100 km by 100km: hence, the validation unit is a 100 km by 100 km region of a L8/S2 scene, centred on the L8 frame or S2 tile.

Temporally, a **short unit** is assumed to be composed of two consecutive cloud-free L8/S2 scenes (pairs of images) whereas a **long unit** is composed of more than two consecutive pairs (consecutive short units) (Figure 1). Notice that time between consecutive images is set a-priori but does not necessarily coincide with theoretical data acquisition frequency of the satellite mission (L8 and/or S2); hence, the term "consecutive" indicates temporally adjacent images with a preliminarily defined maximum temporal distance (i.e. time step). The issue of using short and long reference units has been addressed by Padilla et al. (2018) and Franquesa et al. (2022) to show that using long temporal units provides an alternative approach to estimate spatial accuracy by disentangling the spatial from the temporal component of BA detection errors. Using long reference units provides a method to quantify the spatial accuracy of a BA product while accommodating the dating errors ono the product assessment. This is particularly relevant for those products with low temporal resolution or for areas with high cloud coverage (Franquesa et al., 2022).

An L8/S2 scene is assumed to be cloud free when cloud cover percent (from metadata information associated with Level 2 data products) is less than a given threshold. In the case of L8 for the validation of global BA products, the threshold was set to 30% (Figure 1)**.** In the case of S2 for the validation of the regional-Africa BA product, the threshold for single scene maximum cloud cover was lowered to 15%, taking advantage of the higher frequency of acquisition of S2 missions (Figure 2). Moreover, an additional condition on maximum cumulated cloud cover over the long reference unit was introduced to generate clearer units for extracting fire reference perimeters.

The following parameters are therefore defined to identify long temporal sampling units:

- **Cloud cover** for each scene and **cumulated cloud cover** for the long temporal unit (only for S2) [%]: CC and CC_{cum}.
- **Time interval** between consecutive pairs (i.e. short unit) [days]: $\Delta t = t_k - t_{k-1}$
- **Length of the long unit** (i.e. long unit) [days]: $L = t_n - t_1$

where within the temporal series of L8/S2 images: $t_1$is the acquisition date of the first scene, $t_n$is the acquisition date of the last scene, $t_k$is the acquisition date of any scene within the unit. Validation unit length (L) is the time between first and last L8/S2 scenes, where the last date is when the temporal series ends due to a time interval greater than the maximum allowed between images with cloudiness below the threshold.

The long unit is identified by setting the **maximum time interval** (Δt_{max}) and the **minimum unit length** (L_{min}) for time series of consecutive scenes with cloud cover below the threshold. By setting these parameters, the population of L8/S2 sample units within the L8/S2 data archives can be identified. Ideally, a time step Δt_{max}=16 days guarantees the greatest discrimination between burned and unburned surfaces; this condition becomes particularly relevant in those ecosystems where the duration of the post-fire spectral signal is short (Padilla et al., 2014). But in case of persistent cloud coverage, a conservative value of Δt_{max}=16 days could significantly reduce the length of the validation unit. Therefore, the time interval between pre- and postfire images was initially set at 16 days and it was extended when cloudy images were present in the time series (Franquesa et al., 2022). Since Δt_{max} and L_{min} significantly influence the number of suitable sample

units depending on L8/S2 image availability, a preliminary analysis of L8/S2 archives was carried out to investigate data availability.

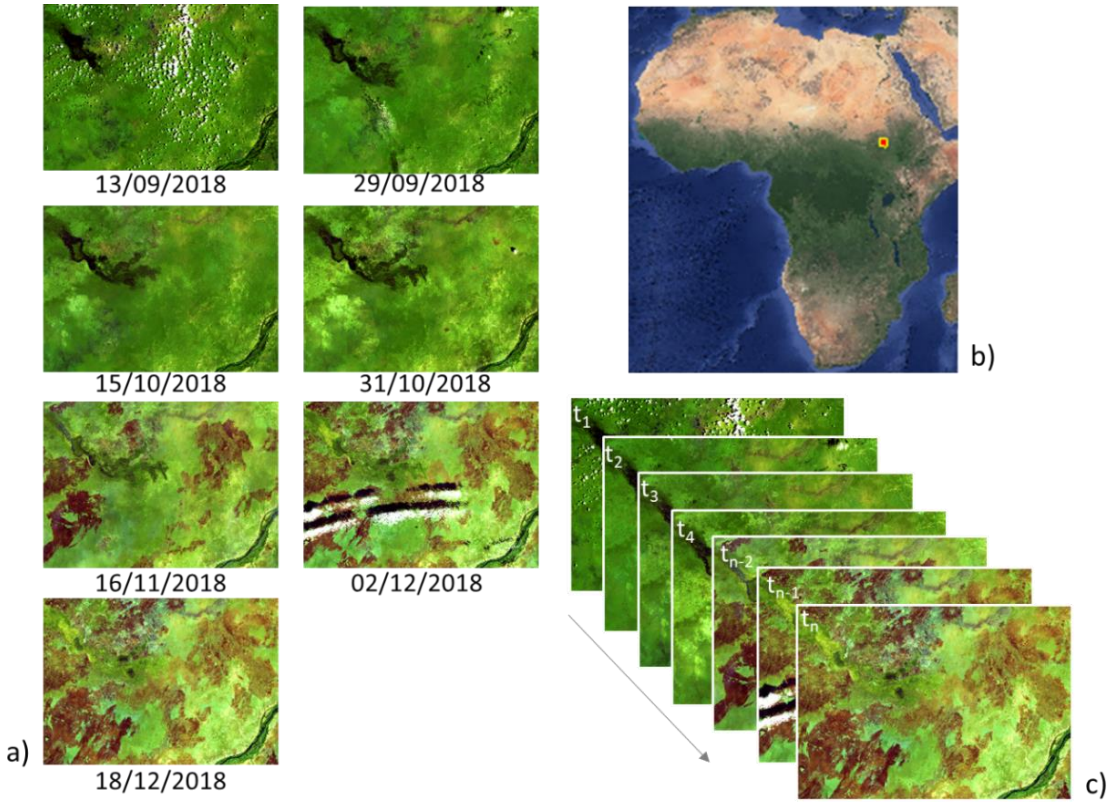

Figure 1: Example of short and long validation units for Landsat frame 173/053 a) consecutive scenes available; b) location of the validation unit; and c) long unit composed of consecutive L8 pairs acquired with a 16-day time step between two dates: t_1 and t_n.

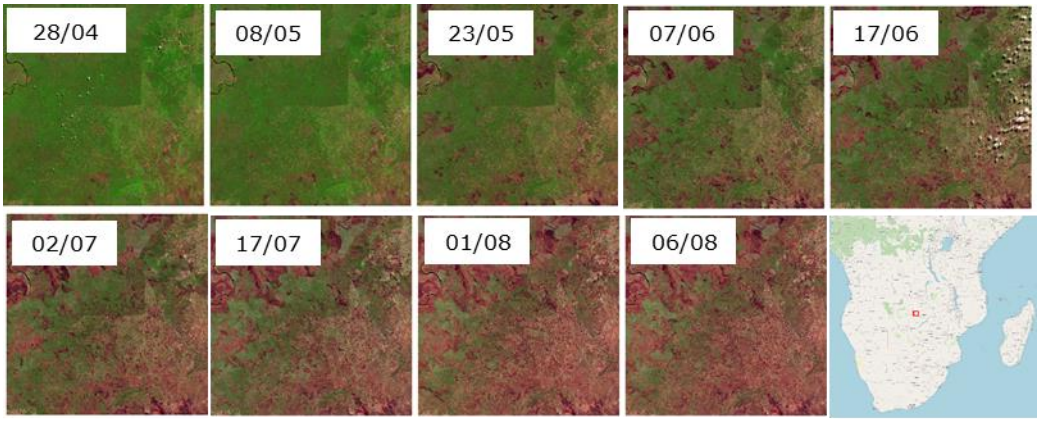

Figure 2: Example of long sampling unit for Sentinel-2 tile 35LMD, Zambia (Savanna biome) where consecutive image pairs are selected with cloud cover less than 15%.

2.3 Planetscope image characteristics

PlanetScope is a constellation composed by more than 120 optical satellites (also named Doves) operated by Planet since 2016. Each Dove satellite is a CubeSat 3U form factor (10 cm by 10 cm by 30 cm) (https://earth.esa.int/web/guest/missions/3rd-party-

missions/current-missions/planetscope, accessed July 2020). The sensor mounted on this platform is characterized by four acquisition bands: three in the visible wavelengths (b1: 455-515 nm; b2: 500-590 nm; b3: 590-670 nm) and one in the NIR wavelengths (b4: 780-860 nm). Planetscope has a swath of about 25 Km. The spatial resolution of PlanetScope images is 3 m for all bands. Imagery is captured as a continuous strip of single frame images known as "scenes". Table 1 summarizes PlanetScope mission and sensor characteristics (Lemajic et al., 2018).

Table 1. PlanetScope Constellation and Sensor Specifications (Source: Lemajic et al., 2018).

MISSION CHARACTERISTICS	SUN SYNCHRONOUS ORBIT
Orbit altitude (reference)	475 Km (-98° inclination)
Max/min latitude coverage	± 81.5° (depending on season)
Equator crossing time	9:30 – 11:30 am (local solar time)
Sensor type	Three-band frame imager or four-band frame imager with a split-frame NIR filter
Spectral bands	Blue 455-515 nm Green 500-590 nm Red 590-670 nm NIR 780-860 nm
Ground sampling distance (nadir)	3.7 m (at reference altitude 475 Km)
Swath width	24.6 Km x 16.4 Km (at reference altitude)
Maximum image strip per orbit	20.000 Km2
Revisit time	Daily at Nadir (early 2017)
Image capture capacity	150 million Km2/day (early 2017)
Camera dynamic range	12-bit

2.4 Validation of global BA products

In this section, the approach implemented for building reference fire perimeters for the validation of FireCCI BA global products is described in detail. Figure 3 shows the flowchart of all the steps, from the selection of the sample units to the extraction of fire perimeters over the long unit.

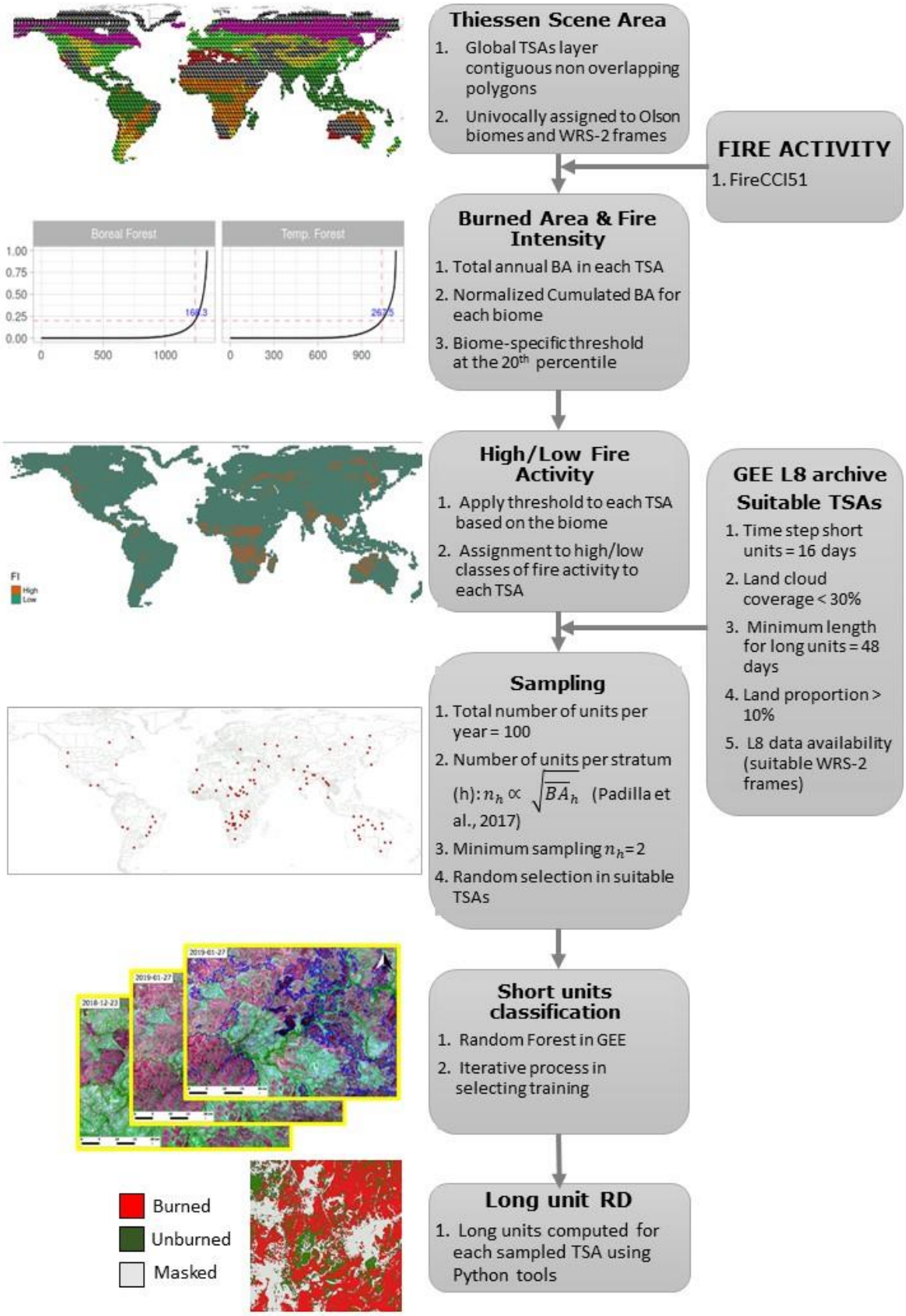

Figure 3: Flowchart of steps for extracting fire reference perimeters for the year 2018 for the global validation: from sampling to classification of the short units and building long units files.

2.4.1 L8 Sampling scheme definition and implementation

Since sampling sites should be selected to properly represent the variety of conditions that affect the accuracy of BA products, both in time and space, a stratified random sampling scheme is adopted. In particular, following Franquesa et al. (2022), stratification has been based be based on i) major Olson biomes (Olson et al. 2001) and ii) areas with high and low fire activity, as derived within each biome from FireCCI51 BA product (Chuvieco et al., 2018) (Figure 4). Olson biomes were aggregated into eight major vegetated biomes as shown in Table 1 in Franquesa et al. (2022) and in Figure 4a. The eight vegetated classes are shown in the legend of the figure together with class "Rock & Ice" that was not considered in the sampling stratification since it does not burn. The applied biome aggregation scheme was the same as used by Boschetti et al. (2016), with the exception that in this study tundra was not aggregated to boreal forest.

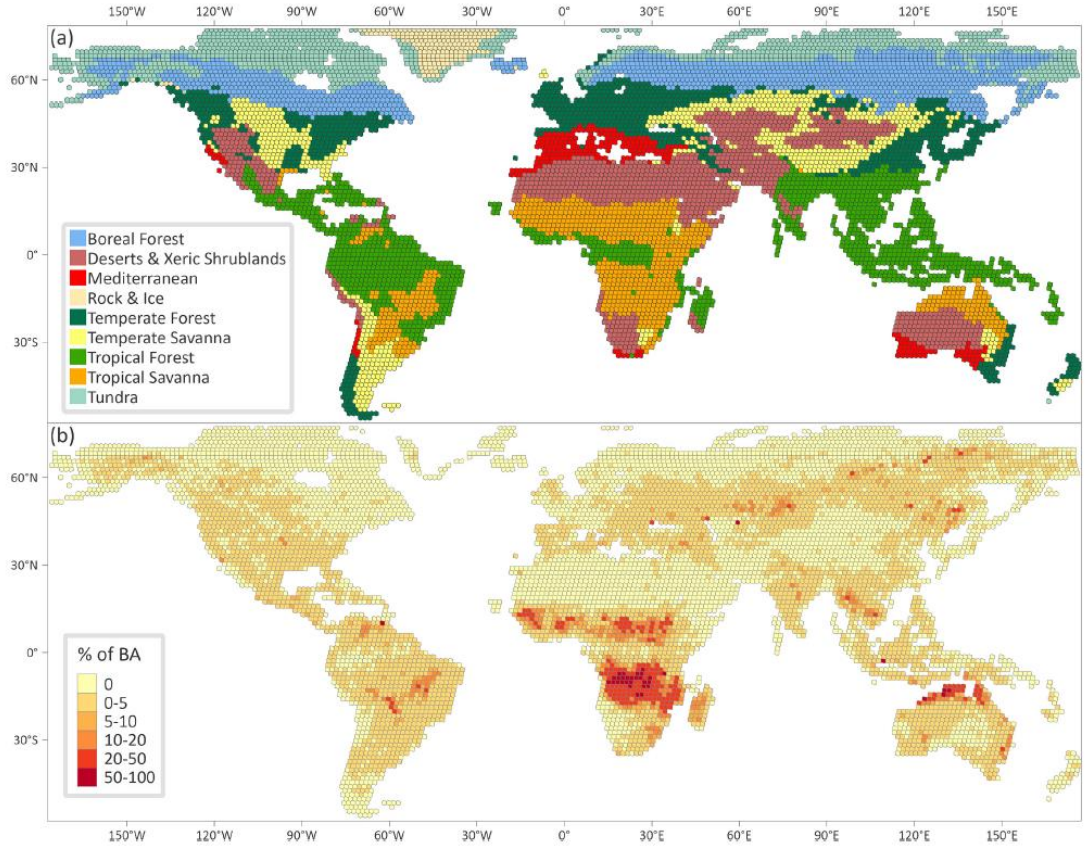

Figure 4: Two layers were used for the definition of the strata in the random stratification sampling: Olson biomes (top panel) reclassified into 8 major biomes (plus Rock & Ice shown as beige), and percentage of BA with respect to the total land area of each TSA, based on the FireCCI51 BA extent for the year 2019 (bottom panel) (Franquesa et al., 2022).

Specifically, for global BA products, the spatial dimension of sampling units is based on Thiessen Scene Areas (TSAs), as explained in Section 2.2. The number of selected units for each year is equal to 100 TSAs.y^{-1} over the period 2017-2019 (tot = 300).

For each TSA, a long sampling unit is composed of consecutive pairs of Landsat images ("short units"). By doing so, the time period covered by the reference fire perimeters will be sufficiently long to reduce the impact of dating errors on the spatial accuracy estimates. This use of long units rather than short units (pairs of L8 images) for extracting reference perimeters, represents an improvement over previous Fire_cci phases. Additionally, for

each sampled TSA an area of approximately 100 km by 100 km was selected (i.e. region) for defining reference fire perimeters; with respect to previous Fire_cci, the area has been increased from 20 km x 30 km to allow a more robust analysis of fire size and patches.

2.4.1.1 L8 sampling stratification

As stated above, strata for implementing stratified random sampling of validation units were derived from the intersection of Olson biomes and fire activity layers. For each TSA, the major Olson biome and high/low fire activity class were assigned (Figure 4). For each TSA, the amount of burned area is derived from the annual FireCCI51 BA product (Lizundia-Loiola et al., 2020). From year to year, fire area and fire activity might significantly change thus requiring a specific source of information for each considered year to be validated. With respect to previous Fire_cci phases, the source of information for classifying into high/low fire activity was updated from lower spatial resolution MCD64A1 Version 6 Burned Area data product to the higher resolution FireCCI51 BA product.

In order to assign the high/low fire activity class, the total annual burned area (TotBA, m^2) was computed in each TSA and these values were then sorted in increasing order; the cumulated sum was computed and normalized with respect to the biome's maximum value of total annual BA. The TotBA value corresponding to the 20th percentile of the normalized cumulated sum (Padilla et al. 2017; Boschetti et al., 2016; Franquesa et al., 2022) provides the threshold for assigning each TSA to either the ***high*** (total annual BA > threshold) or ***low*** (total annual BA <= threshold) fire activity classes. In Figure 5, the threshold values (highlighted in red) for each biome for the year 2017 are reported. Threshold values express the 80th percentile of the BA proportion of the population of TSAs in each biome. In the case of biomes with very low burned area, this value is close to zero; the lowest is estimated for the "Tundra" class were no burns are detected (see also Figure 5, panel b).

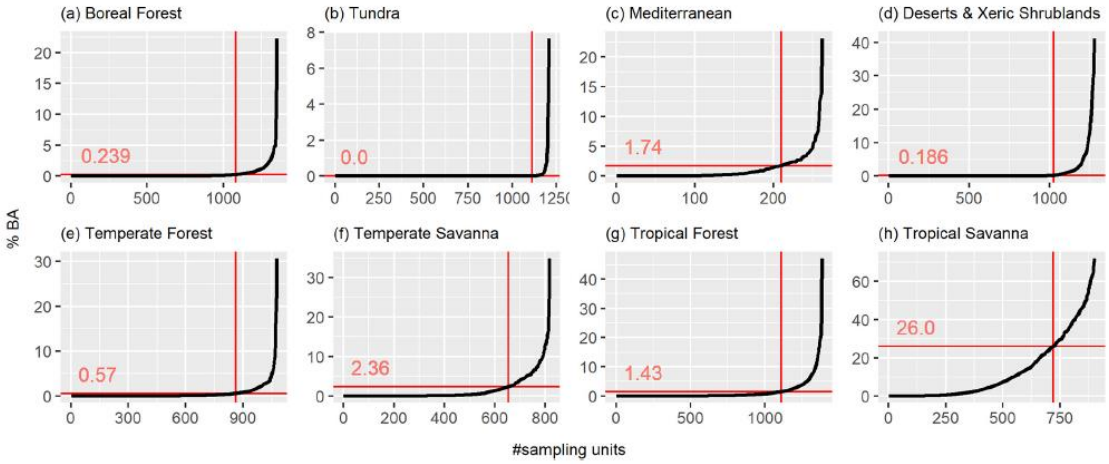

Figure 5: Example Thiessen scene areas (TSAs) for each biome plotted by increasing normalized total burned area for the year 2017; on the x-axis the increasing cumulated number of TSA (#sampling units). Red lines shows the 20th percentile and the corresponding value of total annual burned area used as threshold for assigning high/low fire activity class (Franquesa et al., 2022).

2.4.1.2 Analysis of L8 data availability 2018

Preliminary analysis of data availability at the global scale was carried out for the year **2018** to investigate the distribution of L8 scenes suitable for sampling. The global L8

archive available in Google Earth Engine (GEE) was investigated to identify, over each validation unit, L8 temporal series **with maximum time step of 16 and/or 32 days between consecutive clear-sky images**.

In this context, L8 images are considered to be "clear-sky" if their cloud coverage percentage is lower than 30%. Information on cloud coverage percentage was extracted from metadata and, specifically, the "Land Cloud Cover" value percentage was used. In order to investigate L8 data availability, the length of the available temporal series (L) over the L8 TSAs was analysed by setting time steps (Δt_{max}) of either 16 or 32 days (see Section 2.2).

Figure 6 and Figure 7 show maximum length and starting month of the long units when Δt_{max}=16 days and Δt_{max}=32 days, respectively. As for previous figures, grey TSAs are those for which no suitable data is available. The figure confirms that Δt_{max} and L are correlated, hence relaxing the condition on maximum time step, from 16 to 32 days, produces longer series (light green and yellow colours).

Figure 8 shows as an example the number and location of available TSAs for each biome with a **maximum time step of 16 days (Δt_{max}=16)** for three values of minimum length (L_{min}): 32, 48 and 112 days. An analysis of the figure highlights that, with the exception of savanna and desert areas (the latter included in the "Other" category), availability of TSAs with length greater than 100 days is limited, in particular when Δt_{max}=16 days. Clearly, increasing the required minimum length of the validation unit, decreases the number of available TSAs for each biome especially for those ecosystems mostly affected by cloud cover (*e.g.* tropical regions). In particular, the number of suitable TSAs significantly decreases when L_{min}=112 days for almost all biomes except savannas (blue regions) thus leading to a spatially bias in TSAs availability.

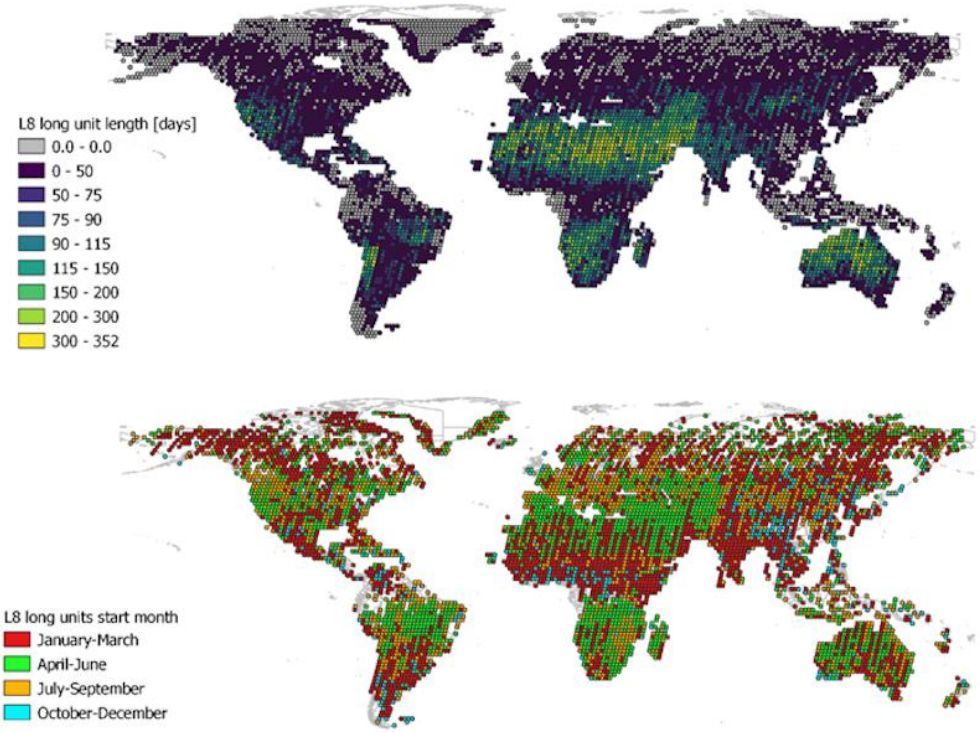

Figure 6: a) Length [days] (top) and starting month (grouped in three-month seasons) (bottom) of the longest temporal series for each TSA (validation unit) with a 16 days maximum time step between consecutive clear sky images.

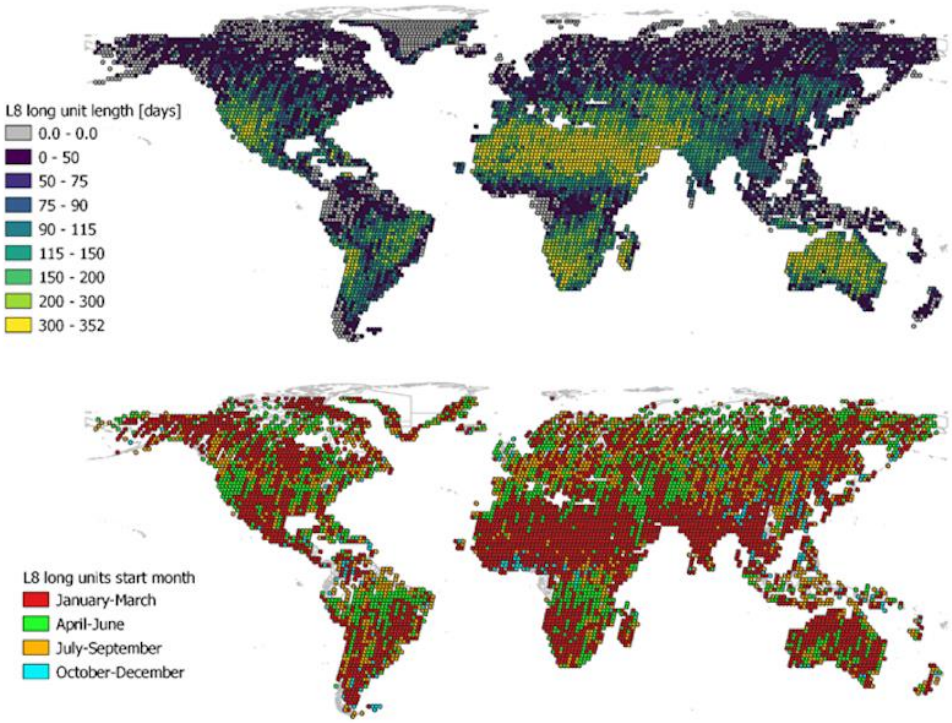

Figure 7: a) Length [days] (top) and starting month (grouped in three-month seasons) (bottom) of the longest temporal series for each TSA (validation unit) with a 32 days maximum time step between consecutive clear sky images.

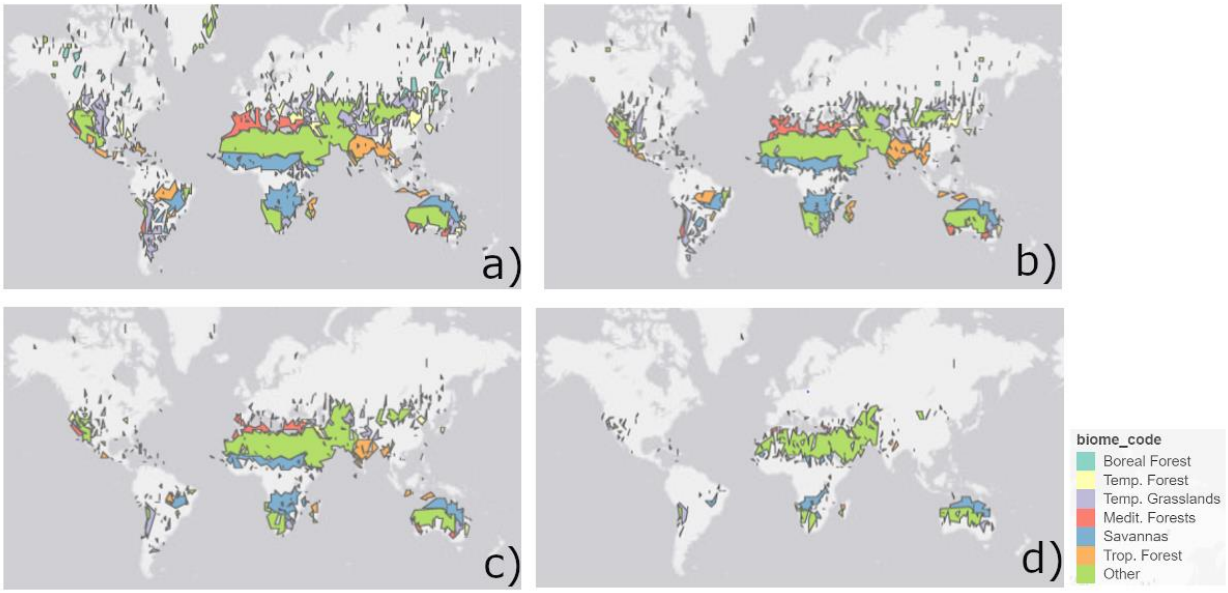

Figure 8: L8 TSAs (validation units) suitable for sampling with a 16 days maximum time step between consecutive scenes and variable total length of the temporal dataset [number of days] for each biome: 32 (a), 48 (b), 64 (c), 112 (d).

This analysis was further carried out by splitting biomes into sub-strata of high/low fire activity by applying threshold values. Figure 9 and Figure 10 show the number of TSAs available for sampling as a function of L_{min} (values on top strips) with Δt_{max}=16 days.

Based on the outcome of this analysis, in order to keep a sufficient number and an unbiased global spatial distribution for TSAs of the least represented biomes, the **minimum length of the validation unit was set to 48 days**.

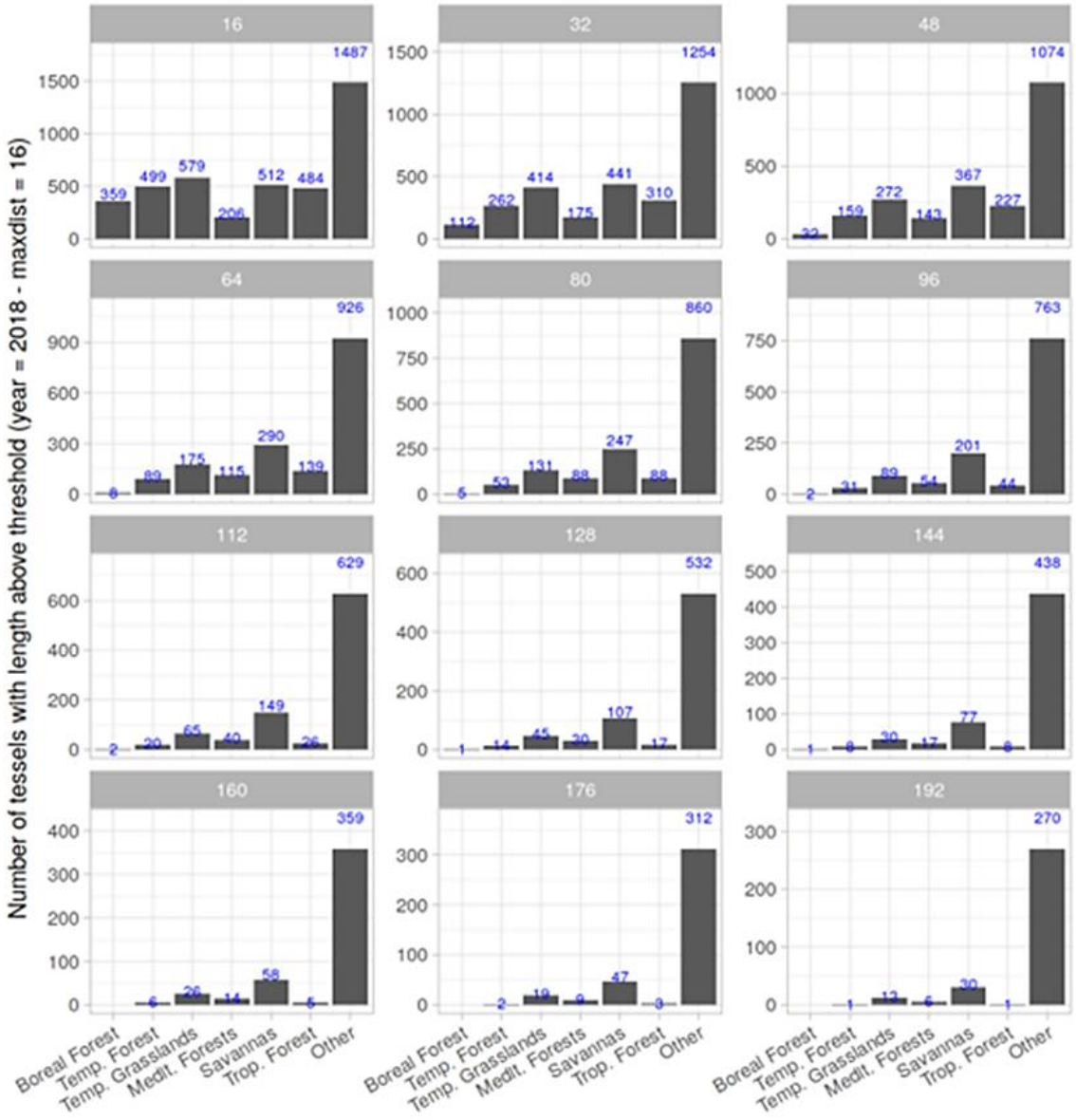

Figure 9: Number of L8 TSAs available for sampling for each biome in the Low Fire Activity stratum as a function of the length of the minimum temporal dataset/validation unit required (number on top of each panel) with maximum time step set to 16 days.

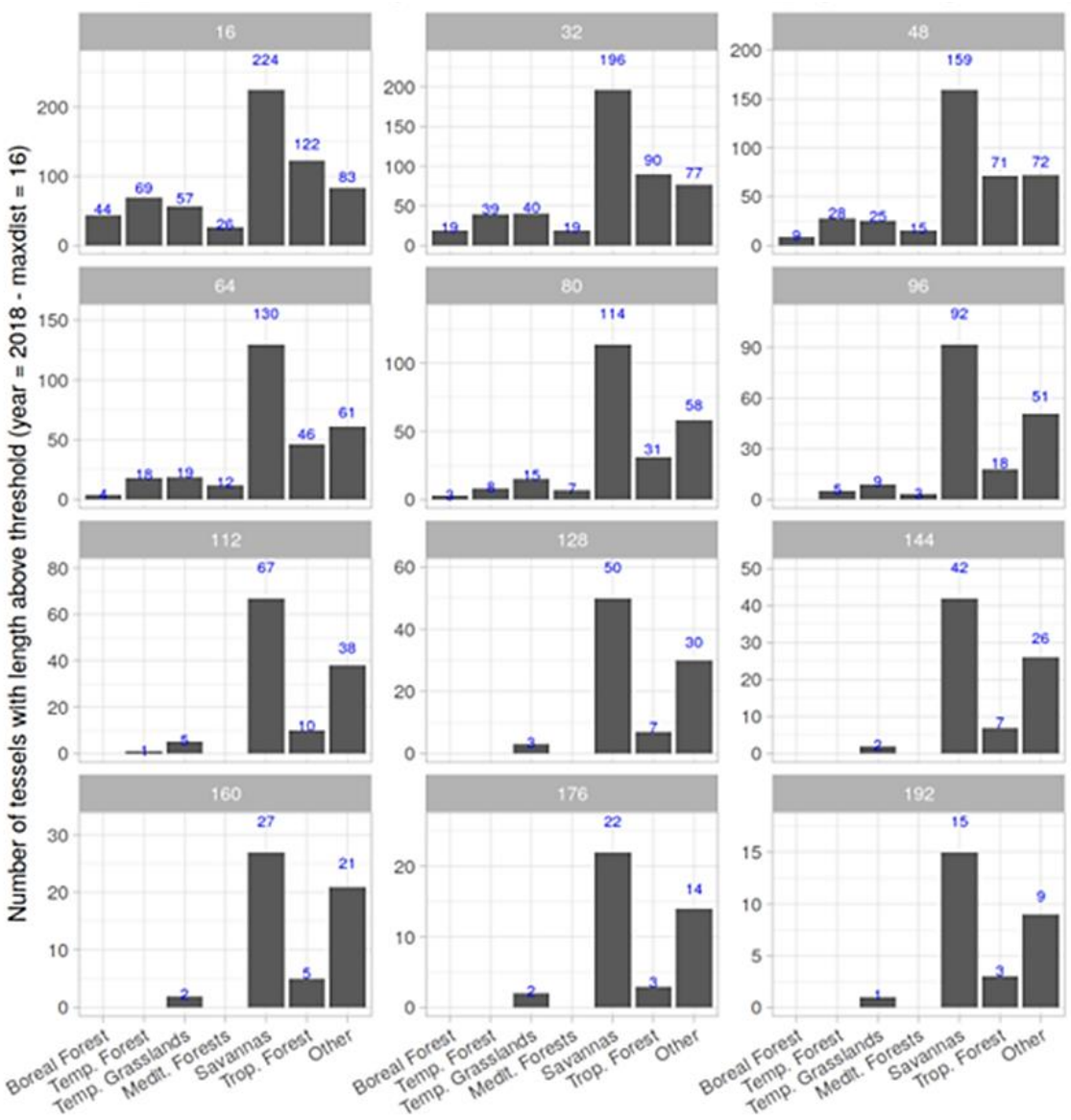

Figure 10: Number of L8 TSAs available for sampling for each biome in the High Fire Activity stratum as a function of the length of the minimum temporal dataset/validation unit required (number on top of each panel) with maximum time step set to 16 days.

2.4.1.3 Identification of sampling cardinality per stratum

An optimal stratum allocation was set to **100 validation** units per year and distributed among strata based on Eq. 1.

$$n_h \propto N_h \sqrt{\overline{BA}_h} \qquad \text{Eq. 1}$$

where n_h is the number of TSAs to be sampled for each stratum h, $\overline{BA}_h$ is the average total annual burned area for stratum h and N_h is the total amount of TSAs available for sampling for each stratum h. An iterative process was applied to ensure that $n_h \geq 2$ in all strata while preserving as much as possible the optimal sample allocation (Padilla et al., 2017; Franquesa et al., 2022). Once the sample size allocation was determined, the sample units were selected using simple random sampling without replacement from the

population within each stratum. This process was applied separately to each of the three years of the validation period (2017 to 2019) to select three samples, one for each year.

Results obtained with this procedure are summarized in Table 2 where for each stratum (combination of biomes and fire activity) we show the number of TSAs sampled.

Table 2. Number of TSAs sampled in each stratum and for each year in high (N_{HIGH}) and low (N_{LOW}) fire activity strata.

BIOMES	2017		2018		2019	
	N_{HIGH}	N_{LOW}	N_{HIGH}	N_{LOW}	N_{HIGH}	N_{LOW}
Boreal Forest	3	2	4	2	4	2
Deserts & xeric shrublands	9	2	7	2	6	2
Mediterranean Forests	2	2	2	2	2	2
Temperate Forest	4	2	5	2	4	2
Temperate savanna	5	5	4	3	5	4
Tropical forest	10	5	10	6	11	7
Tropical savanna	21	29	22	31	21	29
Tundra	2	2	2	2	2	2
Global	56	49	56	50	55	50
Sample size by year	105		106		**105**	

A random sampling algorithm is then applied to each stratum to extract the number of validation units shown in Table 2. The location of the sampled TSAs for the period 2017-2019 is shown in Figure 11.

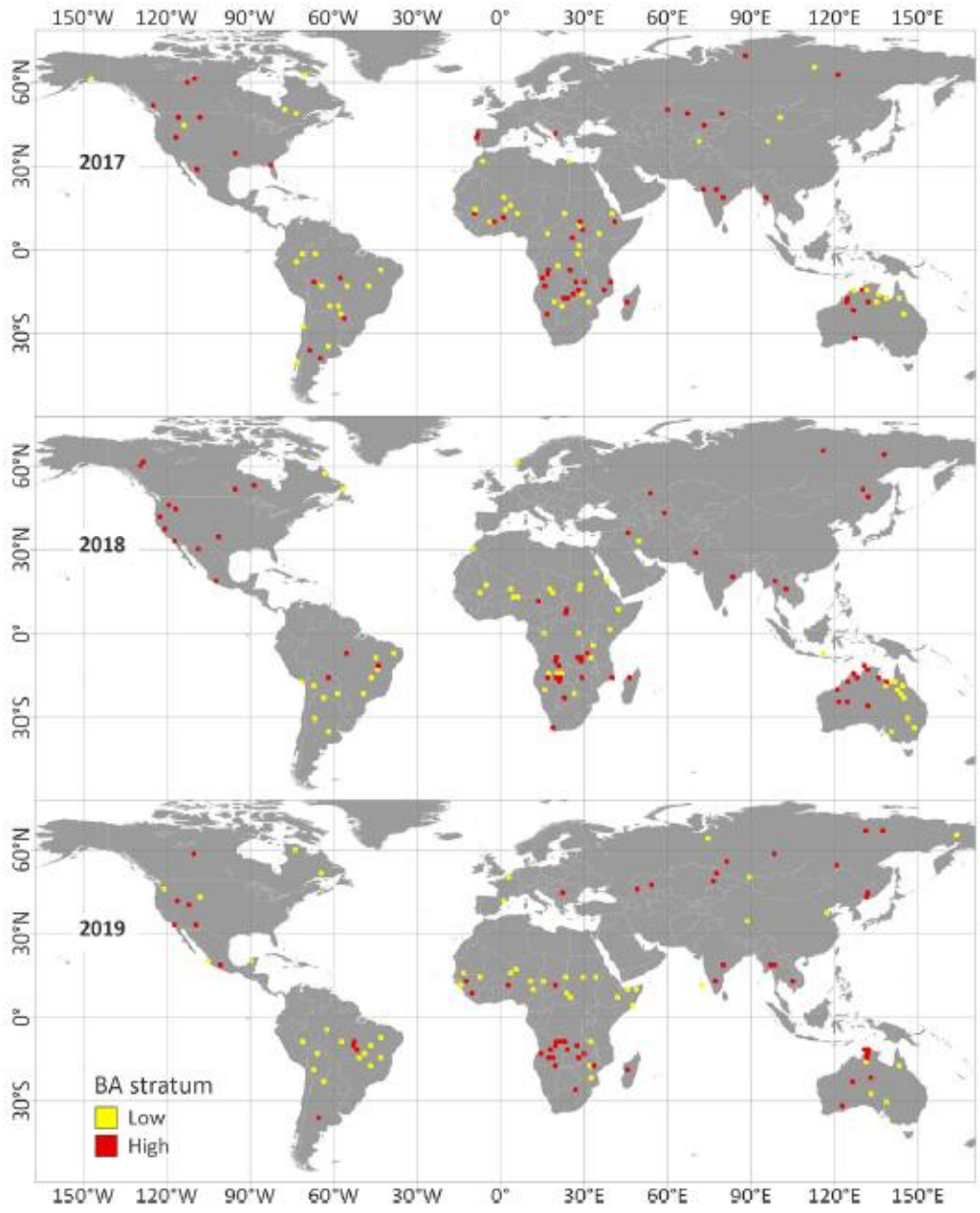

Figure 11: Spatial distribution of the TSAs sampled randomly worldwide for each year and stratum of fire insentity.

2.4.2 Extraction of L8 fire reference perimeters

Figure 12 shows the flowchart of the steps for extracting fire perimeters over L8 validation units composed of six major steps (Step 1 to Step 6 in the figure and in the text); input to the processing are L8 short units (consecutive L8 images) to extract areas that burned between the two dates (t_1, t_2). All short units over the same area are combined to derive fire perimeters over the L8 long unit. The steps are described below.

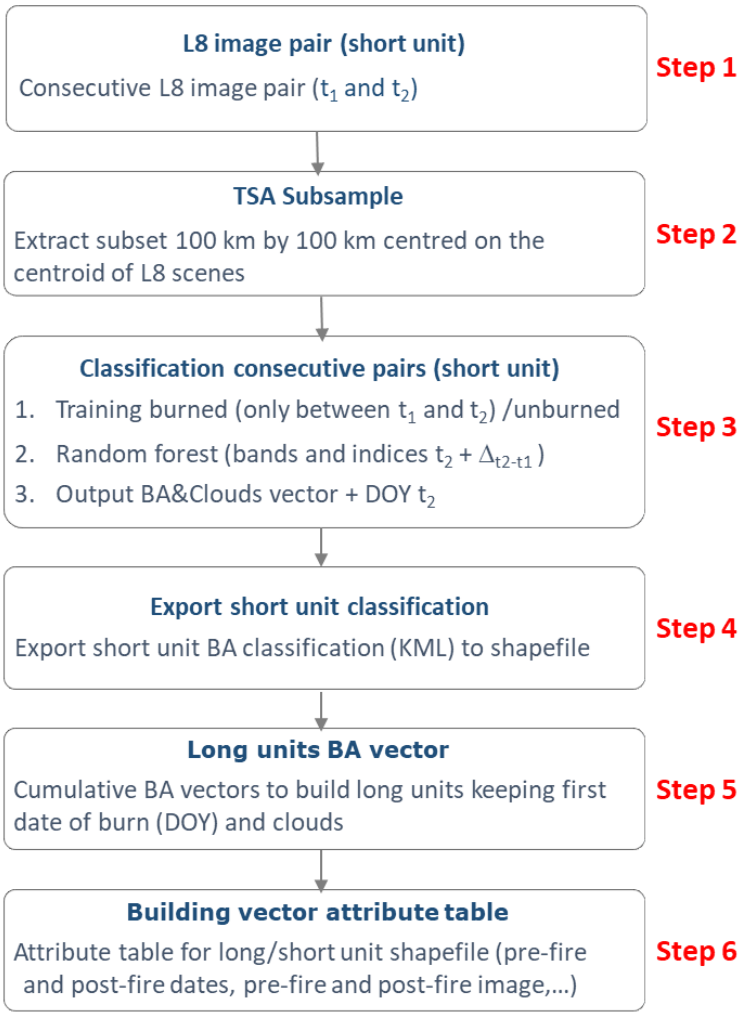

Figure 12: Flowchart showing steps for extraction of fire reference perimeters over sampled TSAs / validation units.

1. **L8 image pair (short unit)**: over each TSA, consecutive L8 scenes (Δt_{max}=16 days, condition relaxed in case of persistent cloud cover) made available as Level 2 product (surface reflectance, USGS Landsat 8 Surface Reflectance Tier 1 collection dataset) with cloud cover layer used for masking not observable pixels.
2. **TSA Subsample**: extraction of validation area of size 100 km by 100 km centred at L8 frame's centroid.
3. **Classification consecutive pairs (short unit)**: extraction of perimeters of the areas that burned between the two dates (t_1, t_2) by applying a Random Forest (RF) algorithm. This step requires the selection of training areas over burned and unburned surfaces carried out in GEE and by visual interpretation of the expert. The RF algorithm is run using all L8 spectral bands and the Normalized Burn Ratio (NBR, Key and Benson 1999), Normalized Burned Ratio 2 (NBR2, Roy et al. 2005) and Normalized Difference Vegetation Index (NDVI, Rouse et al. 1974). Details of the RF algorithm applied are described in Franquesa et al. (2022).
4. **Export short unit classification**: the layer of fire perimeters for the short unit is exported as a KML file; this file contains polygons for burned areas and clouds together with information on burn detection date (t_2 stored as Day of the Year-DOY).
5. **Long Unit BA vector**: all short unit vector files for a specific TSA are stored and converted to shapefile (.shp) and cumulative fire perimeters over the long unit are

derived by retaining first date of burn detection for each polygon; in this step, cumulated cloud cover is used for masking those pixels that have been observed as cloudy at least once during the long unit period. That is, only pixels that have been cloud free over the entire time period covered by the long unit are classified into burned/unburned. This step is carried out with a script coded in Arcpy.

6. **Building vector attribute table**: for both short and long units, attribute tables are built containing all information relative to each polygon about pre- (t_1) and post-fire (t_2) dates and L8 scene identifier (Scene ID), detection dates, category (burned, unburned and masked).

The core of the processing of the L8 short units (short unit classification) was implemented in GEE. A script was coded to perimeter existing burned areas between two consecutive images, identified by acquisition dates (Start, End). Input parameters to run the program are:

- L8 frame path/row;
- L8 dataset (the image collection to be filtered)
- Year;
- Starting date (t1);
- Ending date (t2);

The script needs to be run twice. The first time the dataset is filtered by the dates of input images (Pre_Image, Post_Image) and a cloud mask is applied. Then spectral indices NBR, NBR2 and NDVI and temporal differences are computed.

The images are displayed as RGB false colour composites (SWIR2, NIR, Red) and the training regions over burned and unburned areas are defined by the user as vector format. If available, training polygons can be uploaded as asset on the GEE platform.

The script is then run a second time to apply a Random Forest algorithm for classification of burned areas, using the input layers identified above (spectral bands, indices and their temporal difference).

The output layers consist of:

- Burned areas in vector KML format;
- The validation region of 100 km x 100 km obtained by a buffering of the centroid of the L8 frames;
- Cloud Mask in vector format;
- Training polygons in vector format;

The classification of the L8 short units is then converted to shapefile format and processed to extract fire reference perimeters, clouds and burn date.

Following the procedure to generate the reference data described above, we produced a total of 316 long temporal reference shapefiles for the three validation years (2017–2019).

Figure 13 shows an example fire reference perimeter product extracted from a L8 long validation unit over frame 173/053 (path/row, Africa). Masking cloud areas is necessary to achieve the highest accuracy in the detection on the burn dates; although burned surface and burned signal can be persistent and last over time, the occurrence of clouds over the area might prevent the accurate detection of the burn date. However, since cumulated cloud cover might significantly reduce the total area sampled at the end of the validation period over each TSA, additional criteria are under investigation to combine maximum

time step, unit length and cumulated cloud cover. These criteria will be tested for the S2 sampling approach over continental Africa.

Figure 14 shows an example of attribute table for fire reference perimeters over the long unit.

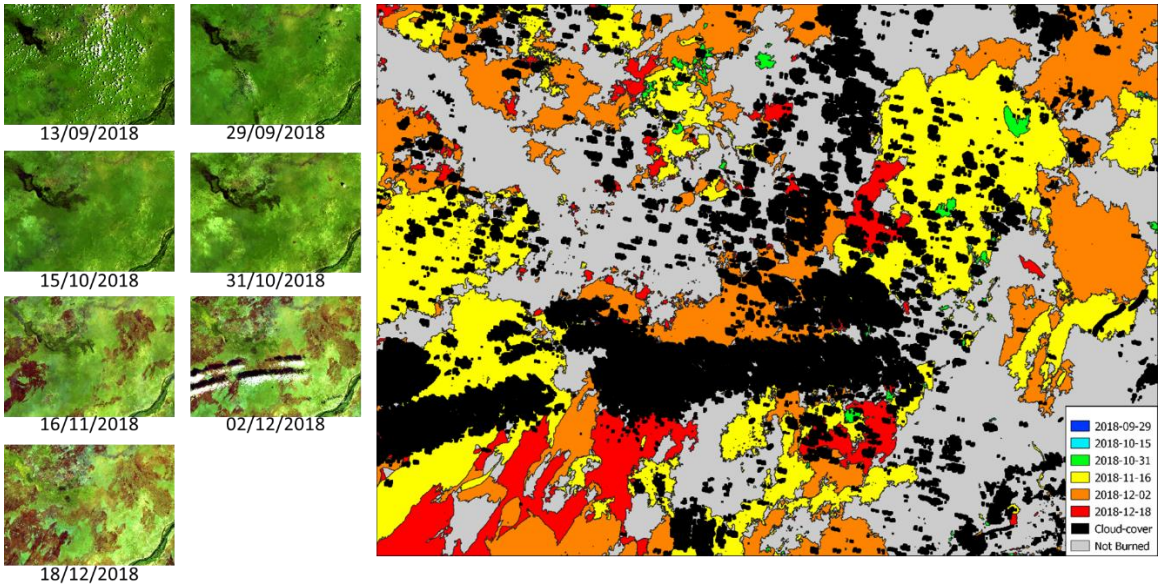

Figure 13: Example of reference fire perimeters extracted over L8 frame 173/053 (Path/Row), Africa; on the left RGB false color composites of the L8 scenes that are part of the validation long unit, in the right the reference burned area perimeters extracted by RF classification with reference to the date of detection (color of the polygons). Black regions are regions masked for cloud cover and grey areas are unburned.

category	preDate	postDate	preImg	postImg	path	row	year	area
3	2018-06-14	2018-10-20	LC81760712018165LGN00	LC81760712018293LGN00	176	71	2018	4139444454,420...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	291918,1587459...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	16721,86342880...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	5400,00000000000
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	193468,5110210...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	36000,00000000...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	43026,44852310...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	24176,49163130...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	186999,5500189...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	1609138,969650...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	71100,00000000...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	62173,63504820...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	31500,00000000...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	6334,84491956000
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	60300,00000000...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	7243335,088220...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	1350938,461050...
1	2018-10-04	2018-10-20	LC81760712018277LGN00	LC81760712018293LGN00	176	71	2018	43200,00000000...

Figure 14: Example of the attribute table of a reference fire perimeters shapefile over validation long units: category can be assigned to burned (1), cloud (2) and unburned (3), preDate and postDate are the pre-fire and post-fire dates of the short unit from which the polygon was extracted, preImg and postImg are the L8 scene ID of pre-fire and post-fire L8 images, path and row the WRS-2 L8 frame identifiers, year is the reference year and area is the area of each polygon.

2.4.3 Spatial accuracy of global BA maps

This Section summarizes the results obtained from the validation of FireCCI51 BA product for the year 2017-2019 and of FireCCIS310 for 2019.

2.4.3.1 Computation of accuracy metrics

Reference fire perimeters and Fire_cci global BA products were intersected to estimate the following accuracy metrics: commission error ratio, omission error ratio, Dice Coefficient (DC) (Dice 1945), bias and relative bias (Table 3). The area of agreement and disagreement between the BA product (map) and the reference data (Padilla et al. 2014; Padilla et al., 2017) (e_{ij}) used to estimate the accuracy metrics were derived from the confusion matrix (Table 4). Global accuracy metrics were estimated according to the sampling design and applying the formulas described in Franquesa et al. (2022).

Table 3. Accuracy metrics computed from the error matrix

Accuracy metric name	Formula
Commission error	$Ce = \frac{e_{12}}{e_{1+}}$
Omission Error	$Oe = \frac{e_{21}}{e_{+1}}$
Dice Coefficient	$DC = \frac{2e_{11}}{2e_{11} + e_{12} + e_{21}}$
Bias	$bias = e_{12} - e_{21}$
Relative Bias	$\boldsymbol{relB = \frac{e_{21} - e_{12}}{e_{+1}}}$

Table 4. Sampled error matrix on a sampling unit. e_{ij} express the proportion of agreements (diagonal cells) or disagreements (off diagonal cells) between the BA product (map) class and the reference class. Proportions for all pixels is derived by summing up the proportion of agreement/disagreement for each pixel at the resolution of the BA products (lower spatial resolution).

Product classification	Reference classification		Row total
	Burned	Unburned	
Burned	e_{11}	e_{12}	e_{1+}
Unburned	e_{21}	e_{22}	e_{2+}
Col. total	e_{+1}	e_{+2}	

2.4.3.2 Results of accuracy metrics

Global accuracy estimates of the three BA products and their standard errors (SE, results reported as ±SE) are presented for the entire validated period (2017–2019) in Table 5. Results at biome level are fully presented and discussed in Franquesa et al. (2022) and Lizundia-Loiola et al. (2022).

Table 5. Global accuracy estimates [%] for each validation year for FireCCI51 and FireCCIS310. Standard error is shown in parenthesis.

BA product	Year	OE	CE	DC	relB
FireCCI51	2017	41.8 (3.1)	21.4 (2.2.)	66.9 (2.3)	-26.0 (4.1)
	2018	41.3 (3.4)	15.7 (1.4)	69.2 (2.7)	-30.4 (3.3)
	2019	46.5 (3.4)	20.8 (1.7)	63.9 (2.8)	-32.5 (3.4)
FireCCIS310	2019	41.2 (3.0)	19.2 (1.7)	68.1 (2.5)	-27.2 (2.7)

The FireCCI51 shows a DC > 63% over the three-year period with a relatively small variability of the accuracy metrics across the three years of the validation period (2017–2019). FireCCS310 had a value of 68.1%.

Globally, omission error is, as expected, greater than commission due to the spatial resolution of the input data; detection of small fires is always an issue. However, omission error is slightly lower than other global BA products (Franquesa et al., 2022), especially in the case of FireCCIS310 (Lizundia-Loiola et al., 2022).

At biome level, differences were found from year to year and across biomes for FireCCI51, with highest accuracies achieved in Boreal Forest (data not shown, available in Table 6 of Franquesa et al., 2022). In the case of FireCCIS310, the accuracy metrics by biome for 2019, presented in Table 12, show that the highest accuracy was found for the Mediterranean biome. FireCCIS310 outperformed the BA results of FireCCI51 for 2019 in almost all biomes and accuracy metrics (in 28 out of 32 cases), as shown by the numbers in bold of Table 12. The DC values were higher for FireCCIS310 in all biomes, and the omission and commission values were lower in all biomes except in the case of the Ce in Tropical Forest.

As highlighted by results presented here and by Boschetti et al. (2019), a significant variability of accuracy metrics can be observed for the different biomes as a consequence of the fire/burns characteristics (e.g. type of fire, fire intensity, fragmentation of the fire patches, etc.). Further analysis on the global and per-biome accuracy estimates can be found in Franquesa et al. (2022).

2.5 Validation of regional/continental BA products

In this phase of the Fire_cci project, the **Sentinel-2 Small Fire Dataset (SFD)** (SFD Fire_cci v2.0, FireCCISFD20) (Pettinari and Roteta, 2021) has been produced for the year **2019** over the Sub-Saharan African continent. The FireCCISFD20 product is derived by processing Multi Spectral Instrument (MSI) Level-2A images, acquired by the Sentinel-2 A&B satellites (https://sentinel.esa.int/web/sentinel/missions/sentinel-2, accessed on April 2022) images acquired during a single year at continental scale (all Africa below latitude 25° N). This product is an extension of the previous SFD dataset FireCCISFD11, which had been processed for the year 2016. In the new product for 2019, the algorithm was able to ingest input surface reflectance from both Sentinel-2 A and B, duplicating the amount of input data. In order to validate the FireCCISFD20 product, a methodology was designed to build a burned area reference dataset from S2 images and

by implementing a stratified random sampling scheme. The proposed scheme builds on the experience inherited from the L8 sampling approach implemented at global scale (Section 2.4.1): the L8 protocol for the selection of the validation units was adapted to consider the peculiarities of S2 acquisitions and the ESA tile format. The proposed scheme implements the selection of S2 time series least affected by clouds for the definition of long temporal reference units (Franquesa et al., 2022). The same random stratified sampling approach was applied and strata were identified from the same sources of information: biomes and fire activity.

Since the size of the validation units was set to 100 km by 100 km, S2 tiles were directly used for spatial sampling. In order to identify, within the S2 archive, the population of tiles suitable for sampling and covering Sub-Saharan Africa, two additional selection criteria were defined to cope with S2 tiles encompassing different orbits and overlapping at the border of UTM zones (Section 2.5.1).

2.5.1 S2 sampling scheme and implementation

2.5.1.1 S2 sampling stratification

Sampling units, identified over sub-Saharan Africa (latitude range 25°N- 35°S), cover the area of a S2 tile (~10 000 km^2) where image time series are selected as ortho-images in UTM/WGS84 projection. The S2 tiling grid was directly used for spatial selection of the sample units for faster download and processing (e.g. cloud cover computation, archive access). However, in order to provide a robust statistical sampling and avoid overlapping between units, two criteria were applied to select suitable S2 tiles.

First, only S2 tiles fully covered by a single orbit, hence with same acquisition date over the tile's footprint, are retained (green polygons in Figure 15); tiles not entirely covered by almost one of the orbits are discarded (red polygons in Figure 15). An example of the 100 x100 km S2 tile area fully and partially covered by a single orbit is given in Figure 16. For the two tiles shown in the example, only the second one (Figure 16 bottom) is retained since data are acquired on a single date.

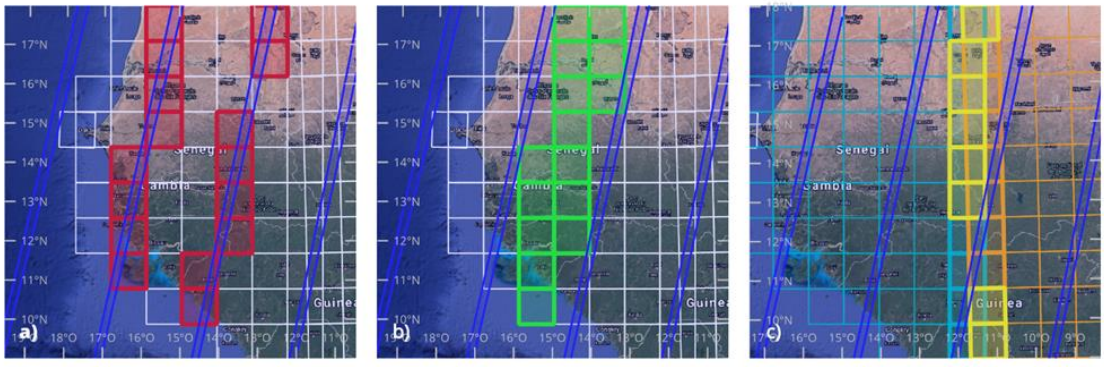

Figure 15: a) example S2 tiles not suitable (red) as TSAs for sampling since they cover different orbits; b) example S2 tiles suitable (green) as TSAs for sampling since they cover the same orbit; c) example S2 tiles overlapping adjacent UTM zones (yellow and orange).

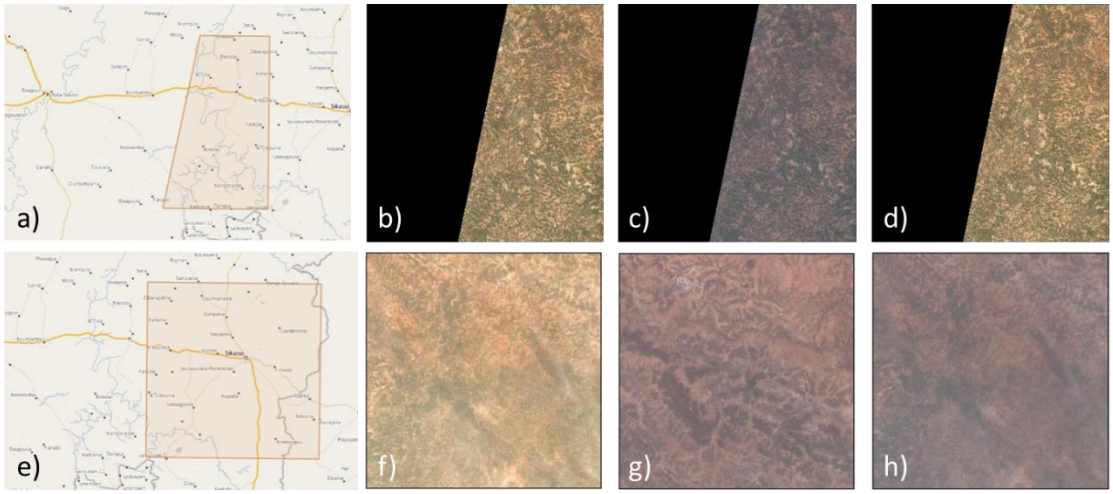

Figure 16: Examples of S2 tile images, as available from Copernicus Open Access Hub (https://scihub.copernicus.eu), for data acquired over Relative Orbit 008 (R008). Top row shows S2 images over tile T29PQN (partially covered by the orbit R008) acquired at different dates in 2020; bottom row shows tile T29PRN (fully covered by R008).

The second problem to be solved was found when S2 tiles overlap adjacent UTM zones (yellow polygons in Figure 15c). In these conditions, keeping both overlapping tiles would increase sampling probability of the common area. In order to address this issue, among the pairs of S2 tiles overlapping due to different UTM zones, only one was randomly selected and retained.

Once the two previous filtering conditions are applied to the S2 tiling grid system, the population of tiles suitable as validation units is composed of the subset of tiles shown in Figure 17a. For each of these tiles, i) the total annual burned area from the 2019 FireCCI51 BA product (Figure 17b) and ii) the major Olson biomes (Figure 17c) were computed. As in the case of the Global BA validation (Section 2.4.1.1), the intersection of these two layers provides strata for implementing stratified random sampling (Figure 19).

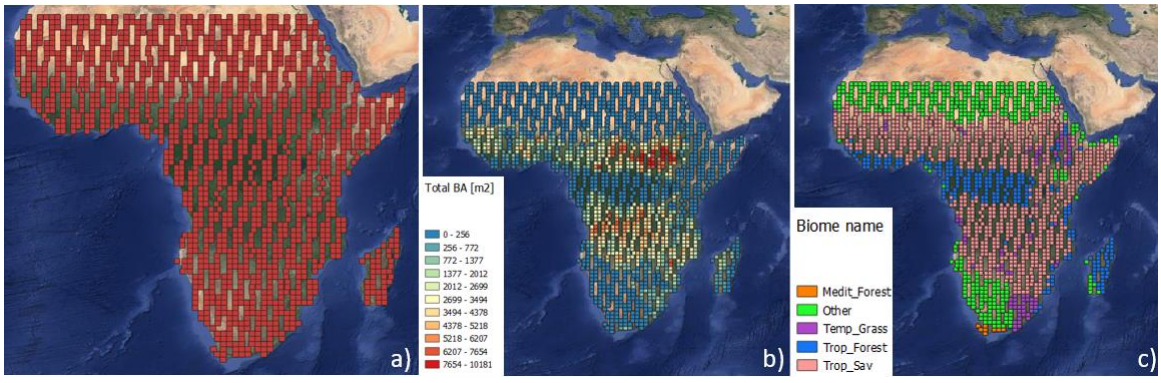

Figure 17: a) S2 tiles available for sampling after applying criteria outlined previously in this section; b) total burned area for each S2 [m^2]; c) Major Olson biome for each S2 tile.

In particular, the 2019 FireCCI51 BA product was exploited to divide each Olson biome into sub-strata of high and low fire activity by applying a threshold derived as in section 2.4.1.1.

In order to assign the high/low fire activity class, the total annual burned area (TotBA, m^2) was computed in each S2 tile and these values were then sorted in increasing order;

the cumulated sum was computed and normalized with respect to the biome's maximum value of total annual BA. The TotBA value corresponding to the 20th percentile of the normalized cumulated sum (Padilla et al. 2017, Boschetti et al., 2016) provides the threshold for assigning each S2 to either the ***high*** (total annual BA > threshold) or ***low*** (total annual BA <= threshold) fire activity classes. In Figure 18, threshold values for each biome and the year 2019 are reported in blue colour; values are also summarized in Table 6.

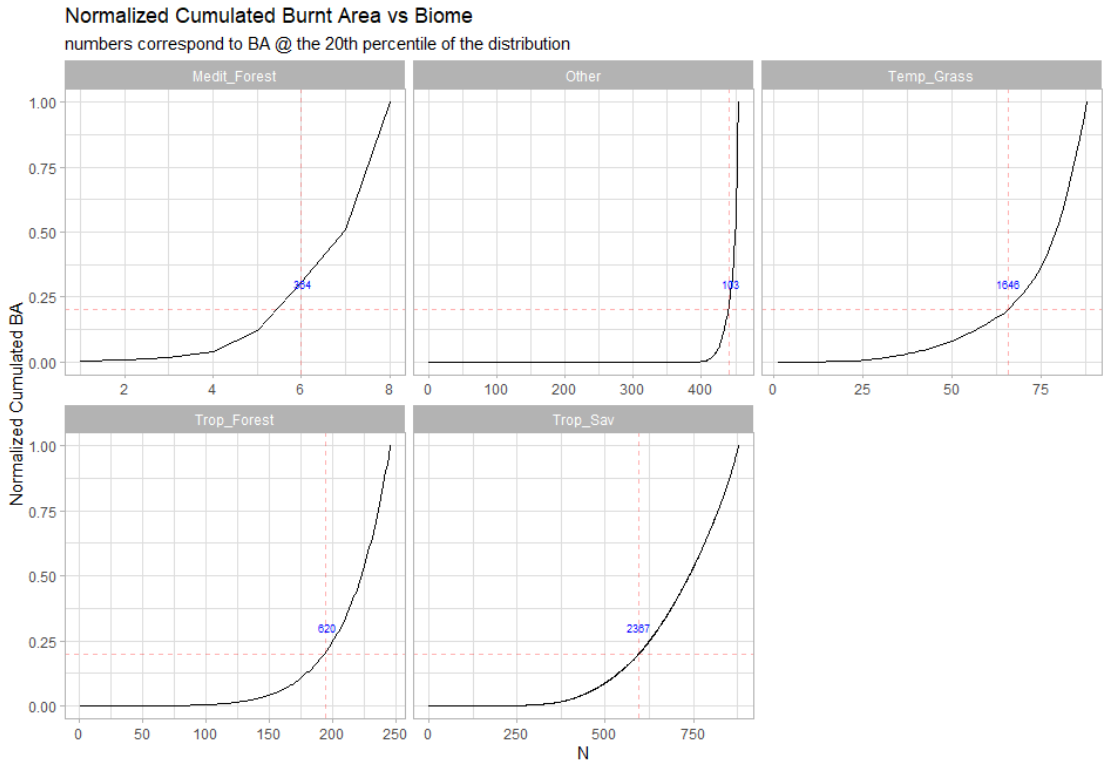

Figure 18: S2 tiles for each biome plotted by increasing value of normalized total annual burned area; on the x-axis the cumulated number of S2 tiles (N). The red dashed horizontal line shows the 20th percentile and the corresponding value of total annual burned area used for assigning each tile to high/low activity fire classes is highlighted in blue.

Table 6. Threshold values identified for stratification of each biome of the African continent into high/low fire activity strata.

Biomes	Threshold [km^2]
Mediterranean forest	364
Tropical savanna	2367
Temperate grassland and savanna	1646
Tropical forest	620
Others	103

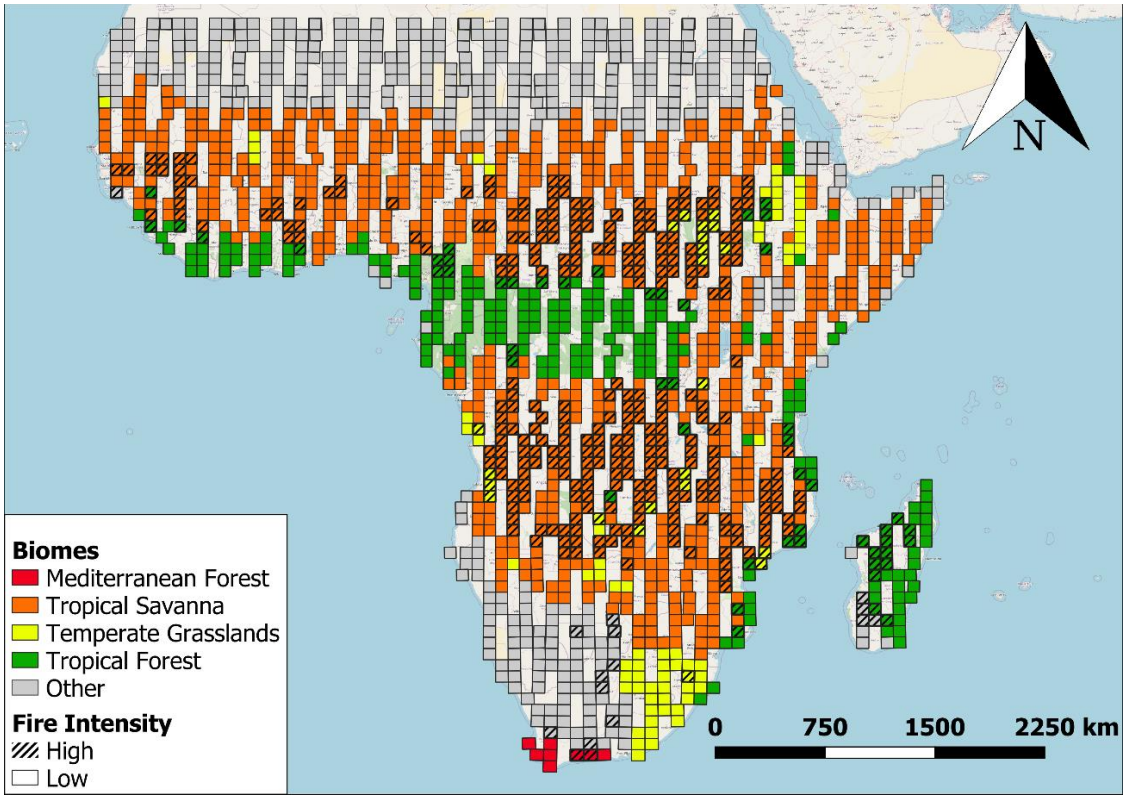

Figure 19: S2 tiles suitable for sampling for each stratum (intesection between biome and high/low fire activity).

2.5.1.2 Analysis of S2 data availability 2019

Preliminary analysis of data availability over the population of S2 tiles available for sampling was carried out for the year 2019 (as done for global L8, see section 2.4.1.2). The S2 archive was inspected to identify, for each tile, the S2 temporal series **with maximum time step (Δt_{max}) of 16 and/or 32 days between pairs of consecutive clear-sky images.** In this context, S2 images are considered to be "clear-sky" if their cloud coverage percentage is lower than 30%. Information on cloud coverage percentage is extracted from the S2 metadata and, specifically, the sum of *"High probability clouds percentage"* and *"Medium probability clouds percentage"* was assumed as total cloud cover.

Figure 20 shows the length [L, days] of the validation long units over S2 tiles suitable for sampling computed by considering Δt_{max}=16 days (panel a) and Δt_{max}=32 days (panel b). The length of the unit ranges between minimum values in the tropical regions (below 30 days) and maximum values for the northern and southernmost areas and desertic regions (greater than 300 days).

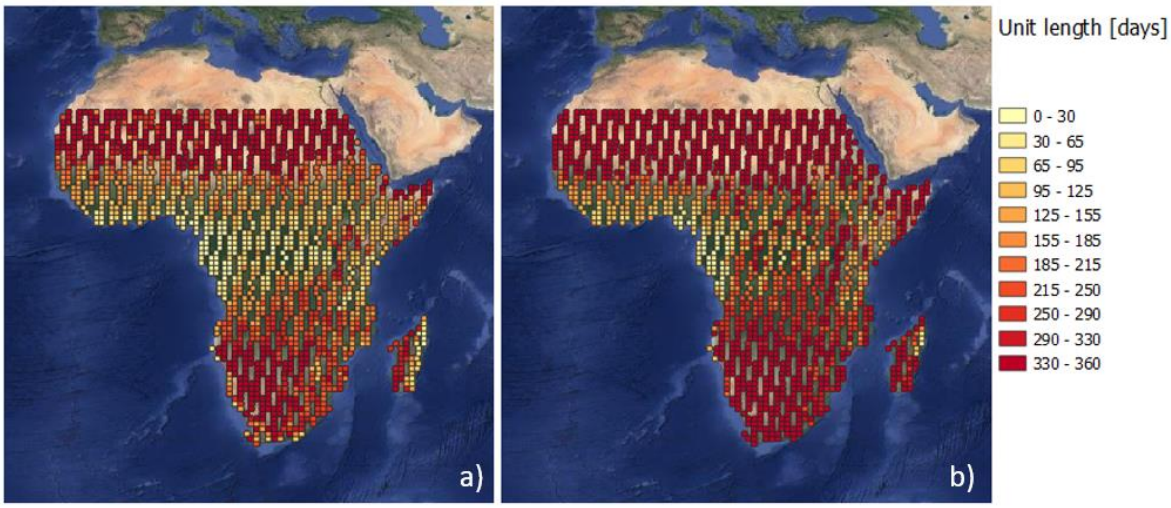

Figure 20: Length [days] of the long validation units over S2 tiles available for sampling, computed considering maximum time step between consecutive scenes (Δt_{max}) of 16 (panel a) and 32 days (panel b).

Results shown in Figure 20a are also shown as histogram plot in Figure 21 where the total number of tiles available for sampling is plotted against series length L [days]. If biome and fire activity stratification is considered, S2 tiles available for sampling are distributed across biomes as a function of the minimum series length L [days] as shown in Figure 22 and Figure 23; series length values are shown on top of each panel.

As observed for global L8 validation units, the number of units available for sampling is greater in the low activity fire stratum and decreases with the increase of minimum length L.

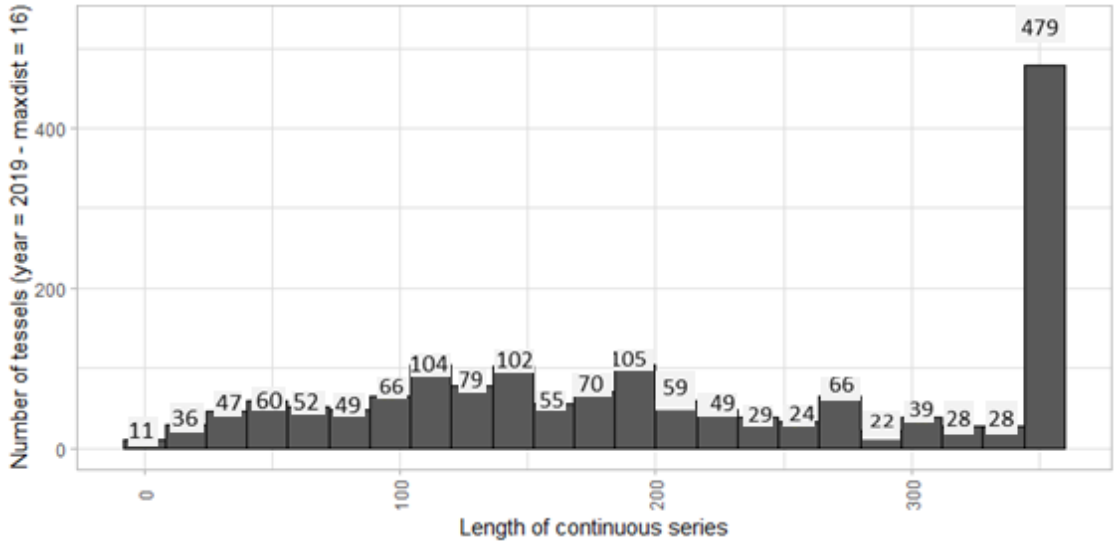

Figure 21: The total number of S2 tiles available for sampling over Africa as a function of the length L of the long unit [days] and computed for consecutive scenes with maximum time step of 16 days.

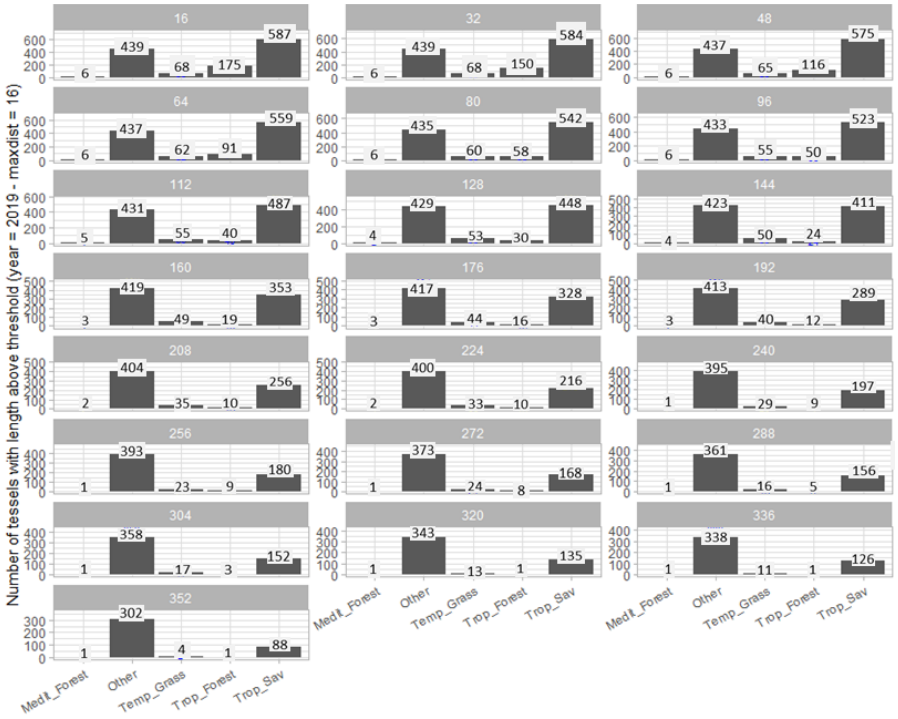

Figure 22: Number of suitable S2 tiles for each biome in the Low Fire Activity class as a function of the minimum length of the of the unit (L) (shown on top of each panel as number of days). Maximum time step between S2 consecutive pairs is set to 16 days and blue values show the number of available S2 tiles when length L is greater than the value shown on the top bar.

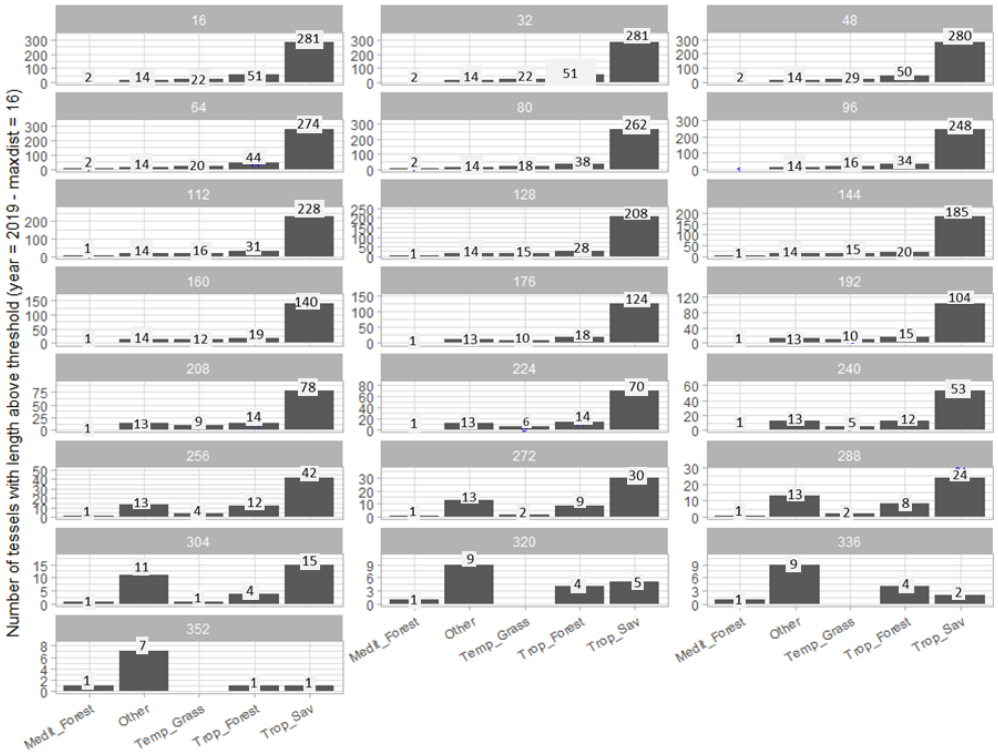

Figure 23: Number of suitable S2 tiles for each biome in the High Fire Activity class as a function of the minimum length of the of the unit (L) (shown on top of each panel as number of days). Maximum time step between S2 consecutive pairs is set to 16 days and blue values show the number of available S2 tiles when length L is greater than the value shown on the top bar.

Further analyses were carried out on S2 data availability highlighting that the frequency of acquisition of S2 A&B satellite missions allows lowering the condition on maximum cloud cover percentage for each scene to (**CC ≤ 15%**) (Figure 24). The percentage of land cloud coverage for each S2 image was extracted from metadata. An additional upper limit for the maximum allowed cumulated cloud cover (**CC_{cum}<30%**) over the series was set to preserve the greatest possible clear area within the sample unit. This value is a trade-off between clear sky time series and time series length: the greater cloud cover is allowed, the longer the time series (since it is combined with maximum allowed time interval) and the smaller the resulting portion of land surface that can be used for extracting reference perimeters of cumulative burned area.

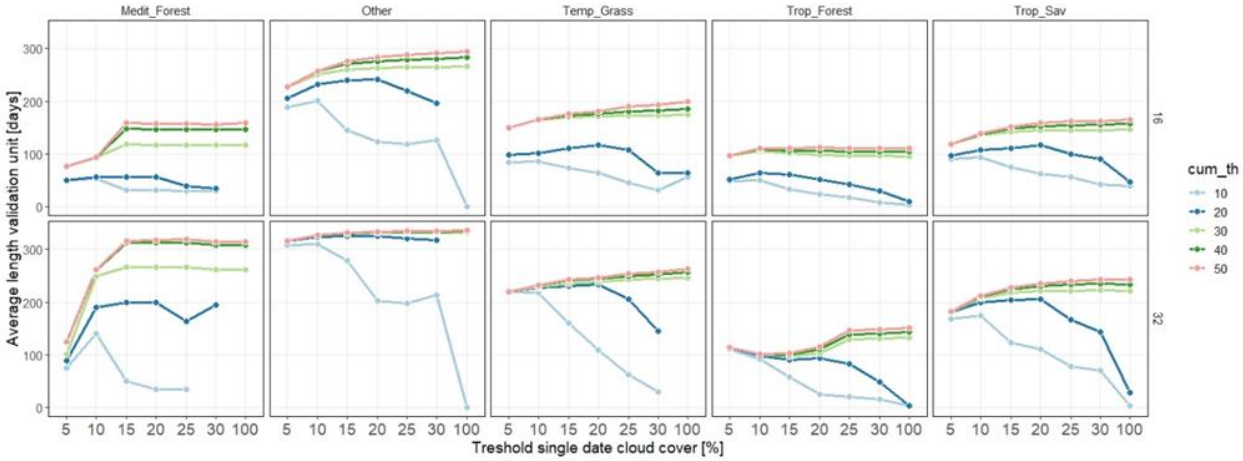

Figure 24. The length of S2 time series [Days, y-axis] over the tiles in Sub-Saharan Africa for the year 2019 as a function of the thresholds set on single scene maximum cloud cover [%, x-axis] and maximum cumulated cloud cover [%, colour line].

2.5.1.3 S2 Sampling cardinality

A total of **50 validation** units/tiles were extracted from the suitable population identified in 2.5.1.1 (CC<15%, CC_{cum}<30%, Δt_{max}=16 days, L_{min}=100 days) and they were distributed among strata based on Eq. 2.

$$n_h \propto N_h \sqrt{\overline{BA}_h} \qquad \text{Eq. 2}$$

where n_h is the number of S2 tiles to be sampled for stratum h, $\overline{BA}_h$ is the average total annual (2019) burned area for stratum h and N_h is the total amount of S2 tiles available for sampling for stratum h. For smaller strata, a minimum of n_h= 2 is assigned. Results obtained with this procedure are summarized in Table 7: the reported number of S2 tiles for each stratum was randomly sampled from the S2 tiles suitable for sampling. The location of the sampled tiles is shown in Figure 25.

Table 7. The number of S2 tiles suitable for sampling in 2019 N for each biome and high/low fire activity stratum (N_{high}, N_{low}) and the number of S2 tiles to be sampled n based on Eq. 1 for each biome and stratum of high/low Fire Activity.

Biomes	N (N_{high} + N_{low})	n (FA_{high} + FA_{low})
Mediterranean Forest	6 (2 + 4)	4 (2 + 2)
Tropical Savanna	646 (192 + 454)	34 (20+14)
Temperate Grassland and Savanna	68 (13 + 55)	4 (2 + 2)
Tropical Forest	65 (29 + 36)	4 (2 + 2)
Others	446 (14 + 432)	4 (2 + 2)
Total	1231 (250 + 981)	50 (28+22)

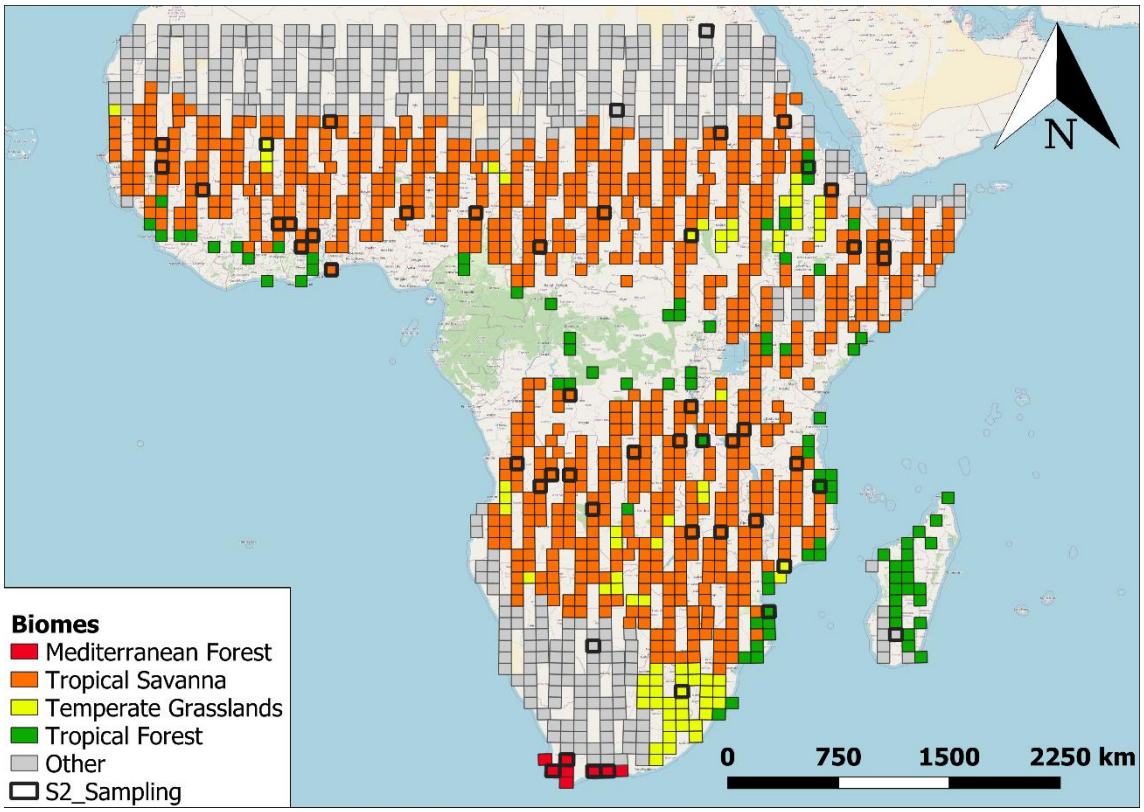

Figure 25: Spatial distribution of the 50 S2 tiles sampled randomly over Africa for each stratum (biome/fire insentity) highlighted in black and overlaid on the S2 tiles suitable for sampling as extracted from the S2 archive by setting Δtmax=16 days and Lmin=100 days.

2.5.2 Extraction of S2 fire reference perimeters

Fire perimeters for S2 validation units were extracted for short units (consecutive S2 images) to map areas burned between the two dates (t1, t2); processing steps are the same as L8 and the flowchart is shown in Figure 3. All short units over the same area are combined to derive fire perimeters over the S2 long unit with a script coded in Arcpy.

S2 short unit classification was implemented in GEE, which applies a RF algorithm to identify burned polygons; the script is coded to perimeter existing burned areas between two consecutive images. Details on script input/output are provided in Section 2.4.2. The output layers consist of:

- Burned areas in vector KML format;

- The validation region of 100 km x 100 km obtained by a buffering of the centroid of the S2 frames;
- Cloud Mask in vector format;
- Training polygons as vector shapefiles;

The classification of the S2 short units is then converted to shapefile format and processed to extract fire reference perimeters, clouds and burn date. An example S2 tile is given in Figure 26 and Figure 27. Figure 26 shows a time series of cloud-free S2 images with maximum time step between consecutive dates of 16 days for S2 tile 35MLN (Democratic Republic of Congo, DRC). The corresponding BA product covering the entire validation period is shown in Figure 27.

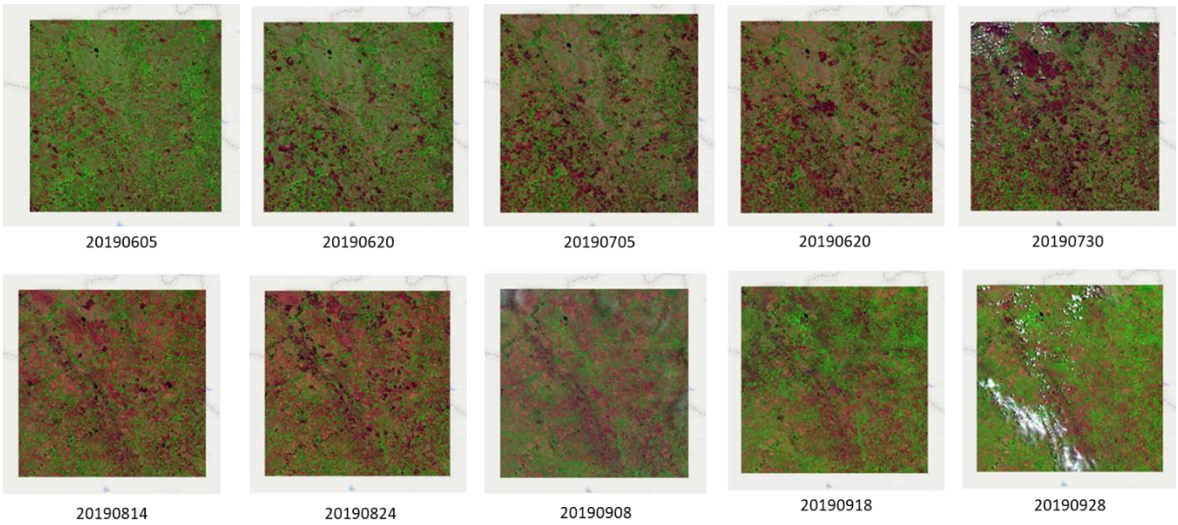

Figure 26: Time series of cloud-free S2 images for S2 tile 35 MLN (DRC) and displayed as RGB false colour composites (SWIR-NIR-Red).

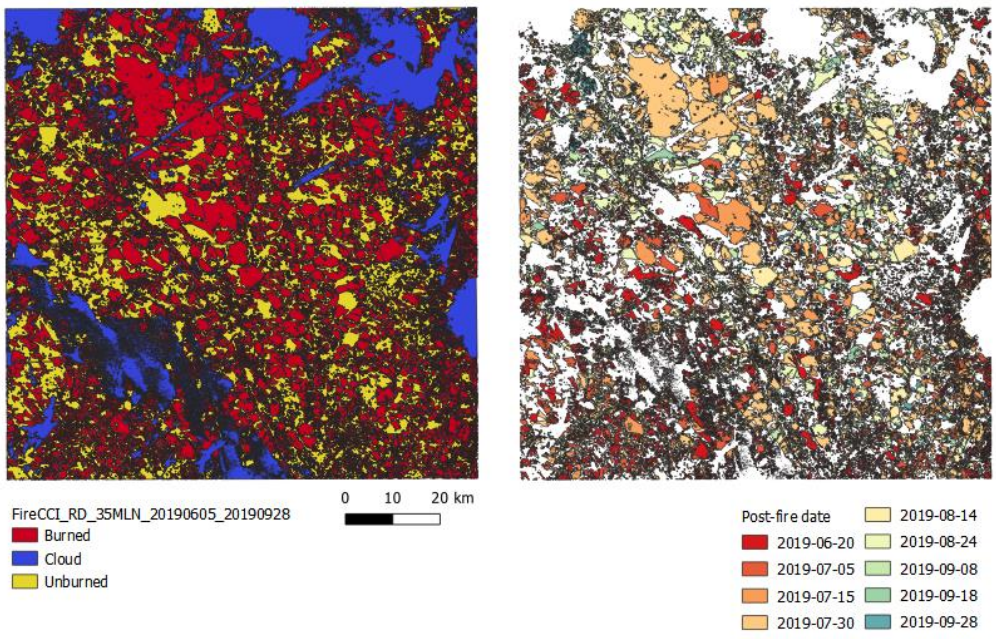

Figure 27: BA reference perimeters for S2 tile 35 MLN (DRC, Africa) as obtained from time series of cloud-free S2 images (Figure 26) and a RF algorithm: left panel: synthetic final BA map showing over the period June 05 to September 28, 2019, burned polygons (red), clouds (blue) and unburned polygons (yellow); right panel: BA polygons displayed with post-fire date attribute.

The reference files generated for this article have been added to the Burned Area Reference Database (BARD) (Franquesa et al., 2020) and are freely available at https://doi.org/10.21950/VKFLCH (Stroppiana et al. 2022a).

2.5.3 Computation of regional-Africa BA accuracy metrics

2.5.3.1 Evaluation of reference perimeters by comparison with Planet data

The accuracy of the fire perimeters obtained from S2 was assessed by comparison with perimeters obtained from very High Resolution (VHR) multi-spectral Planetscope images downloaded for this purpose. Over five sites shown by the red squared symbols in Figure 28), cloud-free time series of Planet images covering about 2/3 months for the year 2019 were downloaded and processed to extract fire perimeters. Planet time series were processed similarly to S2 images with a supervised RF algorithm applied to each image pair within each site with training polygons collected over burned and unburned areas.

To quantify the agreement, we computed the confusion matrix for each site and accuracy metrics for the Burned Area (BA) category: omission (Oe) and commission error (Ce), Dice Coefficient (DC) and the Relative Bias (relB). The results (not shown here) are available in Table 7 of Stroppiana et al. (2022b).

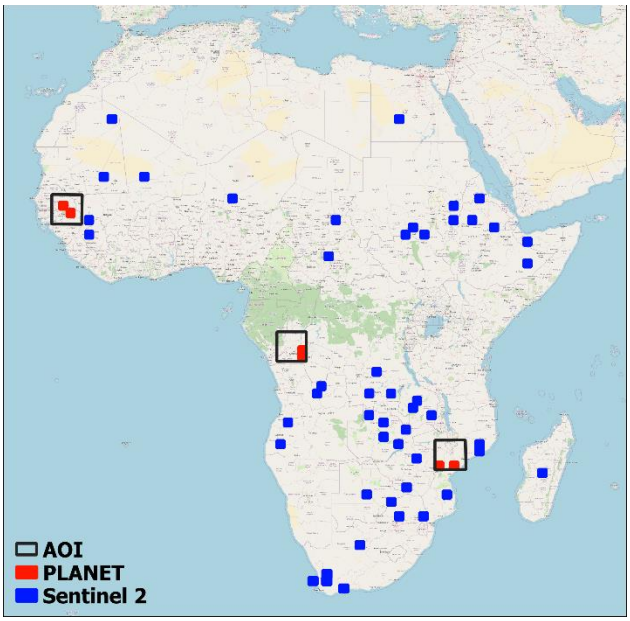

Figure 28: S2 tiles sampled for extracting fire perimeters from S2 times series (n=50, blue square polygons) and the S2 tiles used for comparison with Planet fire perimeters (red square polygons).

The comparison of S2 fire perimeters with Planet reference showed a general good and satisfying agreement between the two datasets, considering that products from Plant have a very high resolution and can be assumed as “ground truth” across the investigated period. Figure 29 shows fire polygons extracted from Planet over two example sites in Northern Africa and Central Africa, and the agreement maps (bottom row). In the figure, Planet perimeters are shown with different colours to highlight DOY of detection within the time series of Planet images for each site; the agreement maps show spatial distribution of the accurately mapped burned/unburned polygons (green), commission (blue) and omission (red) errors.

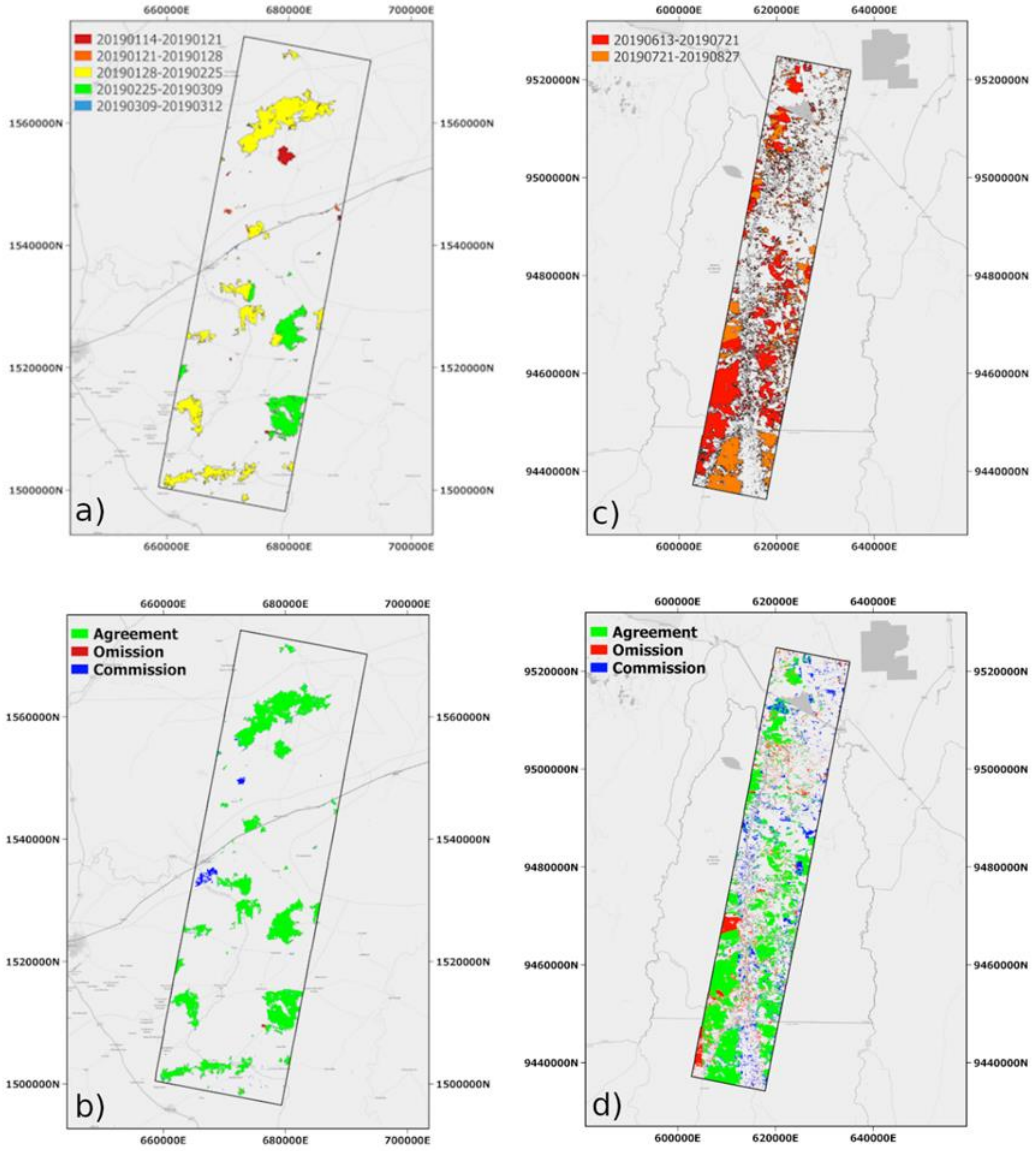

Figure 29: Planet fire perimeters (top) and agreement maps (bottom) for two sites: Northern Africa (a, b) and Central Africa (c, d).

2.5.3.2 Accuracy assessment for the FireCCISFD20 BA product

The comparison between the FireCCISFD20 BA product and S2 fire reference perimeters over the 50 sample tiles generated the global confusion matrix shown in Table 8, where agreement over burned areas is given by the True Positive (TP) entry expressed as km2. In the table, overall metrics are summarized with metrics estimated per biome; notice that with the exception of "Tropical savanna", the other biomes are represented by four S2 tiles sampled in the "High" (two tiles) and "Low" (two tiles) fire activity strata.

Table 8. The overall confusion matrix derived from the comparison of fire reference perimeters and the FireCCISFD20 BA product over the 50 tiles sampled for validation. Areas are given in square kilometres [km^2] while error metrics are given in percentage [%].

Biome	TP [km^2]	FP [km^2]	FN [km^2]	TN [km^2]	Oe [%]	Ce [%]	DC [%]	relB [%]
All	80958.0	15040.77	7572.63	316170.19	8.5	15.0	87.7	8.4
Mediterranean Forest	1.31	2.28	39.43	32945.25	96.8	63.5	5.9	-91.2
Tropical Savanna	69516.72	12176.14	5675.33	198937.18	7.5	14.9	88.6	8.6
Temperate Grassland and Savanna	5662.52	1229,90	298.45	24745.43	5.0	17.8	88.1	15.6
Tropical Forest	5776.22	1629.76	1557,41	25741.43	21.2	22.0	78.4	0.99
Others	1.22	2.69	2.02	33800.91	62.2	68.7	34.2	20.6

Overall, the FireCCISFD20 BA product tends to overestimate BA (Ce=15%, RelB=8.4%), and with the exception of "Mediterranean" tiles, commission errors are greater than omission (relB>0). Savanna and Grassland biomes show the greatest agreement (DC~88%). Lowest metric estimates are observed in the "Mediterranean" and "Other" biomes where very few fires occur and the amount of burned areas is very low. Globally (all sample units together), the FireCCISFD20 BA product is characterized by commission and omission errors of 15% and 8.5%, respectively, and a Dice Coefficient greater than 87%.

2.6 Validation of the FireCCIS1S2AF10 BA product over test sites

In this section, the data (Planet) and method for extracting fire reference perimeters for the validation of the FireCCIS1S2AF10 BA product are described. Results of the validation are also presented.

2.6.1 Planet dataset

Over the three study sites for testing the SAR-O algorithm that exploits the integration of S1 and S2 for high resolution burn mapping (Figure 30), no stratified sampling design is applicable (which means a Stage 1 validation). Hence, within each test area, validation units were selected to cover different fire, land cover and cloud cover conditions. The availability of VHR Planet data was inspected for spot/local evaluation of the FireCCIS1S2AF10 BA products and Planet scenes were downloaded over a set of validation units within the three sites shown in Figure 30. Cloud free images were selected to cover a period of about two/three months over a set of five sites shown in the Figure.

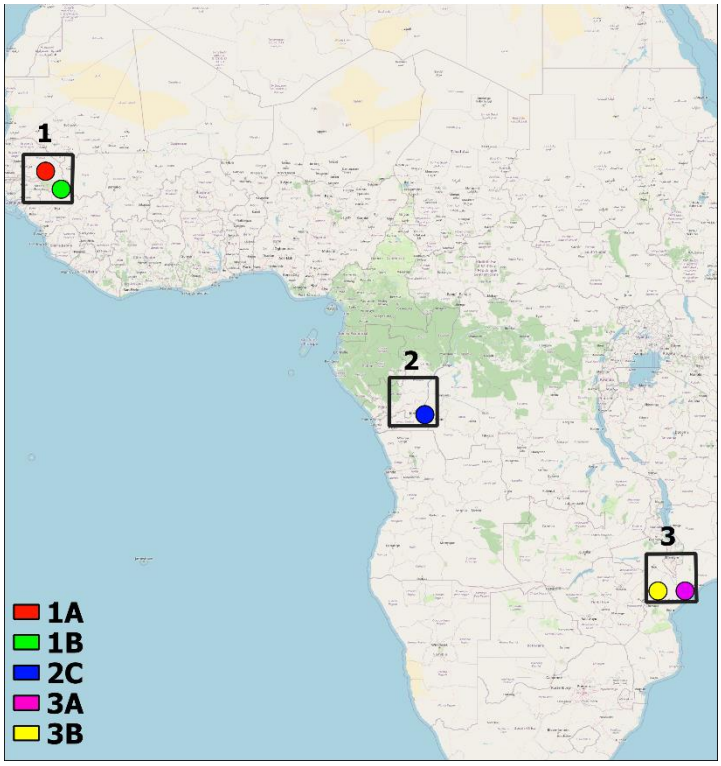

Figure 30: Location of the sites of interest for SAR-O algorithm deployment: Site 1 Northern Africa, Site 2 Central Africa and Site 3 Southern Africa and the sites where Planet time series were downloaded and processed to extract fire reference perimeters.

Table 9 summarizes Planet images downloaded over sites 1 to 3 (Figure 30). For each validation site, smaller regions were identified that were covered by approximately ten Planet scenes. The multi-temporal dataset was selected from the most cloud free dates. Site 2 (Central Africa) is characterized by the least images available due to persistent cloud cover.

Table 9. Planet images downloaded and processed to extract fire reference perimeters: location (country), first and last image of the time series (Period), number of images in the time series (N dates), length of the time series from the first to the last image (N days) and the S-2 tile covered by the Planet sites.

Site	Location	Period	N dates	N days	S2 tile
1A	Tambacounda (Senegal)	14/01/2019 - 12/03/2019	6	57	28PFA
1B	Niokolo_Koba (Senegal)	17/01/2019 - 20/03/2019	4	62	28PGV
2C	Kinshasa, DRC	13/06/2019 - 27/08/2019	3	75	33MXQ
3A	Capenga (Mozambico)	30/07/2019 - 26/10/2019	5	88	36KYE
3B	KWE Chimoio (Mozambique)	23/07/2019 - 14/09/2019	3	53	36KWE

Figure 31 shows along track mosaics of Planet images. In particular, panel a shows the two sites A (Tambacounda) and B (Niokolo_Koba) in Senegal while panel b shows the site A (Capenga) in Mozambique. Figure 32 presents an example of the mosaic images derived from single Planet scenes over Site 1B; images are shown as RGB false colour composites (NIR-Red-Green). Progress of burned surfaces can be observed as the fire season proceeds from January to March.

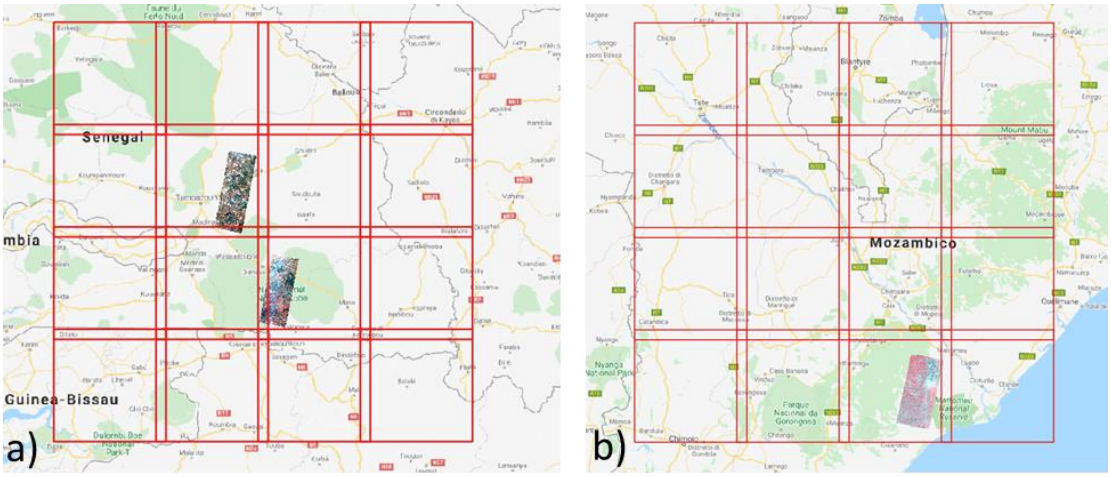

Figure 31: Location of sites 1A, 1B and 3A within the validation sites in Northern Africa (a) and southern Africa(c).

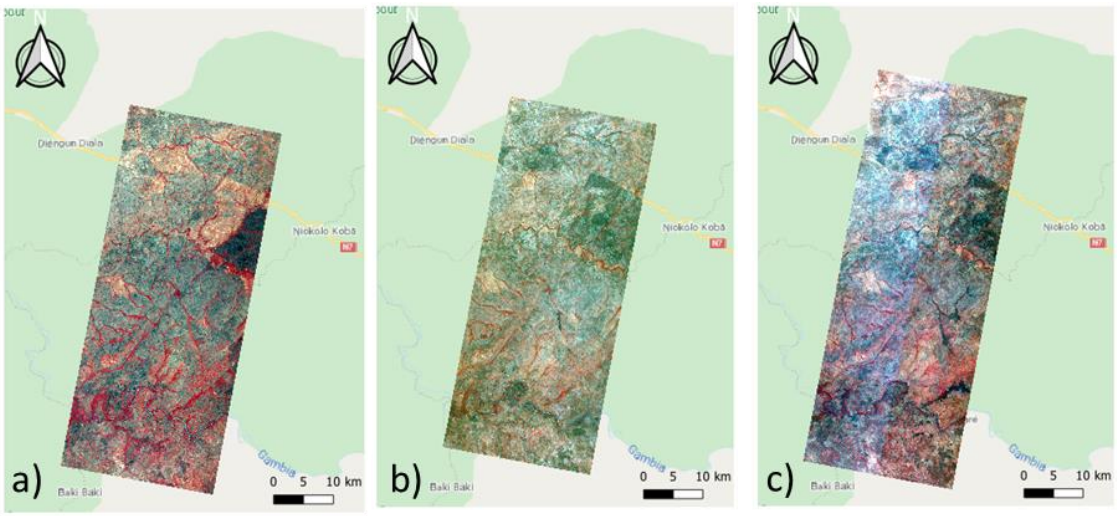

Figure 32: Example of Planet image mosaics displayed as false colour composites (RGB; NIR, Red, Green) over site 1B for the following dates: 30/01/2019 (a), 06/03/2019 (b) and 20/03/2019 (c).

2.6.2 Extraction of Planet fire reference perimeters

Preliminary tests were carried out for defining the best approach for the extraction of fire reference perimeters from Planet images. Finally, a supervised Random Forest algorithm was applied to Planet time series to identify areas burned between two consecutive dates (pre- and post-fire dates); the algorithm is applied at pixel level. Training data over burned and unburned surfaces were extracted for image pair(s) by photointerpretation of RGB false colour composites. RF classification output raster are converted to ESRI© shapefile and cleaning of small polygons was carried out to reduce the impact of isolated pixels, geo-location errors and different spatial resolutions. Shapefiles of consecutive Planet image pairs are summed up over the time period from the first to the last date of the time series to build the fire reference perimeters for each site and they are used as input for the cross-tabulation with the FireCCIS1S2AF10 BA product to be validated. Figure 33 shows an example of fire reference perimeters for sites 1A (left) and 1B (right) where colours

represent the so called short-units (pairs of consecutive Planet images that were classified with RF algorithm).

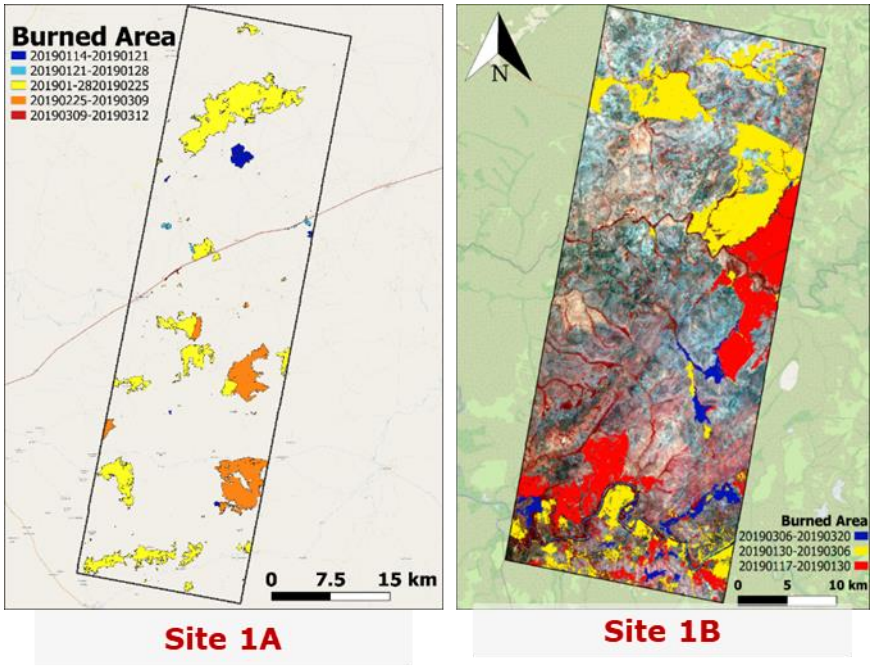

Figure 33: Fire reference perimeters for site 1A (left) and 1B (right): colours represent the consecutive image pairs that were classified to obtain the reference perimeters.

The RF algorithm is implemented in R open source software (https://www.r-project.org/, accessed on April 2022). The algorithm was applied iteratively by adjusting the training areas over burned and unburned areas to derive the best classification (supervised classification). The output raster is characterized by a spatial resolution of 4 meters, and represents the areas burned between two dates. An example of the input image and output classification (burned/unburned) is shown in Figure 34.

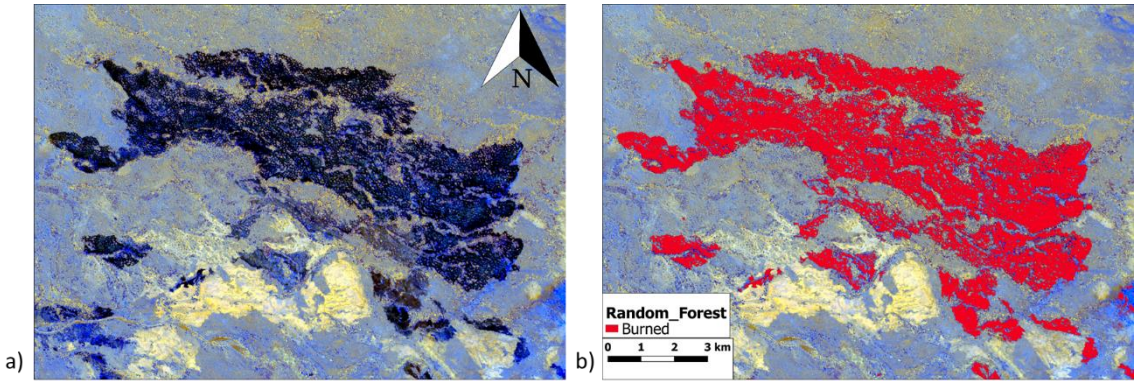

Figure 34: a) Detail of the temporal difference image (RGB: R=NDVI, G=NIR reflectance, B=Red reflectance) between post- and pre-fire dates ($\Delta_{post-pre}$) and b) example of burned area map.

Figure 35 shows example BA maps extracted from Planet mosaic images for site 1B (Niokolo_Koba, Senegal) covering the time period January 30th to March 1st, 2019. Since burned areas are derived for temporal difference images, BAs refer to the post-fire date for each pair of mosaic images: 17/01/2019-30/01/2019 (red) and 30/01/2019-06/03/2019 (orange).

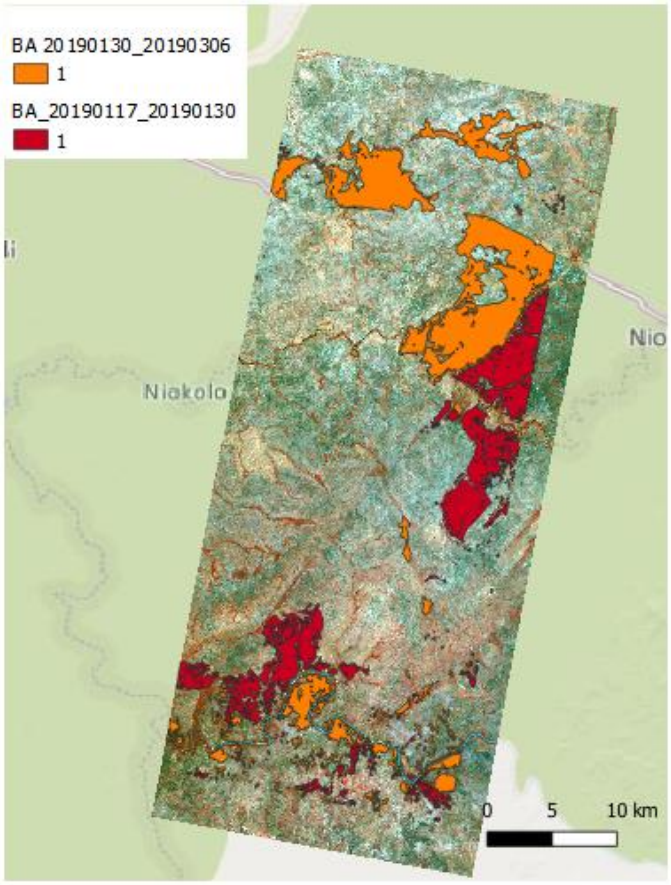

Figure 35: Example classification of burned areas derived from a multi-temporal Planet mosaic image for site 1B (Niokolo_Koba). On the background is shown an RGB Planet image mosaic obtained for 06/03/2019.

2.6.3 Accuracy of the FireCCIS1S2AF10 BA product over test sites

The FireCCIS1S2AF10 BA product is composed by four distinct products:

1. FireCCIS1S2AF10_SAR from Sentinel-1 backscatter (*S1*)
2. FireCCIS1S2AF10_MSI from Sentinel-2 reflectance (*S2*)
3. FireCCIS1S2AF10_SAR_MSI from combined Sentinel-1 and Sentinel-2 (i.e. sensor integration at algorithm input level) (*S1_S2*)
4. FireCCIS1S2AF10_SAR_MSI_Opt by merging two independent BA products (i.e. integration at algorithm output level) (*S1_S2_Optimum*)

Figure 36 shows the comparison between the FireCCIS1S2AF10 BA sub-products and fire reference perimeters extracted from processing Planet time series over site 1A. Greatest accuracy is achieved by the *S2* and *S1_S2_Optimum* products with Dice Coefficient above 80% (81.80% and 81.37%, respectively).

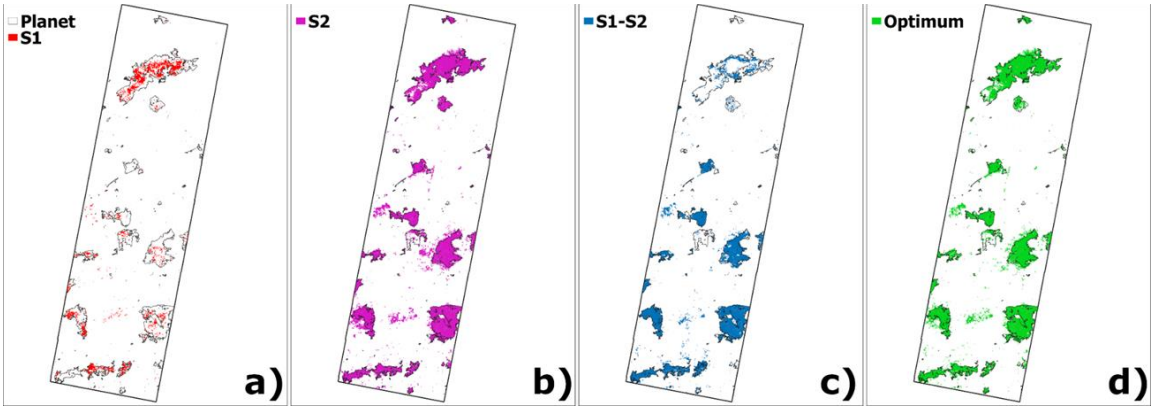

Figure 36. Area classified as burned in the FireCCIS1S2AF10 BA products in the period covered by the Planet time series over site 1A. The area burned is extracted by selecting DOYs in 14/01/2019 (DOY=14) to 12/03/2019 (DOY=71) coincident with dates of acquisition of Planet images.

Figure 37 shows the accuracy metrics computed for the five sites of the FireCCIS1S2AF10 BA products validation. Metrics are computed separately for the sub products. Site 1A is the one where accuracy is greatest for the sub-products that, with different algorithms, exploit information from the multispectral S2 data. As expected, accuracy of BA product derived only from S1 data is the lowest; this happens for all sites although in site 2C the difference it is reduced probably due to the high cloud cover of tropical regions, which limits the availability of S2 data.

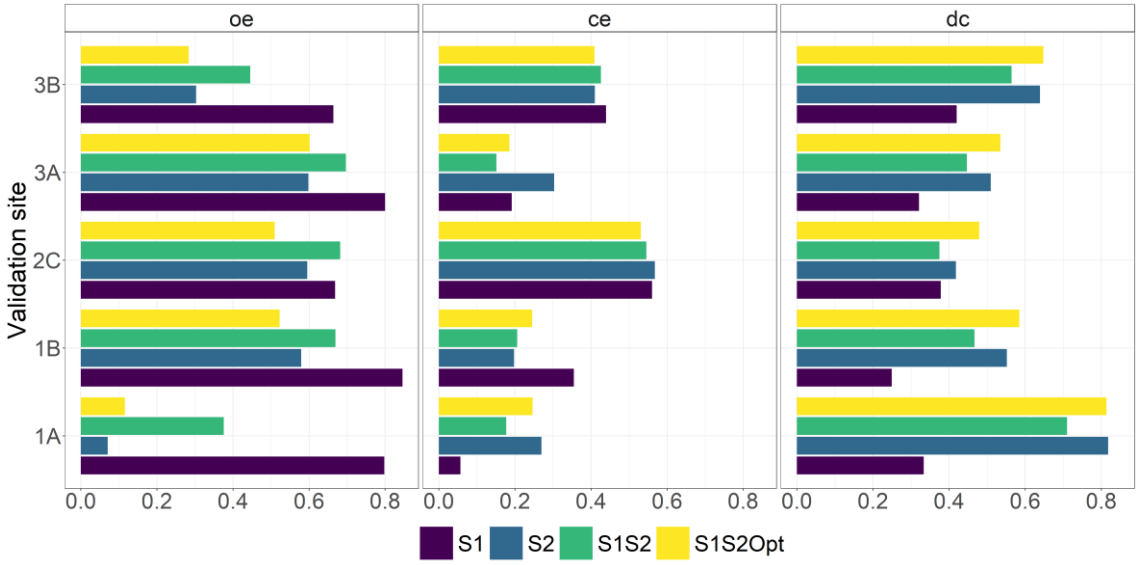

Figure 37. Accuracy metrics for the FireCCIS1S2AF10 BA products over Planet sites.

3 Intercomparison of global BA products

This section presents a comparison between four global burned area products for the year 2019: FireCCI51, FireCCIS310, MCD64A1 collection 6, and C3SBA11.

3.1 Burned area detection of existing products

Figure 38 shows the temporal (yearly) trends of the different current FireCCI products: FireCCILT11, FireCCI51, FireCCIS310, FireCCISFD11, FireCCISFD20, as well as some additional current and/or well-known global BA products: C3SBA11 (Lizundia-

Loiola et al. 2021), NASA's MCD64 Collection 6 (Giglio et al. 2018) and GFED4s (https://www.globalfiredata.org/data.html, accessed on April 2022). The characteristics of each of these products are summarized in Table 10.

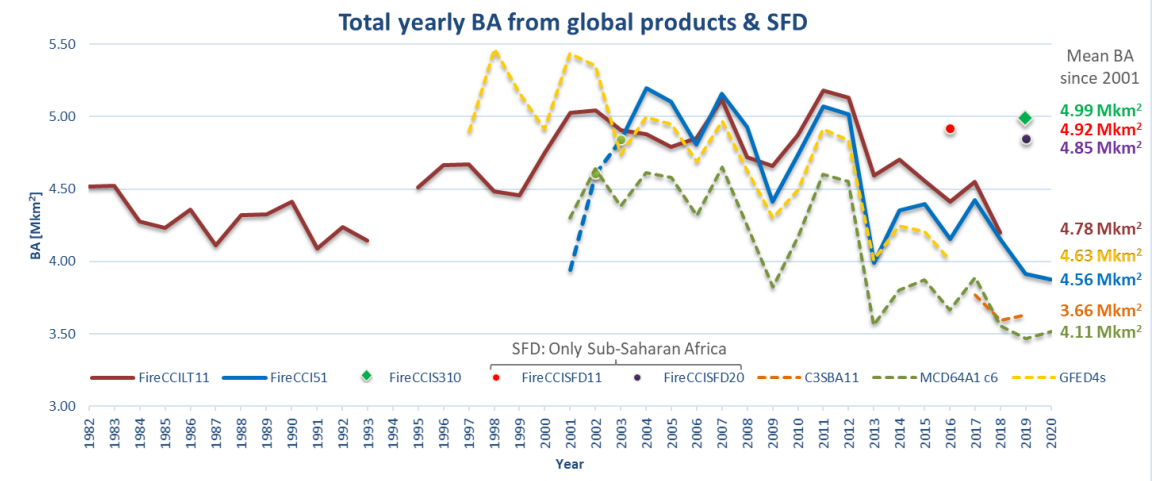

Figure 38: Annual burned area of different global and regional products

Table 10: Summary of BA products characteristics corresponding to the products shown in Figure 38

BA product	Input Data Sensor		Coverage	Spatial resolution		Time Series	Reference
	Surface Reflectance	Active Fires		Pixel	Grid		
FireCCILT11	AVHRR-LTDR	-	Global	0.05 deg.	0.25 deg	1982-2018 (gap 1994)	Otón et al. (2021)
FireCCI51	MODIS	MODIS	Global	250 m	0.25 deg	2001-2020	Lizundia-Loiola et al. (2020)
FireCCIS310	Sentinel-3 SYN	VIIRS	Global	300 m	0.25 deg	2019	Lizundia-Loiola et al. (2022)
FireCCISFD11	Sentinel-2	MODIS	Sub-Saharan Africa	20 m	0.25 deg	2016	Roteta et al. (2019)
FireCCISFD20	Sentinel-2	VIIRS	Sub-Saharan Africa	20 m	0.05 deg	2019	Chuvieco et al. (2022)
C3SBA11	Sentinel-3 OLCI	MODIS	Global	300 m	0.25 deg	2017-Sep. 2020	Lizundia-Loiola et al. (2021)
MCD64A1 c6	MODIS	MODIS	Global	500 m	-	Nov. 2000-Mar. 2022	Giglio et al. (2018)
GFED4s	MODIS	MODIS	Global	0.25 deg.	-	1997-2016	Giglio et al. (2013)

The SFD products detect more burned area for Sub-Saharan Africa than most global products detect for the whole world, with the exception of FireCCIS310, which detects 4.99 Mkm^2 globally. Temporal trends showed an overall good global agreement between FireCCI51, MCD64A1 c6 and GFED4 (all based on MODIS information), particularly starting in 2003 when both MODIS Terra and Aqua data were available.

Considering only the four global medium-resolution products (250 to 500 m), the total BA is highest for FireCCIS310, as stated above, followed by FireCCI51 (3,8 Mkm^2), C3SBA11 (3,6 Mkm^2), and MCD64 c6 (3,4 Mkm^2). FireCCIS310 detects 28% more BA compared to the next highest BA product, FireCCI51.

Regarding the spatial patterns, the stratification of the world in seven major biomes show, as expected, that the biomes that burn the most are the tropical savannas, followed by the tropical forests (Figure 39). While the different products analysed show a similar BA in the biomes that have few fires (i.e. Tundra and Mediterranean), some differences can be observed in the rest of the biomes. The most significant difference is located in the Tropical Forests, where FireCCIS310 detects 116% more burned area (897,703 km^2) than the mean of the other three products (415,194 km^2), which have similar values. In the case of the Tropical Savanna, an increasing trend is found starting with MCD64 c6 (2.5 Mkm^2), followed by C3SBA11 (2.7 Mkm^2) and FireCCI51 (3.0 Mkm^2), with the highest value found for FireCCIS310 (3.3 Mkm^2). FireCCIS310 has the highest BA in all biomes except Tundra and Boreal Forests, where C3SBA11 reach higher values.

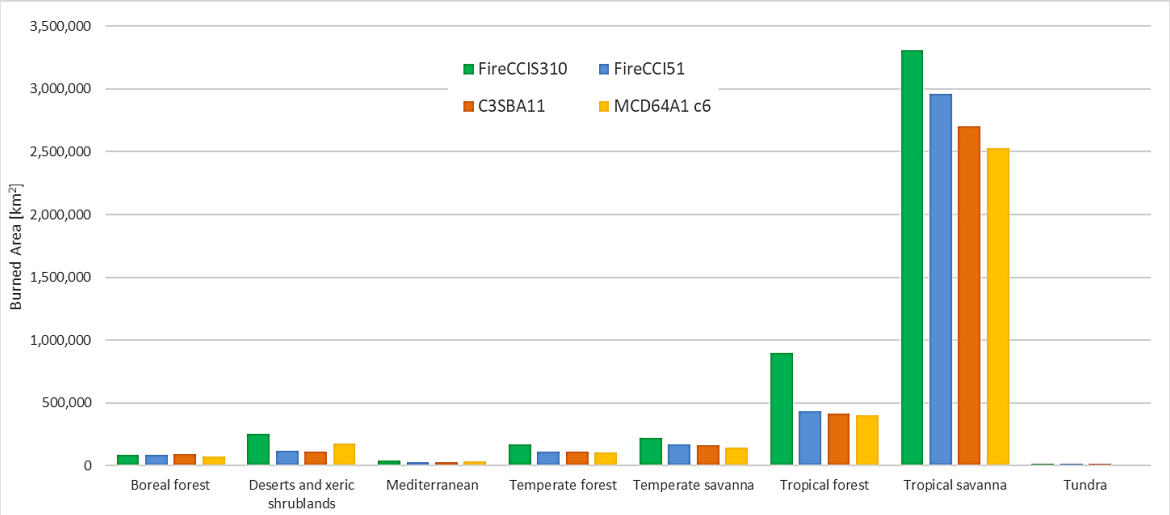

Figure 39: Burned area for the year 2019 stratified per biome in different global products.

3.2 Spatial accuracy assessment

Table 11 presents a comparison between the four global burned area products for the year 2019. The validation results have been extracted from Franquesa et al. (2022) and Lizundia-Loiola et al. (2022). FireCCIS310 has the highest Dice Coefficient (68.1%) and the lowest absolute value of relB (-27.2%). The negative value of relB for all BA products indicate that the datasets underestimate the per-class BA. This product also has the lowest omission error from the different products (41.2%), while MCD64A1 c6 has the lowest commission error (17.5%).

Table 11: Global error estimates [%] for the four products and year 2019, with the standard error in parenthesis. Data for FireCCI51, C3SBA11 and MCD64A1 c6 were extracted from Table 5 of Franquesa et al. (2022), while data for FireCCIS310 were extracted Table 1 of Lizundia-Loiola et al. (2022). In bold the product that shows the highest accuracy in each specific metric.

	FireCCI51	FireCCIS310	C3SBA11	MCD64A1 c6
DC	63.9 (2.8)	**68.1 (2.5)**	61.7 (2.9)	59.8 (3.2)
Ce	20.8 (1.7)	19.2 (1.7)	20.8 (1.7)	**17.5 (1.4)**
Oe	46.5 (3.4)	**41.2 (3.0)**	50.3 (3.4)	53.1 (3.6)
relB	-32.5 (3.4)	**-27.2 (2.7)**	-39.0 (3.59)	-43.1 (3.8)

The accuracy metrics at biome level are shown in Table 12. The highest DC values between the different products are systematically found for FireCCIS310, with the highest values regarding biomes are found in those biomes with low proportion of BA (Mediterranean, Boreal Forests and Tundra). The lowest DC was found in the Tropical Forests. relB values are mostly negative, indicating an underestimation of BA, except in the case of the Boreal Forests, where most relB values are positive. And in the case of the Mediterranean biome, FireCCIS310 has a value close to 0, which shows that the commission and omission errors are well balanced. The highest Ce corresponds to the Mediterranean biome for the C3SBA11 (28.8%), while the lowest belongs to the Tundra biome in FireCCIS310. Regarding omission errors, the highest is found in the MCD64 c6 product in Tropical Forests (70.9%), while the lowest belongs to the Boreal Forests of FireCCIS310.

Table 12: Estimated error metrics [%] at biome level for each product and year 2019, with the standard error in parenthesis. The total BA mapped in the reference data (BAref) is provided per biome in km^2. Data for FireCCI51, and MCD64A1 c6 was extracted from Table 6 of Franquesa et al. (2022), while data for FireCCIS310 were extracted Table 2 of Lizundia-Loiola et al. (2022). In bold the product that shows the highest accuracy in each specific metric.

Biome		FireCCI51	FireCCIS310	C3SBA11	MCD64A1 c6
Boreal forest	*DC*	79.0 (1.6)	**82.2 (1.2)**	76.7 (1.7)	75.7 (4.5)
	Ce	23.2 (3.3)	21.4 (4.0)	24.2 (6.1)	**20.4 (2.2)**
	Oe	18.8 (1.9)	**14.0 (2.5)**	22.3 (5.5)	27.8 (6.9)
	relB	5.7 (5.9)	9.4 (8.5)	**2.5 (14.9)**	-9.3 (7.1)
	BAref		1840.8		
Deserts & xeric shrublands	*DC*	52.5 (7.7)	**76.7 (2.6)**	52.8 (7.7)	75.5 (2.5)
	Ce	20.1 (0.9)	**15.1 (0.8)**	15.5 (1.9)	19.1 (1.5)
	Oe	60.9 (8.4)	30.1 (4.8)	61.6 (8.0)	**29.2 (4.6)**
	relB	-51.0 (10.1)	-17.6 (6.3)	-54.6 (9.0)	**-12.4 (6.3)**
	BAref		439.0		
Mediterranean	*DC*	69.5 (6.2)	**83.4 (4.7)**	57.3 (2.7)	74.5 (6.4)
	Ce	23.7 (6.9)	17.0 (1.8)	28.8 (4.4)	**13.2 (1.7)**
	Oe	36.2 (5.6)	**16.2 (7.6)**	52.1 (1.8)	34.7 (8.8)
	relB	-16.3 (0.7)	**1.0 (7.2)**	-32.7 (1.6)	-24.8 (8.8)
	BAref		283.6		
Temperate forest	*DC*	59.6 (9.8)	**67.5 (10.4)**	62.6 (10.0)	51.5 (10.1)
	Ce	22.3 (10.4)	21.5 (12.6)	**17.0 (7.2)**	19.6 (8.8)
	Oe	51.7 (9.0)	**40.9 (9.0)**	49.7 /10.3)	62.1 (9.0)
	relB	-37.8 (4.7)	**-24.7 (2.9)**	-39.4 (7.4)	-52.9 (6.3)
	BAref		889.1		
Temperate savanna	*DC*	62.0 (1.9)	**67.6 (2.0)**	64.2 (1.9)	55.0 (1.9)
	Ce	19.6 (1.6)	**17.9 (0.9)**	19.2 (1.0)	21.9 (2.1)
	Oe	49.6 (2.4)	**42.6 (2.6)**	46.8 (2.5)	57.6 (2.3)
	relB	-37.3 (3.0)	**-30.1 (2.8)**	-39.4 (7.4)	-45.7 (3.4)
	BAref		3403.1		
Tropical forest	*DC*	50.7 (8.0)	**57.8 (6.6)**	46.3 (8.2)	42.6 (8.4)
	Ce	24.1 (6.1)	27.6 (5.4)	24.1 (6.7)	**20.6 (4.7)**
	Oe	61.9 (7.7)	**51.9 (7.1)**	66.7 (7.4)	70.9 (7.3)
	relB	-49.8 (7.3)	**-33.6 (6.3)**	-56.1 (7.1)	-63.4 (7.6)

Biome		FireCCI51	FireCCIS310	C3SBA11	MCD64A1 c6
	BAref		19204.3		
Tropical savanna	*DC*	68.9 (2.9)	**72.0 (2.5)**	67.0 (3.1)	66.3 (3.4)
	Ce	19.4 (1.7)	**15.5 (1.3)**	16.3 (1.5)	16.0 (1.5)
	Oe	39.9 (3.8)	**37.2 (3.2)**	44.1 (3.9)	45.2 (4.2)
	relB	**-25.4 (4.0)**	-25.7 (3.1)	-33.2 (4.1)	-34.8 (4.5)
	BAref		104793.3		
Tundra	*DC*	69.9 (9.3)	**81.8 (0.9)**	78.8 (3.1)	50.9 (10.7)
	Ce	26.9 (12.7)	**11.8 (5.7)**	18.9 (1.9)	26.0 (15.9)
	Oe	33.1 (6.4)	23.7 (3.1)	**23.4 (7.5)**	61.2 (9.4)
	relB	-8.6 (7.1)	-13.4 (9.0)	**-5.5 (11.5)**	-47.6 (10.8)
	BAref		564.9		

3.3 Temporal reporting accuracy

The most accurate product in terms of temporal reporting accuracy compared to VIIRS active fires (Schroeder and Giglio 2018) was the MCD64A1 c6, followed by the FireCCIS310 (Figure 40). The differences were accounted in absolute terms, e.g. either if the difference is -1 (the product detected the fire one day before the active fire) or +1 (the product detected the fire one day after the active fire) the difference is considered to fall in the 0-1 days frame.

MCD64A1 c6 detected 65.9% of the cases within 0-1 day difference, while FireCCIS310 was able to detect 55.8% of the cases. FireCCI51 and C3SBA10 showed a clear trend to detect burned area after the fire, more specifically within the first 10 days after the active fire. The four products show more similar accuracies within 0-10 day difference, with the highest value corresponding to MCD64A1 c6 (97.7%) and the lowest to FireCCI51 (89.3%).

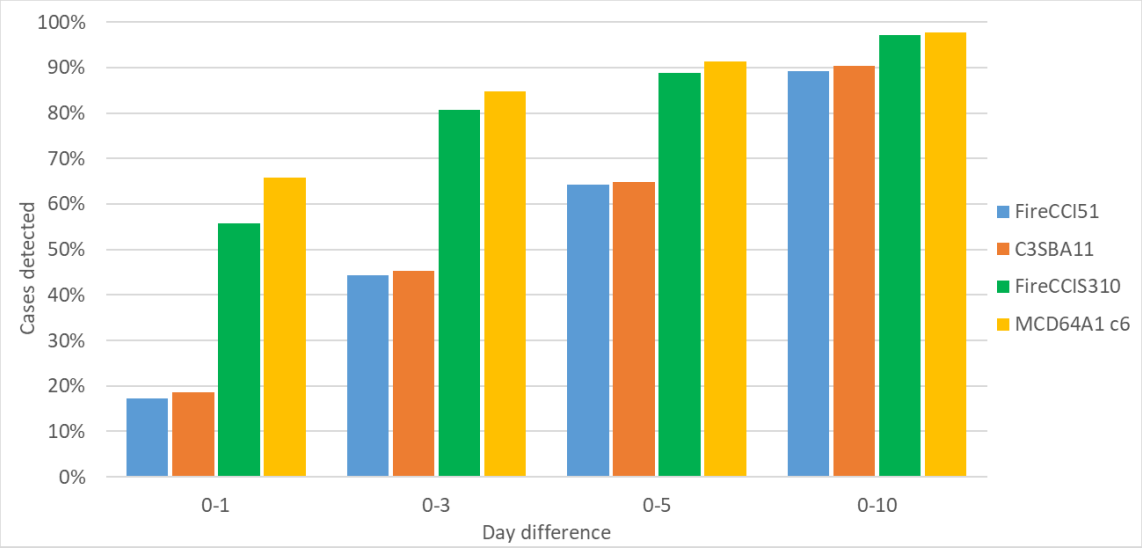

Figure 40: Global temporal reporting accuracy for each product compared to VIIRS active fires (VNP14IMGML) for the year 2019.

Figure 41 show the per-biome temporal reporting accuracy estimates for the different products, extracted from Lizundia-Loiola et al. (2022). A similar pattern as in the global analysis was observed at biome level, where MCD64A1 c6 was the most accurate product

in most of the biomes. FireCCIS310 clearly outperformed the temporal reporting accuracy of the previous FireCCI51 product, as well as C3SBA11.

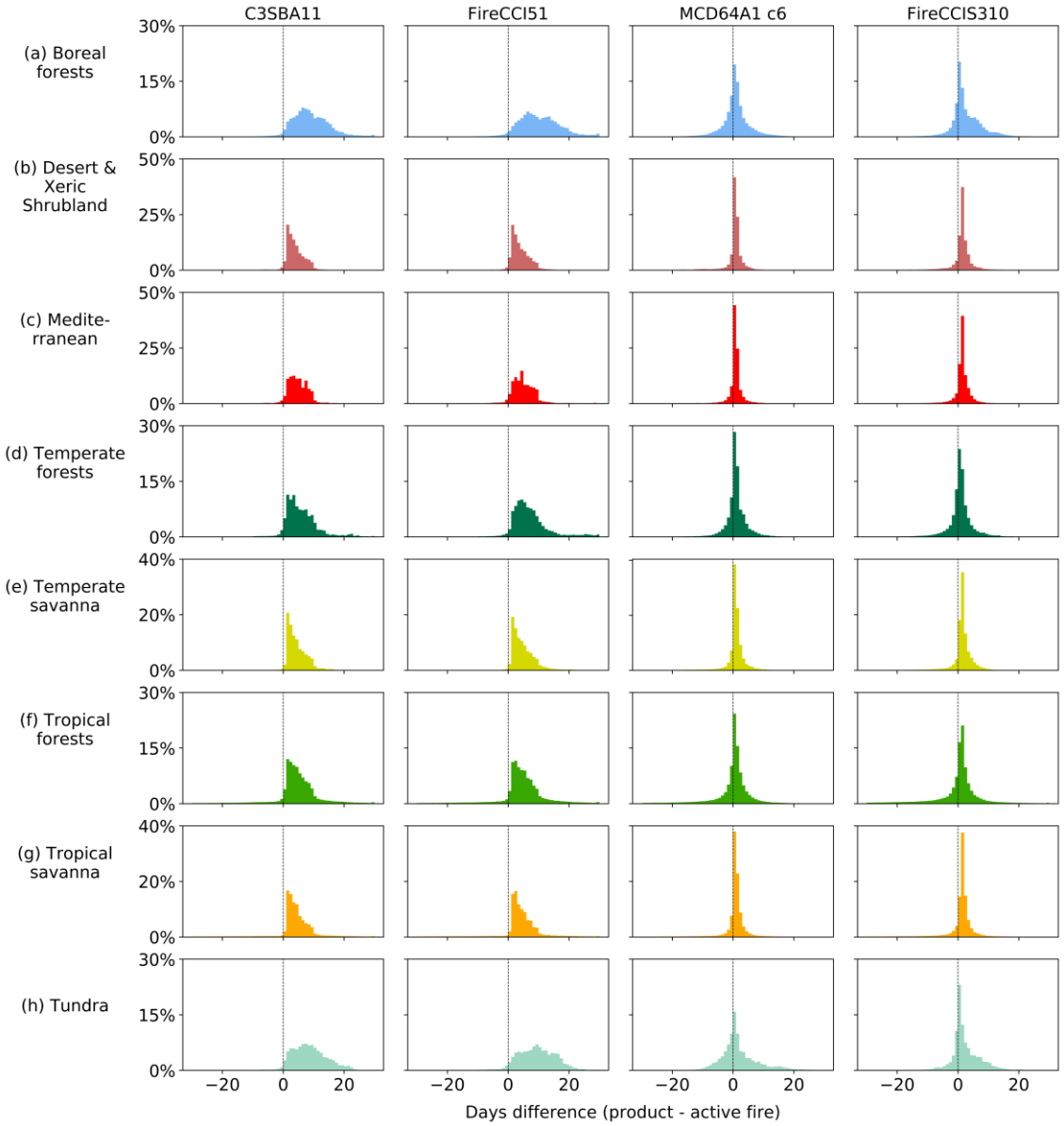

Figure 41: Temporal reporting accuracy for each product and biome compared to VIIRS active fires (VNP14IMGML) for the year 2019. Each bin of the histogram represents a 1-day step.

4 References

Belenguer-Plomer M.A. and Tanase M.A. (2020) ESA CCI ECV Fire Disturbance: D2.1.1 SAR-Optical Algorithm Theoretical Basis Document: version 1.1. Available at: https://climate.esa.int/en/projects/fire/key-documents/

Boschetti, L., Stehman, S., V., & Roy, D.P. (2016). A stratified random sampling design in space and time for regional to global scale burned area product validation. Remote Sensing of Environment, 186, 465-478. https://doi.org/10.1016/j.rse.2016.09.016

Boschetti, L., Roy, D. P., Giglio, L., Huang, H., Zubkova, M., Humber, M. L. (2019) Global validation of the collection 6 MODIS burned area product. Remote Sensing of Environment 235, 111490, https://doi.org/10.1016/j.rse.2019.111490

Chuvieco, E., Lizundia-Loiola, J., Pettinari, M. L., Ramo, R., Padilla, M., Tansey, K., Mouillot, F., Laurent, P., Storm, T., Heil, A., and Plummer, S. (2018). Generation and analysis of a new global burned area product based on MODIS 250 m reflectance bands and thermal anomalies, Earth Syst. Sci. Data, 10, 2015-2031, https://doi.org/10.5194/essd-10-2015-2018.

Chuvieco, E., Roteta, E., Sali, M., Stroppiana, D., Boettcher, M., Kirches, G., Storm, T., Khairoun, A., Pettinari, M.L., Albergel, C. (2022) Building a small fire database for Sub-Saharan Africa from Sentinel-2 high-resolution images. In review.

Cohen, W.B., Yang, Z., & Kennedy, R.E. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync - Tools for calibration and validation. Remote Sensing of Environment, 114, 2911-2924. https://doi.org/10.1016/j.rse.2010.07.010

Congalton, R.G., & Green, K. (1999). Assessing the Accuracy of Remotely Sensed Data: Principles and Applications. Boca Raton: Lewis Publishers.

Dice, L.R. (1945). Measures of the Amount of Ecologic Association Between Species. Ecology, 26, 297-302. https://doi.org/10.2307/1932409

Franquesa, M., Vanderhoof, M.K., Stavrakoudis, D., Gitas, I.Z., Roteta, E., Padilla, M., Chuvieco, E., 2020. Development of a standard database of reference sites for validating global burned area products. Earth Syst. Sci. Data, 12, 3229–3246. https://doi.org/10.5194/essd-12-3229-2020

Franquesa, M., Lizundia-Loiola, J., Stehman, S.V., Chuvieco, E., 2022. Using long temporal reference units to assess the spatial accuracy of global satellite-derived burned area products, Remote Sensing of Environment, 269, https://doi.org/10.1016/j.rse.2021.112823.

Giglio, L., Boschetti, L., Roy, D.P., Humber, M.L., Justice, C.O., 2018. The Collection 6 MODIS burned area mapping algorithm and product. Remote Sens. Environ. 217, 72–85, https://doi.org/10.1016/j.rse.2018.08.005

Giglio, L., Randerson, J. T., and van der Werf, G. R., (2013), Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). Journal of Geophysical Research: Biogeosciences, 118, 317– 328, https://doi.org/10.1002/jgrg.20042.

Kennedy, R.E., Yang, Z., & Cohen, W.B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr - Temporal segmentation algorithms. Remote Sensing of Environment, 114, 2897-2910. https://doi.org/10.1016/j.rse.2010.07.008

Key, C. H. & N. C. Benson (1999) The Normalized Burn Ratio, a Landsat TM radiometric index of burn severity. US Geological Survey, Department of the Interior.

Latifovic, R., Olthof, I. (2004). Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data. Remote Sensing of Environment, 90, 153-165. https://doi.org/10.1016/j.rse.2003.11.016

Lemajic, S., Vajsová, B. and Aastrand, P., (2018) New sensors benchmark report on PlanetScope: Geometric benchmarking test for Common Agricultural Policy (CAP) purposes, EUR 29319 EN, Publications Office of the European Union, Luxembourg, 2018, ISBN 978-92-79-92833-8, https://doi.org/10.2760/178918, JRC111221.

Lizundia-Loiola, J., Franquesa, M., Boettcher, M., Kirches, G., Pettinari, M.L., Chuvieco, E. Implementation of the Burned Area Component of the Copernicus Climate Change Service: From MODIS to OLCI Data. Remote Sensing 13 (21), 4295, 2021. https://doi.org/10.3390/rs13214295

Lizundia-Loiola, J., Franquesa, M., Khaïroun, A., Chuvieco, E. (2022). Global burned area mapping from Sentinel-3 Synergy and VIIRS active fires. Remote Sensing of Environment, in review.

Lizundia-Loiola, J., Otón, G., Ramo, R. and Chuvieco, E. (2020). A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data. Remote Sensing of Environment, 236, 111493, https://doi.org/10.1016/j.rse.2019.111493.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., & Kassem, K.R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. BioScience, 51, 933-938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Otón, G., Lizundia-Loiola, J., Pettinari, M.L., Chuvieco, E. (2021) Development of a consistent global long-term burned area product (1982–2018) based on AVHRR-LTDR data. International Journal of Applied Earth Observation and Geoinformation 103, 102473. https://doi.org/10.1016/j.jag.2021.102473

Padilla, M., Stehman, S.V., & Chuvieco, E. (2014). Validation of the 2008 MODIS-MCD45 global burned area product using stratified random sampling. Remote Sensing of Environment, 144, 187-196. https://doi.org/10.1016/j.rse.2014.01.008

Padilla, M., Stehman, S.V., Ramo, R., Corti, D., Hantson, S., Oliva, P., Alonso, I., Bradley, A., Tansey, K., Mota, B., Pereira, J.M., & Chuvieco, E. (2015). Comparing the Accuracies of Remote Sensing Global Burned Area Products using Stratified Random Sampling and Estimation. Remote Sensing of Environment, 160, 114-121. https://doi.org/10.1016/j.rse.2015.01.005

Padilla M., Olofsson P., Stehman S.V., Tansey K., Chuvieco E. (2017). Stratification and sample allocation for reference burned area data. Remote Sensing of Environment, 203, 240-255. https://doi.org/10.1016/j.rse.2017.06.041

Padilla, M., Wheeler, J., Tansey, K. (2018) ESA CCI ECV Fire Disturbance: D4.1.1. Product Validation Report, version 2.1. Available at: https://climate.esa.int/en/projects/fire/key-documents/.

Pettinari, M.L., Roteta. E. (2021) ESA CCI ECV Fire Disturbance: D2.4.4. Product User Guide – Small Fire Database, version 2.0. Available at: https://climate.esa.int/en/projects/fire/key-documents/

Roteta, E., Bastarrika, A., Padilla, M., Storm, T., Chuvieco, E. (2019). Development of a Sentinel-2 burned area algorithm: Generation of a small fire database for sub-Saharan Africa, Remote Sensing of Environment, 222, 1-17, https://doi.org/10.1016/j.rse.2018.12.011.

Rouse Jr, J. W., R. Haas, J. Schell & D. Deering (1974) Monitoring vegetation systems in the Great Plains with ERTS. Available at: https://ntrs.nasa.gov/citations/19740022614

Roy, D. P., P. G. H. Frost, C. O. Justice, T. Landmann, J. L. Le Roux, K. Gumbo, S. Makungwa, K. Dunham, R. Du Toit, K. Mhwandagara, A. Zacarias, B. Tacheba, O. P. Dube, J. M. C. Pereira, P. Mushove, J. T. Morisette, S. K. Santhana Vannan & D. Davies (2005) The Southern Africa Fire Network (SAFNet) regional burned-area product-validation protocol. International Journal of Remote Sensing, 26**,** 4265-4292. https://doi.org/10.1080/01431160500113096

Roy, D. P., H. Huang, L. Boschetti, L. Giglio, L. Yan, H.H. Zhang, Z. Li, (2019) Landsat-8 and Sentinel-2 burned area mapping - A combined sensor multi-temporal change detection approach. Remote Sensing of Environment, 231, 111254, https://doi.org/10.1016/j.rse.2019.111254.

Schroeder, W., & Giglio, L. (2018). Visible Infrared Imaging Radiometer Suite (VIIRS) 375 m & 750 m Active Fire Products, Product User's Guide Version 1.4. NASA VIIRS Land Science Investigator Processing System (SIPS). Available at: https://lpdaac.usgs.gov/documents/427/VNP14_User_Guide_V1.pdf.

Stroppiana, D., Sali, M., Busetto, L., Boschetti, M., Franquesa, M. (2022a). FireCCI_Africa_2019_S2: reference fire perimeters obtained from Sentinel-2 imagery over Africa continental for the year 2019. In: e-cienciaDatos, https://doi.org/10.21950/VKFLCH

Stroppiana, D., Sali, M., Busetto, L., Boschetti, M., Ranguetti, L., Franquesa, M., Pettinari, M.L., Chuvieco, E. (2022b). Sentinel-2 sampling design and reference perimeters to assess the accuracy of burned area products over Africa for the year 2019. ISPRS Journal of Photogrammetry and Remote Sensing, in review.

Annex 1: Acronyms and abbreviations

AUST	Australia
BA	Burned Area
BOAS	Boreal Asia
BONA	Boreal North America
CCI	Climate Change Initiative
Ce	Commission error
CEAM	Central America
CEAS	Central Asia
DC	Dice Coefficient
DOY	Day of the Year
DRC	Democratic Republic of Congo
ECV	Essential Climate Variables
EO	Earth Observation
EQAS	Equatorial Asia
ESA	European Space Agency
EURO	Europe
FireCCI41	MERIS Fire_cci v4.1
FireCCI50	MODIS Fire_cci v5.0
FireCCI51	MODIS Fire_cci v5.1
FireCCISFD11	Sentinel-2 SFD Fire_cci v1.1
FireCCISFD20	Sentinel-2 SFD Fire_cci v2.0
FireCCIS1S2AF10	Sentinel-1&2 Fire_cci test site in Africa v1.0
FN	False negative
FP	False positive
GEE	Google Earth Engine
HR	High Resolution
KML	Keyhole Markup Language
L	Validation unit Lenght
L8	Landsat 8
MCD64 c6	MODIS Burned Area product collection 6
MERIS	Medium Resolution Imaging Spectrometer
MIDE	Middle East
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration's
NBR	Normalized Burned Ratio
NBR2	Normalized Burned Ratio 2
NDVI	Normalized Difference Vegetation Index
NHAF	Northern Hemisphere Africa
NHSA	Northern Hemisphere South America
NIR	Near InfraRed
NPP	National Polar-orbiting Partnership
Oe	Omission error

PVIR	Product Validation and Intercomparison Report
RF	Random Forest
RGB	Red-Green-Blue composite
S1	Sentinel-1
S2	Sentinel-2
SAR	Synthetic Aperture Radar
SAR-O	SAR-Optical
SE	Standard Error
SEAS	Southern Asia
SFD	Small Fire Database
SHAF	Southern Hemisphere Africa
SHSA	Southern Hemisphere South America
SWIR	Short Wave InfraRed
TENA	Temperate North America
TN	True negative
TP	True positive
TSA	Thiesen Scene Area
UTM	Universal Transverse Mercator
VHR	Very High Resolution
WGS84	World Geodetic System 1984