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	20/07/2024
Version	3.0
Author	Daniela Stroppiana, Bhogendra Mishra, Matteo Sali, Erika Solano, Amin Khairoun, M. Lucrecia Pettinari
Document Ref.	Fire_cci_D4.1_PVIR_v3.0
Document type	Public

To be cited as: Stroppiana D., Mishra B., Sali M., Solano E., Khairoun A., Pettinari M.L. (2024) ESA CCI ECV Fire Disturbance: D4.1 Product Validation and Intercomparison Report, version 3.0. Available at: https://climate.esa.int/en/projects/fire/key-documents/

Project Partners

Prime Contractor/ Scientific Lead & ProjectManagement	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 GmbH (Germany)
Climate Research Group	MPIC – Max Planck Institute for Chemistry (Germany) CNRS – National Center for Scientific Research (France)

Distribution

Affiliation	Name	Address	Copies
ESA	Clément Albergel (ESA)	clement.albergel@esa.int	electronic copy
Project Team	Emilio Chuvieco (UAH) M. Lucrecia Pettinari (UAH) Amin Khairoun (UAH) Erika Solano (UAH) Consuelo Gonzalo (UPM) Dionisio Rodríguez (UPM) Ángel García Pedrero (UPM) Daniela Stroppiana (CNR) Thomas Storm (BC) Martin Böttcher (BC) Florent Mouillot (CNRS) Philippe Ciais (CNRS) Guido van der Werf (VU)	emilio.chuvieco@uah.es mlucrecia.pettinari@uah.es amin.khairoun@uah.es erika.solano@uah.es consuelo.gonzalo@upm.es dionisio.rodriguez@ulpgc.es angelmario.garcia@upm.es stroppiana.d@irea.cnr.it thomas.storm@brockmann-consult.de martin.boettcher@brockmann-cons... florent.mouillot@cefe.cnrs.fr philippe.ciais@lsce.ipsl.fr g.r.vander.werf@vu.nl	electronic copy

Summary

This Product Validation and Intercomparison Report (PVIR) describes the implementation of the validation methods and results for assessing the accuracy of global burned area products and regional high resolution burned area products (Amazonia, Sahel, Siberia), as well as the intercomparison between the currently existing global burned area products.

	Affiliation/Function	Name	Date
Prepared	CNR UAH	Daniela Stroppiana Boghendra Mishra Matteo Sali Erika Solano Romero Amin Khairoun M. Lucrecia Pettinari	10/07/2024
Reviewed	UAH	M. Lucrecia Pettinari Emilio Chuvieco	20/07/2024
Authorized	UAH	Emilio Chuvieco	20/07/2024
Accepted	ESA	Clément Albergel	21/09/2024

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
3.0	20/07/2024	Third version of the document

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	Text updated Small changes in the text Sections added

Issue	Date	Request	Location	Details
			Section 4	New references added
2.1	07/06/2022	CNR UAH	Section 2.2 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	Figure 2 updated. Small changes in the text. Text expanded. Figures 21, 22 and 23 updated. Text expanded, figure 38 updated.
3.0	20/07/2024	UAH	All document	Text 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 global and regional burned area (BA) products derived by applying FireCCI algorithms. This document describes (see also Table 1):

1. the generation of the reference datasets used for the validation;
2. the validation results for global (FireCCI51 and FireCCIS311) and regional high resolution (FireCCISFDL) BA products;
3. the comparison with coarser resolution products such as FireCCI51, FireCCIS311, MCD64A1 C6 and C3SBA11 (inter-comparison activities).

Table 1. BA product pair comparisons (product vs. reference) carried out in the validation activities.

ROI	FireCCISFDL	FireCCI51	FireCCIS311	MCD64A1 C6
Africa Sahel	1990-2019	2001-2019	-	2000-2019
Russia Siberia	1990-2019	2001-2019	-	2000-2019
South America	1990-2019	2001-2019	-	2000-2019
Global	-	2017-2021	2019-2022	2017-2022

In order to assess the accuracy of the 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, we relied on Landsat 8 (L8) and Landsat 9 (L9) for the validation of the global products and on Landsat 5 (L5), L8 and Sentinel-2 (S2) for the validation of the regional high resolution products.

2 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 from EO data since systematic collection of reliable and representative ground truth/in situ fire data is hardly feasible. Validation of moderate spatial resolution BA products (e.g., e.g. FireCCI51, FireCCIS311) can be therefore carried by comparison with reference perimeters generated from decametric spatial resolution data as Landsat and Sentinel (Chuvieco et al., 2018; Lizundia-Loiola et al., 2021). In the case of BA products derived from high/medium resolution data higher remotely sensed data should be used for generating reference perimeters. However, systematic acquisition of high/very high data over large areas might be not easily accessible and/or might have high costs of acquisition and processing. In these cases, satellite images with comparable spatial resolution are commonly used (e.g. Roteta et al., 2019, Roy et al., 2019; Stroppiana et al., 2022a) and independence of the reference dataset is achieved by independent processing. For this reason, in this project the validation of the multi-annual high resolution FireCCISFDL BA product was carried out by generating reference perimeters from Landsat and Sentinel-2 data, also considering that only Landsat missions cover past years.

The protocol implemented for validation of global and regional BA products follows these steps:

- Identification of validation units: validation units are the regions where the reference and the product were compared. They are selected by stratified random sampling in each calendar year, considering the major biomes (Dinerstein et al., 2017) and regions with high and low fire activity as identified from yearly-cumulated burned area.
- Generation of reference fire perimeters: reference fire perimeters are generated from supervised classification of consecutive satellite images, i.e. temporal series of images (Landsat for global and regional validation, and S2 for regional validation) called reference long units.
- Computation of accuracy metrics: the error matrix (Congalton and Green 1999; Latifovic and Olthof 2004) is extracted by comparing burned area products and reference fire perimeters over the validation units and the accuracy metrics are computed: commission error ratio, omission error ratio, Dice Coefficient (DC) (Dice 1945), bias and relative bias. Accuracy metrics are complemented by agreement maps showing the spatial distribution of agreement and disagreement regions over the validation unit.
- Computation of spatially distributed agreement: in the case of regional validation, a 5 km by 5 km grid layer is used to compare proportion of burned cells from reference and FireCCI BA products to derive regressive parameters and metrics such as slope, intercept, coefficient of determination (R2), Root Mean Squared Error (RMSE) (Boschetti et al. 2019, Sali et al. 2021).

2.1 Reference long units

A **long temporal reference unit** is a spatio-temporal partition of the source EO data archive where time series of images (Landsat and/or Sentinel-2) are acquired and processed to generate fire perimeters. The time series over each unit is composed of consecutive images where each image pair (i.e., **short unit**) is classified to extract burned polygons. A maximum time step between consecutive images is set to exploit the greatest discrimination of recent burns especially 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 is increased to guarantee a sufficient length of the time series. The burned polygons preserve as date of burning the acquisition date of the short unit and are assumed to be burned only once within the time span covered by the long unit. Short units fire perimeters are cumulated between the first and last date of the time period covered by the long unit (Franquesa et al., 2020; Franquesa et al., 2022, Stroppiana et al., 2022b).

Spatially, each unit covers an area of ~10 000 km2 (100 km x 100 km) to be consistent with previous reference fire datasets derived for the validation of FireCCI BA products at regional and global scales (Stroppiana et al., 2022b; Chuvieco et al., 2022; Franquesa et al., 2022; Lizundia-Loiola et al., 2021). The sampling design implemented for the selection of the validation units differs for the regional and global products and depend on the source of EO data (Landsat/Sentinel-2).

Temporally, a **short unit** is composed of two consecutive cloud-free images/scenes (pairs of images) whereas a **long unit** is composed of more than two consecutive pairs (consecutive short units) over the same unit (Figure 1). Long temporal units have shown to be a robust approach to estimate spatial accuracy by disentangling spatial from temporal component of BA detection errors. Reducing the impact of date mismatch is particularly relevant for BA products derived from EO data with low temporal resolution or for areas with high cloud coverage (Franquesa et al., 2022). Reference long temporal units are identified from the available images based on **cloud cover**, **time interval between consecutive images**, and **total length of the unit** (Franquesa et al., 2022; Stroppiana et al. 2022b).

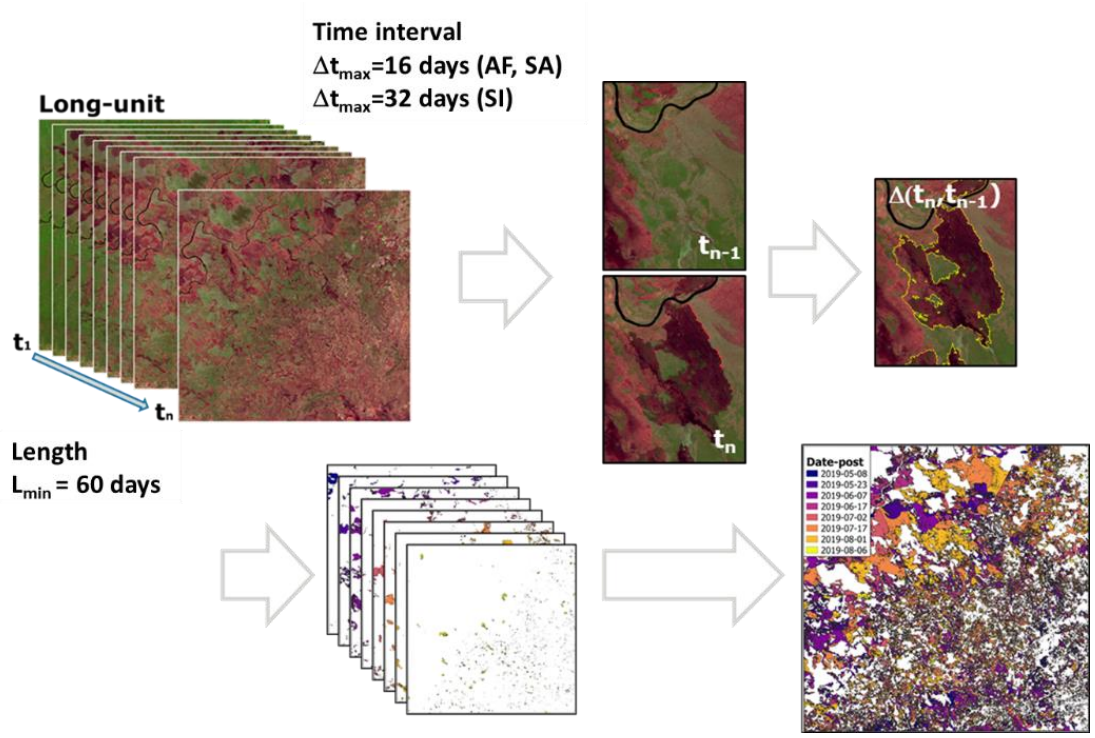

Figure 1. Example of the implementation of the algorithm for the extraction of reference validation units from the classification of consecutive images pairs. In this figure source images are Sentinel-2.

3 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 2 shows the flowchart of all the steps, from the selection of the sample units to the extraction of fire perimeters over the long unit.

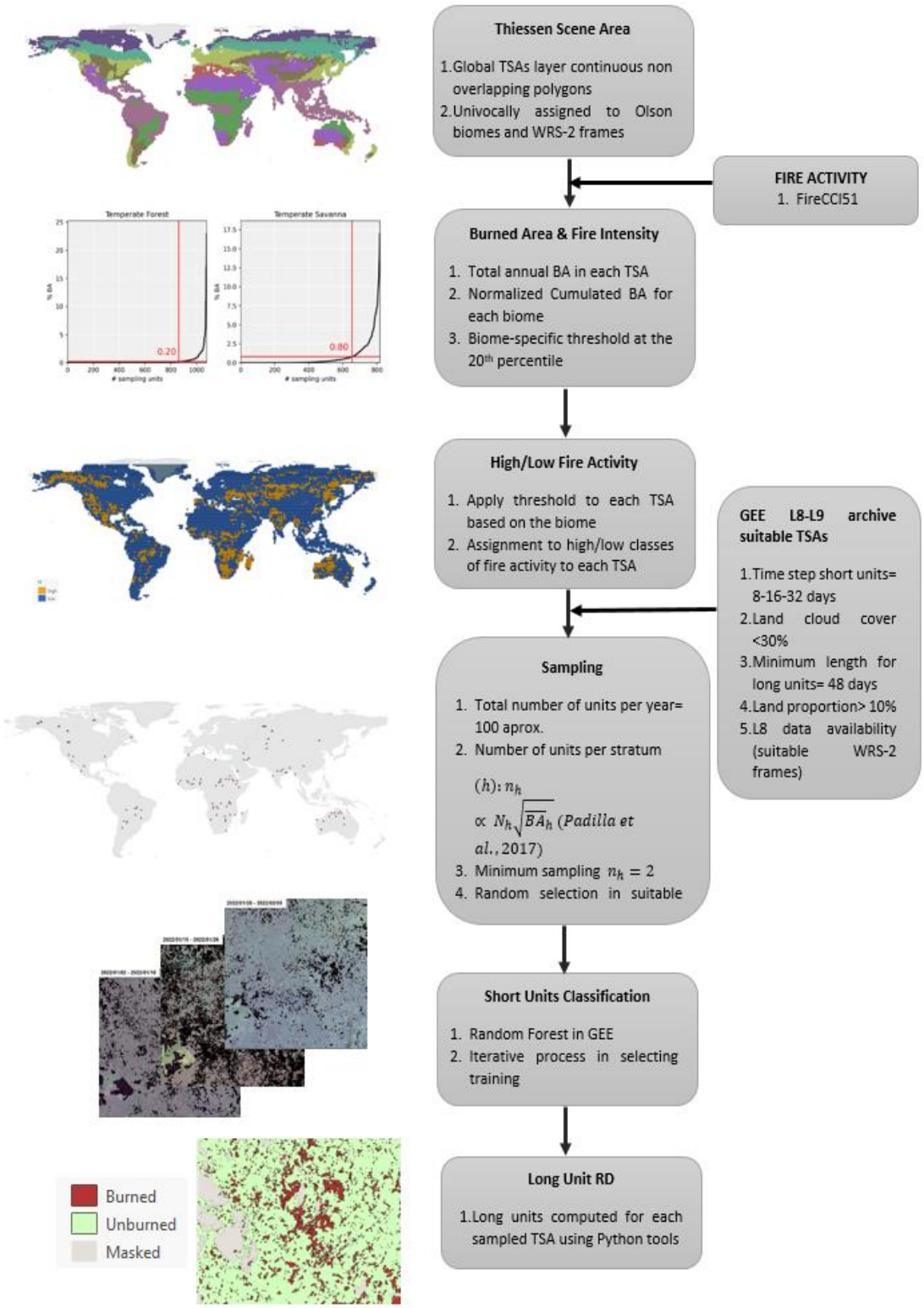

Figure 2. Flowchart of steps for extracting fire reference perimeters for the global validation: from sampling to classification of the short units and building long units files.

3.1 Global sampling scheme of long temporal reference unit

3.1.1 Spatial definition of global validation units

For global BA products, the spatial dimension of sampling units is based on Thiessen Scene Areas (TSAs). The number of selected units for each year is approximately 100 TSAs. As described in Section 2.1, for each TSA, a long sampling unit covering an area of 100 km x 100 km is composed of consecutive pairs of Landsat images (short units).

3.1.2 Temporal definition of global validation units

In the case of validation of global products, reference long units are composed of time series of Landsat (L8, L9) images defined based on the following criteria:

1. **Cloud cover**: scene cloud cover (CC<10%) and cumulated cloud cover (CC_{cum}<30%)
2. **Time interval** between consecutive images ($\Delta t \leq 16$ days)
3. **Length** of the time series (L>48 days)

Maximum cloud cover for each image was set to **10%** while **maximum cumulated cloud cover** to 30% in order to generate long reference units with the least possible cloud cover. The **time interval** between consecutive images was set to **16 days** to consider the minimum duration of the burned area spectral signal. The minimum **length** of the long unit was set to **48 days** to consider the reduced frequency of clear sky observation in some world biomes.

3.1.3 Stratification of global units

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 biomes (Dinerstein et al. 2017) and ii) areas with high and low fire activity. Biomes were aggregated into eight major vegetated biomes as shown in Table 1 in Franquesa et al. (2022). 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.

For each TSA, the major biome was assigned (Figure 3 top). The high/low fire activity class was assigned based on the total annual burned area derived from the MCD64A1 c6 BA product (Giglio et al. 2018, Figure 3 bottom). 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 20^{th} percentile of the normalized cumulated sum 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 4, the threshold values (highlighted in red) for each biome for the year 2022 are reported.

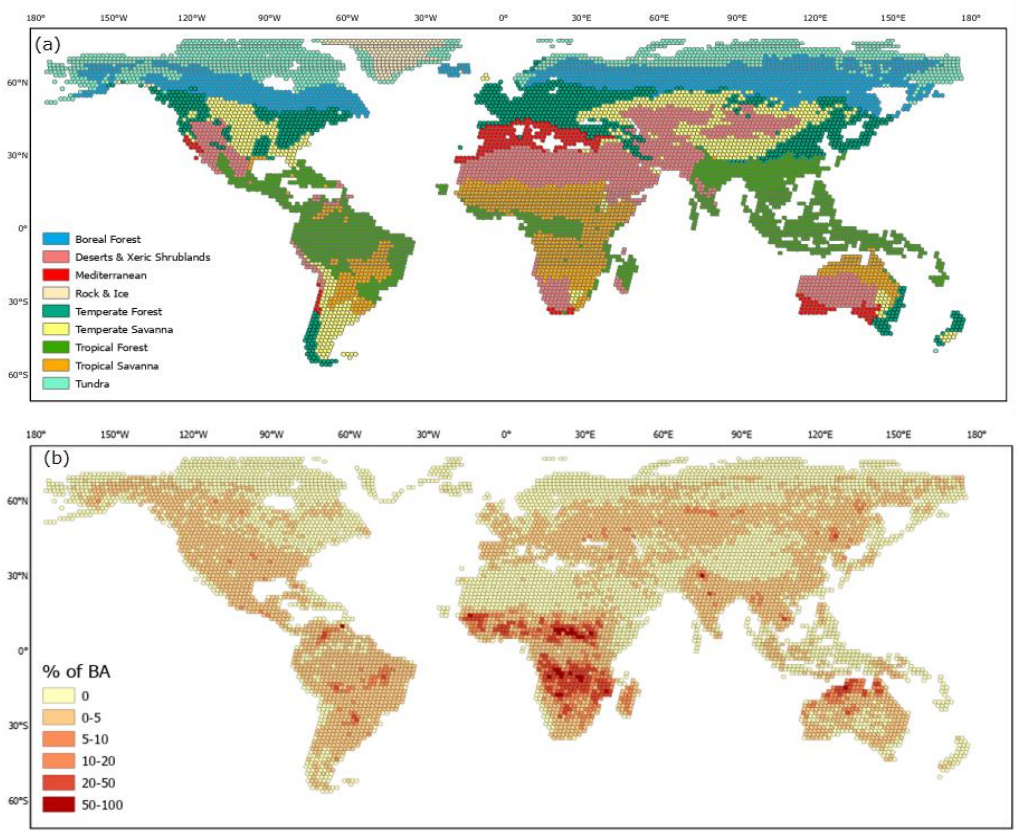

Figure 3. Two layers were used for the definition of the strata in the random stratification sampling: biomes (top panel) reclassified into 8 major biomes (plus Rock & Ice shown as beige), and percentage of BA with respect to the total land area of each TSA, based on the MCD64A1 c6 BA extent for the year 2022 (bottom panel) (Franquesa et al., 2022).

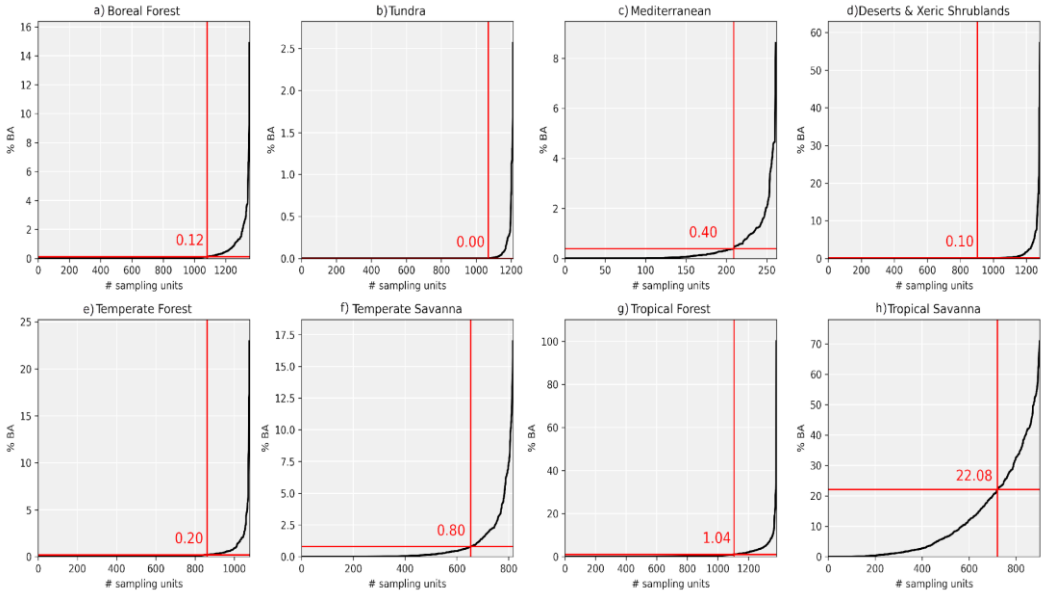

Figure 4. Example Thiessen scene areas (TSAs) for each biome plotted by increasing normalized total burned area for the year 2022; on the x-axis the increasing cumulated number of TSA (#sampling units). Red lines show the 20th percentile and the corresponding value of total annual burned area used as threshold for assigning high/low fire activity class (Franquesa et al., 2022).

3.1.4 Analysis of L8-L9 data availability 2022

Preliminary analysis of data availability at the global scale was carried out for the year **2022** to investigate the distribution of L8-L9 scenes suitable for sampling. The global L8-L9 archive available in Google Earth Engine (GEE) was investigated to identify, over each validation unit, L8-L9 temporal series **with maximum time step of 16 and/or 32 days** and cloud coverage percentage lower than 30% (from the metadata attribute *Land Cloud Cover*). Figure 5 and Figure 6 show maximum length and starting month of the long units when Δtmax=16 days and Δtmax=32 days, respectively. The figures confirm that Δtmax and the length are correlated, hence relaxing the condition on maximum time step, from 16 to 32 days, produces longer series (light green and yellow colours).

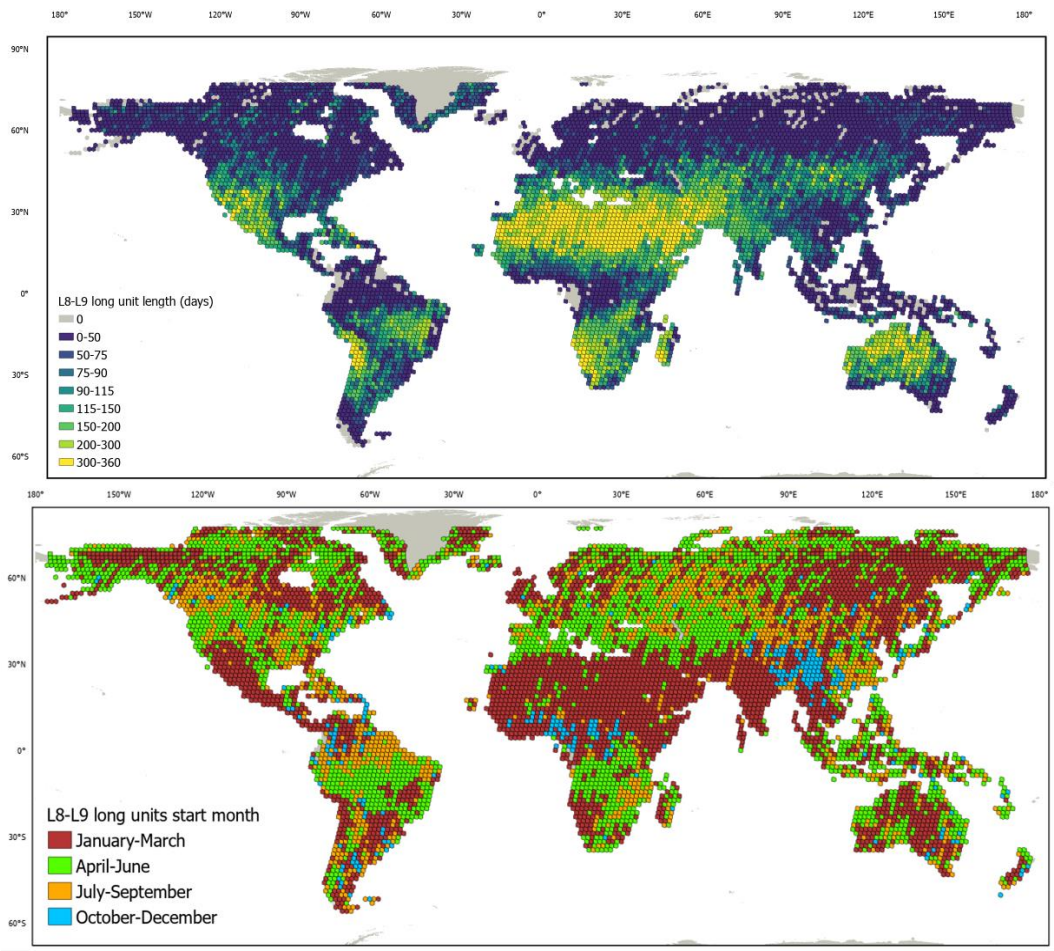

Figure 5. Length (days) (top) and starting month (grouped in three-month seasons) (bottom) of the longest temporal series for each TSA (validation unit) with a 16-day maximum time step between consecutive clear sky images. Grey areas show regions where no suitable data are available.

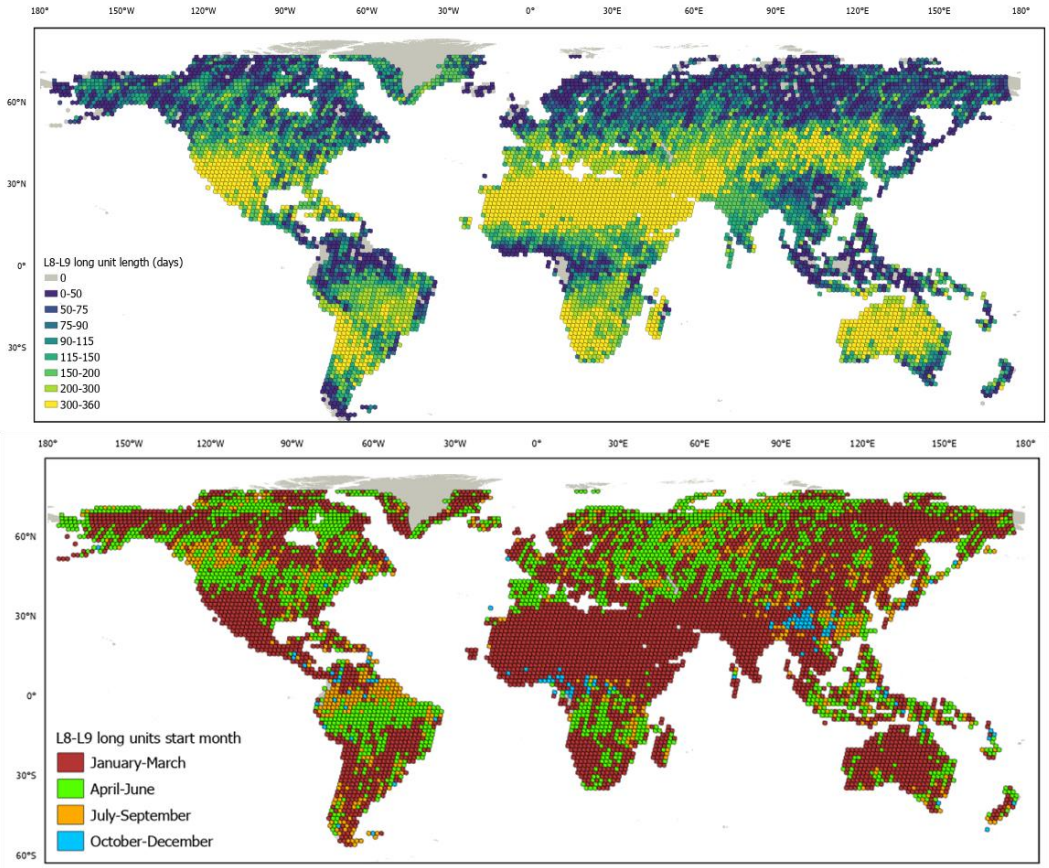

Figure 6. Length (days) (top) and starting month (grouped in three-month seasons) (bottom) of the longest temporal series for each TSA (validation unit) with a 32-day maximum time step between consecutive clear sky images. Grey areas show regions where no suitable data are available.

An analysis was also carried out, dividing the biomes into high/low fire activity sub-strata by applying threshold values. Figure 7 shows the number of TSAs available for sampling as a function of Lmin (values on top strips) with Δtmax=16 days. In order to have 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**.

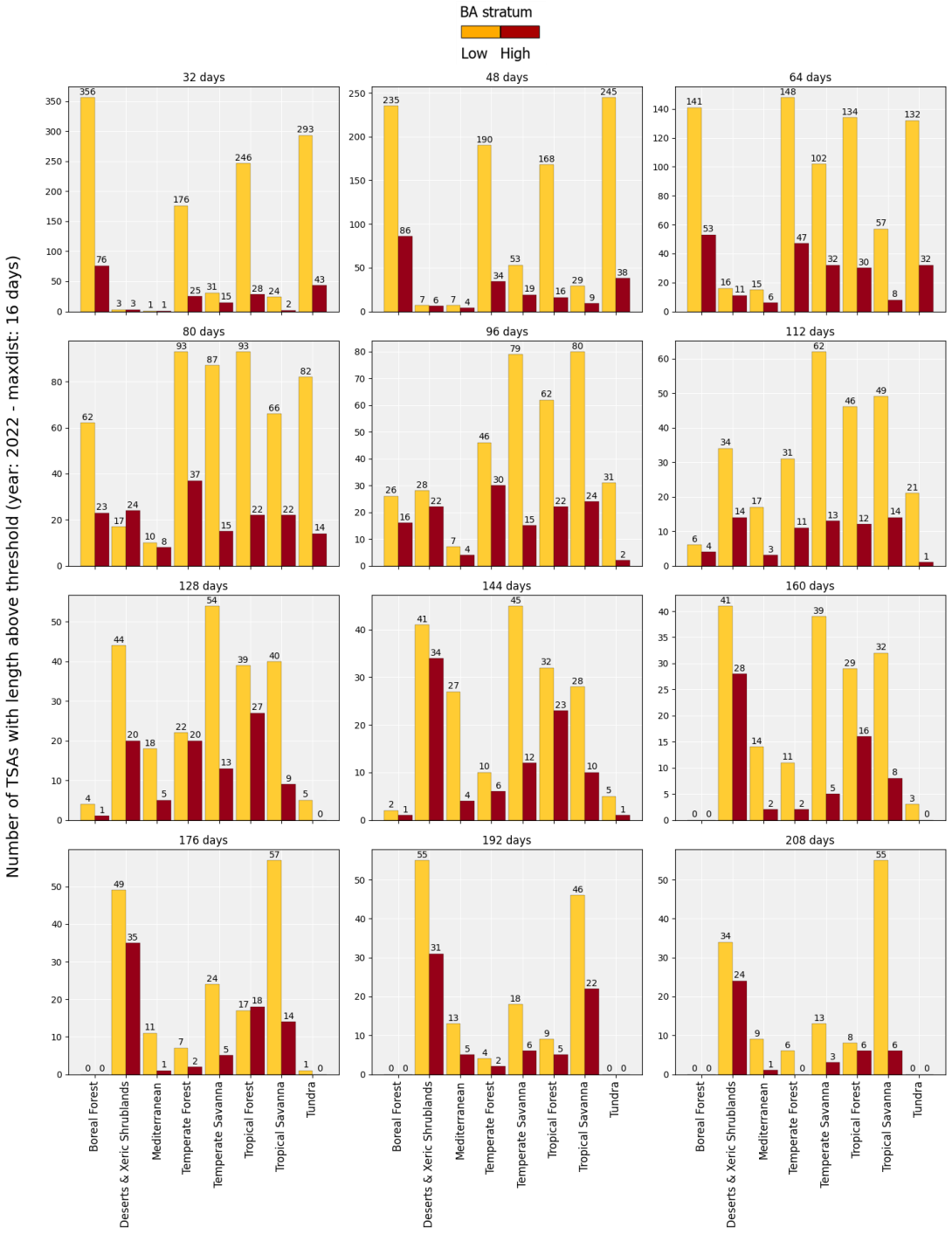

Figure 7. Number of L8-L9 TSAs available for sampling for each biome in the Low (yellow) and High (red) Fire Activity stratum as a function of the length of the minimum temporal dataset/validation unit required (number on top of each panel) with maximum time step set to 16 days.

3.1.5 Sampling cardinality of reference units

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

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

where n_h is the number of units to be sampled for stratum h (intersection biome/fire activity layers), $\underline{BA}_h$ is the average burned area for stratum h and N_h is the total amount of units available for sampling in stratum h. An iterative process was applied to ensure that $n_{h>} > 2$ in all strata while preserving as much as possible the optimal sample allocation (Franquesa et al., 2022; Padilla et al., 2017). 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 six years of the validation period (2017 to 2022) to select six samples, one for each year. A random sampling algorithm is then applied to each stratum to extract the number of validation units that are summarized in Table 2 and Table 3, where for each stratum (combination of biomes and fire activity) we show the number of TSAs sampled. The location of the sampled TSAs for the period 2017-2019 is shown in Figure 8 and for 2020-2022 is shown in Figure 9.

Table 2. Number of TSAs sampled in each stratum and for each year (2017-2019) 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	

Table 3. Number of TSAs sampled in each stratum and for each year (2020-2022) in high (N_{HIGH}) and low (N_{LOW}) fire activity strata.

Biomes	2020		2021		2022	
	N_{HIGH}	N_{LOW}	N_{HIGH}	N_{LOW}	N_{HIGH}	N_{LOW}
Boreal Forest	4	2	5	2	3	2
Deserts & xeric shrublands	6	2	8	2	9	2
Mediterranean Forests	2	2	2	2	2	2
Temperate Forest	4	2	4	2	4	2
Temperate savanna	4	4	4	3	4	3
Tropical forest	12	6	11	5	10	6

Biomes	2020		2021		2022	
	N_{HIGH}	N_{LOW}	N_{HIGH}	N_{LOW}	N_{HIGH}	N_{LOW}
Tropical savanna	21	31	21	31	21	31
Tundra	2	2	2	2	2	2
Global	55	51	57	49	55	50
Sample size by year	106		106		105	

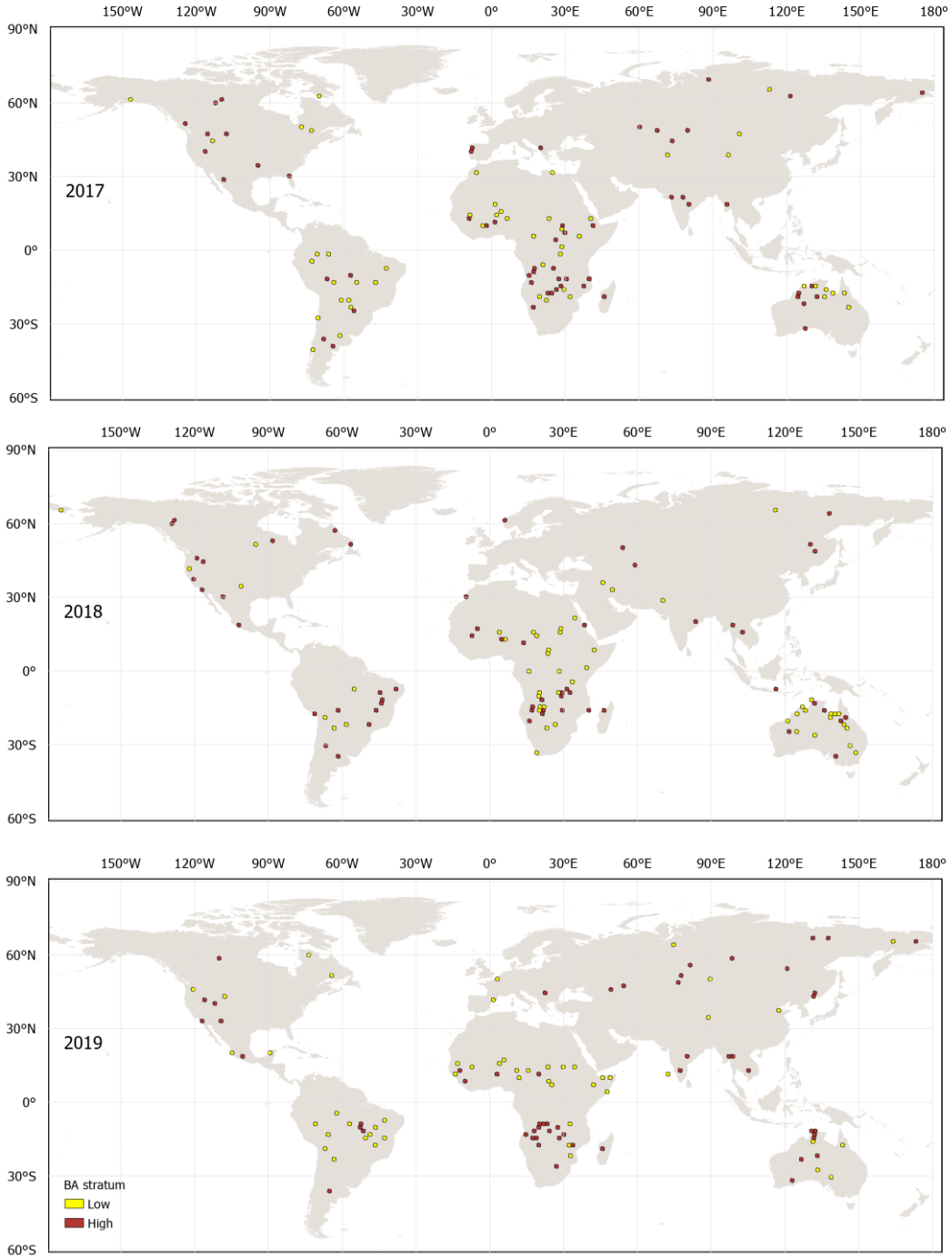

Figure 8. Spatial distribution of the TSAs sampled randomly worldwide for each year (2017-2019) and stratum of fire intensity.

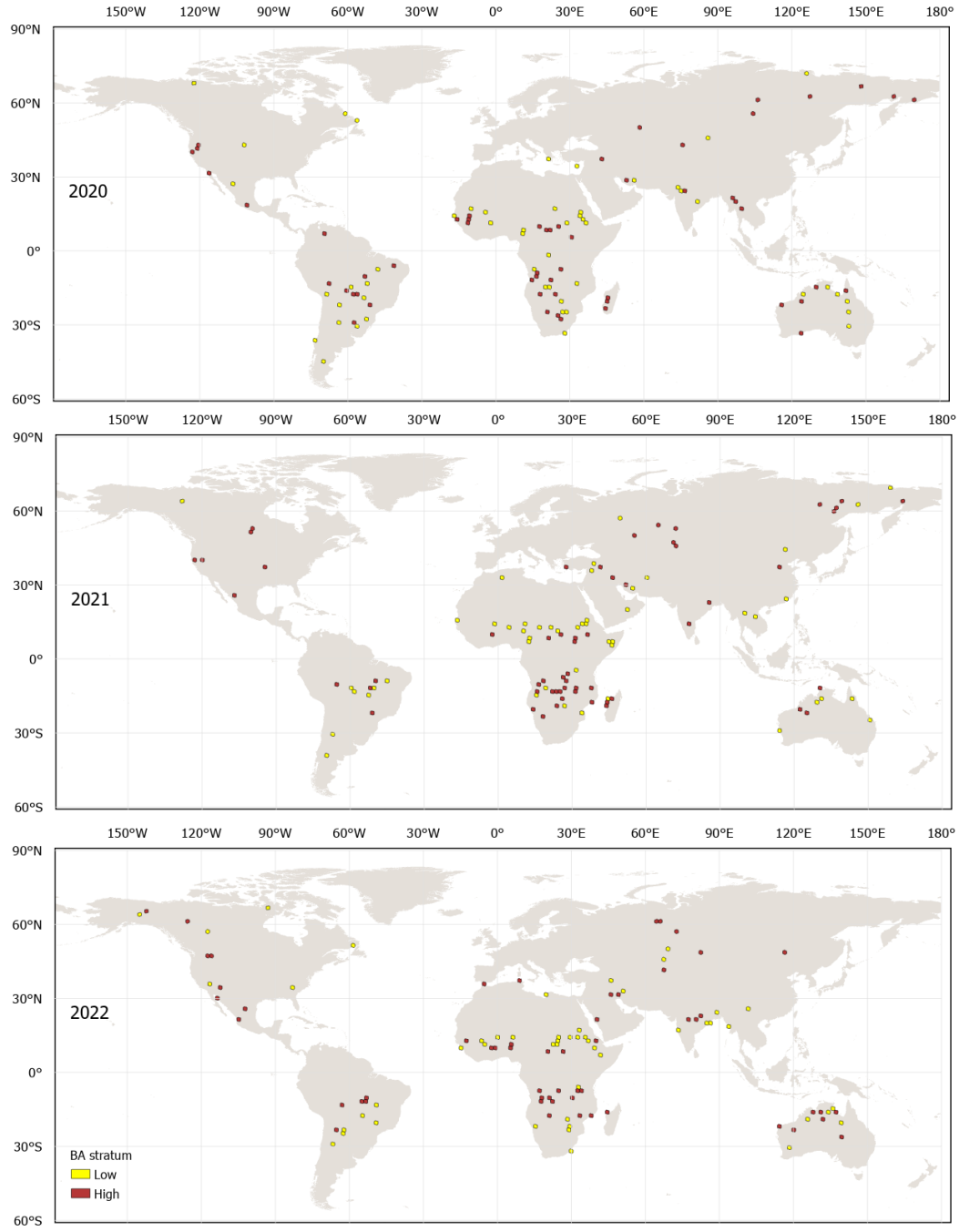

Figure 9. Spatial distribution of the TSAs sampled randomly worldwide for each year (2020-2022) and stratum of fire intensity.

3.1.6 Generation of the global fire reference perimeters

Figure 10 shows the flowchart of the steps for extracting fire perimeters over L8-L9 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-L9 short units (consecutive L8-L9 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-L9 long unit. The steps are described below.

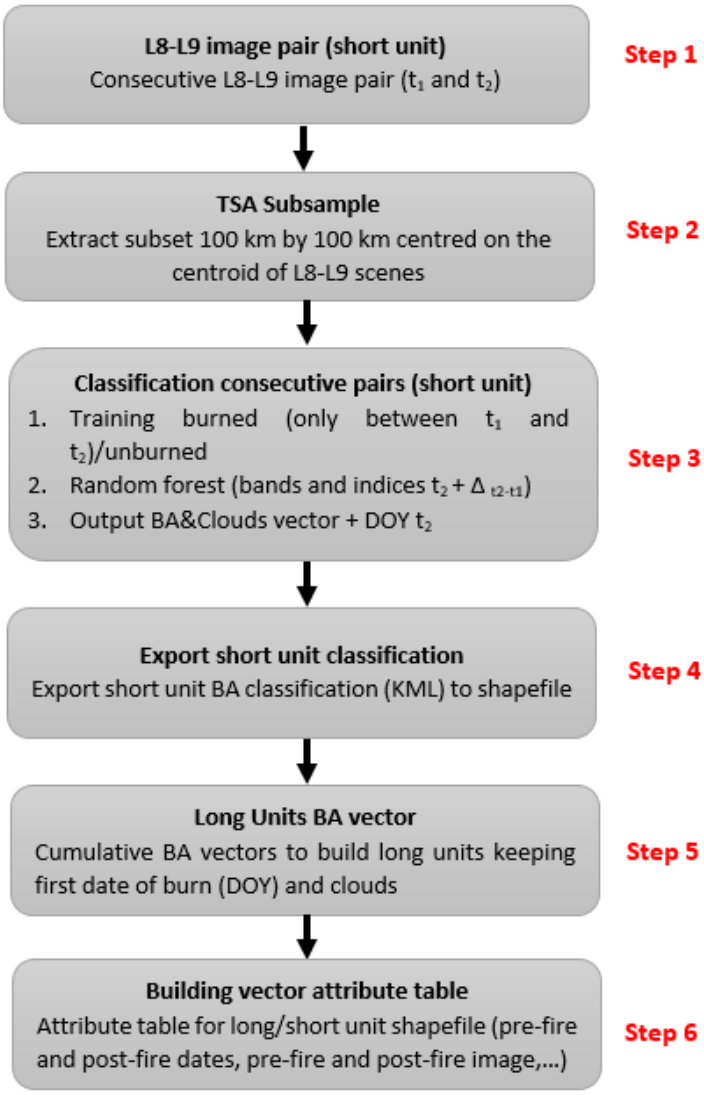

Figure 10. Flowchart showing steps for extraction of fire reference perimeters over sampled TSAs / validation units.

1. **L8-L9 image pair (short unit)**: over each TSA, consecutive L8-L9 scenes (Δtmax=16 days, condition relaxed in case of persistent cloud cover) made available as Level 2 product (surface reflectance, USGS Landsat 8-9 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-L9 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 & 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-L9 scene identifier (Scene ID), detection dates, category (burned, unburned and masked).

The core of the processing of the L8-L9 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-L9 frame path/row;
- L8-L9 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-L9 frames;
- Cloud Mask in vector format;
- Training polygons in vector format;

The classification of the L8-L9 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) and 317 for the years 2020 to 2022.

Figure 11 shows an example fire reference perimeter product extracted from a L8-L9 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. Figure 12 shows an example of attribute table for fire reference perimeters over the long unit.

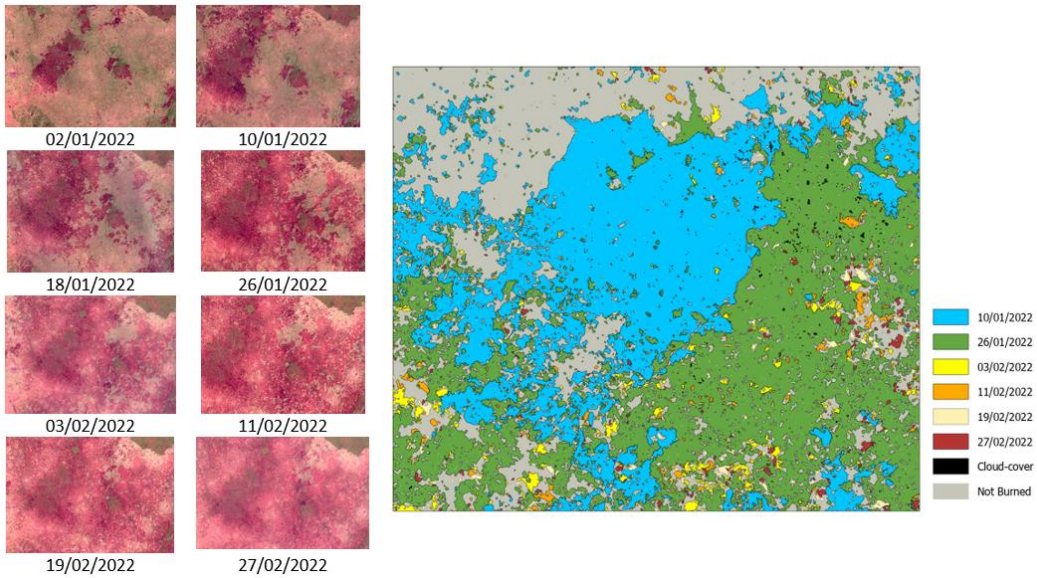

Figure 11. Example of reference fire perimeters extracted over L8-L9 frame 190/053 (Path/Row), Africa; on the left RGB false color composites of the L8-L9 scenes that are part of the validation long unit, in the right the reference burned area perimeters extracted by RF classification with reference 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	02/01/2022	27/02/2022	LC81900532022002LG...	LC91900532022058LG...	190	53	2022	5762077891
2	02/01/2022	27/02/2022	LC81900532022002LG...	LC91900532022058LG...	190	53	2022	85201042,3305
1	19/02/2022	27/02/2022	LC81900532022050LG...	LC91900532022058LG...	190	53	2022	25200
1	03/02/2022	11/02/2022	LC81900532022034LG...	LC91900532022042LG...	190	53	2022	10800,0015
1	10/01/2022	26/01/2022	LC91900532022010LG...	LC91900532022026LG...	190	53	2022	99000
1	03/02/2022	11/02/2022	LC81900532022034LG...	LC91900532022042LG...	190	53	2022	401400,0015
1	26/01/2022	03/02/2022	LC91900532022026LG...	LC81900532022034LG...	190	53	2022	1302299,9985
1	10/01/2022	26/01/2022	LC91900532022010LG...	LC91900532022026LG...	190	53	2022	900
1	11/02/2022	19/02/2022	LC91900532022042LG...	LC81900532022050LG...	190	53	2022	18000
1	19/02/2022	27/02/2022	LC81900532022050LG...	LC91900532022058LG...	190	53	2022	94599,696355
1	26/01/2022	03/02/2022	LC91900532022026LG...	LC81900532022034LG...	190	53	2022	499500,003
1	26/01/2022	03/02/2022	LC91900532022026LG...	LC81900532022034LG...	190	53	2022	3425542,18362
1	03/02/2022	11/02/2022	LC81900532022034LG...	LC91900532022042LG...	190	53	2022	109800
1	26/01/2022	03/02/2022	LC91900532022026LG...	LC81900532022034LG...	190	53	2022	4231800,006
1	10/01/2022	26/01/2022	LC91900532022010LG...	LC91900532022026LG...	190	53	2022	9900
1	11/02/2022	19/02/2022	LC91900532022042LG...	LC81900532022050LG...	190	53	2022	15300

Figure 12. 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-L9 scene ID of pre-fire and post-fire L8-L9 images, path and row the WRS-2 L8-L9 frame identifiers, year is the reference year and area is the area of each polygon.

3.1.7 Computation of accuracy metrics for global products

Reference fire perimeters and FireCCI 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 4). The area of agreement and disagreement between the BA product (map) and the reference data (Padilla et al., 2017; Padilla et al., 2014) (e_{ij}) used to estimate the accuracy metrics were derived from the confusion matrix (Table 5).

Table 4. Accuracy metrics computed from the error matrix

Accuracy metric name	Equation
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	$relB = \frac{e_{12} - e_{21}}{e_{+1}}$

Table 5. 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}	

3.1.8 Results of accuracy metrics FireCCI products

Global accuracy estimates of the two BA products (FireCCI51 for 2017-2021 and FireCCIS311 for 2019-2022) and their standard errors (SE, results reported as ±SE) are presented for the entire validated period (2017–2022) in Table 6.

Table 6. Global accuracy estimates (%) for each validation year for FireCCI51 and FireCCIS311. 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 (3.0)	-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)
	2020	65.5 (5.2)	21.7 (1.7)	47.9 (5.0)	-55.9 (6.7)
	2021	62.8 (4.9)	21.9 (2.8)	50.4 (5.0)	-52.4 (4.8)
FireCCIS311	**2019**	41.2 (3.0)	19.2 (1.7)	68.1 (2.5)	-27.2 (2.7)
	2020	57.2 (5.6)	19.1 (1.6)	56.0 (4.9)	-47.1 (6.8)
	2021	53.2 (3.4)	20.3 (3.1)	58.9 (3.5)	-41.3 (2.4)
	2022	53.0 (3.3)	18.8 (2.4)	59.5 (3.0)	-42.1 (3.5)

Figure 13 represents the precision on a global scale of the evaluated metrics and shows that the FireCCI51 product presented the highest estimated accuracy in the 2017 to 2019 period, with a DC value > 63%. In relation to this same parameter, the year 2020 presented the lowest accuracy with a value of 47.9%. Meanwhile, FireCCIS311 product presented the highest DC in 2019 with 68% and for the period from 2020 to 2022 the values were in a range from 56 to 59.5%.

It is observed that for the period 2020-2022 there is an increase of OE in FireCCI51 with respect to the years 2017 to 2019, which derives in a lower accuracy in the metrics of the products, because an increase of omission errors decreases the accuracy of the DC and increases the underestimation with respect to relB. The same happens in the case of FireCCIS311, which presents higher omission values in the period 2020 to 2022 with respect to 2019. In relation to the commission errors, the values are more similar for all years in both products, remaining in general at values of 15% and 22%.

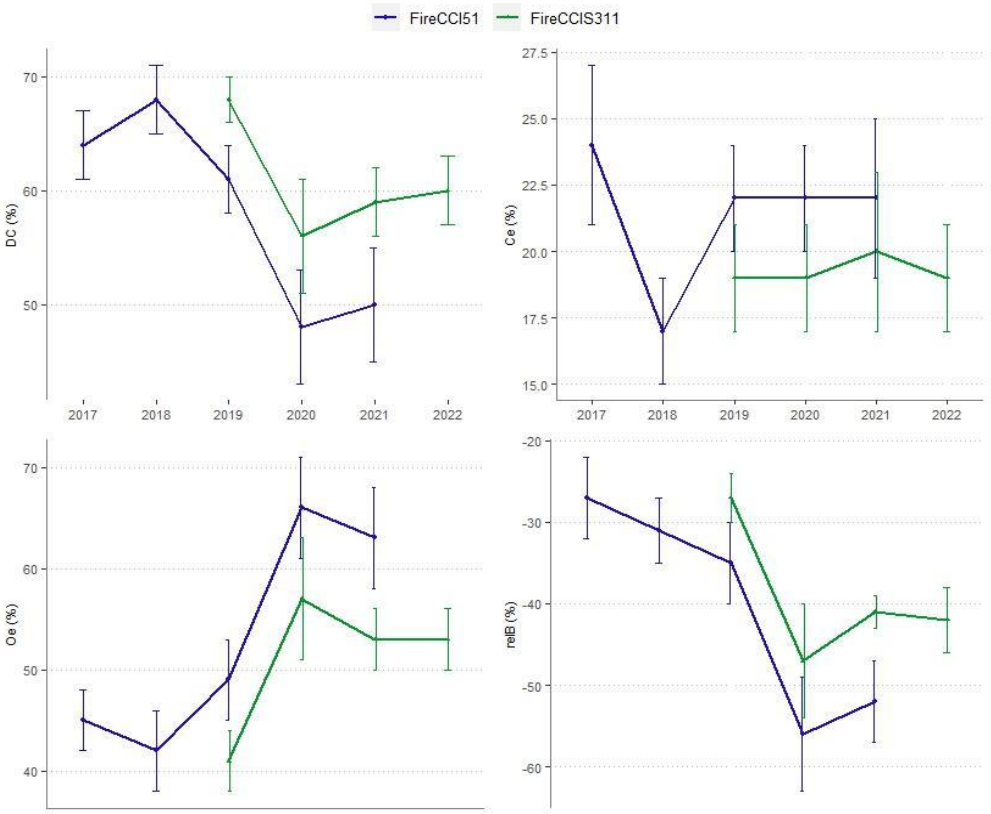

Figure 13. Accuracy estimates (%) for each calendar year for the period 2017-2021 for FireCCI51 and 2019-2022 for FireCCIS311.

The estimated accuracy metrics for each BA product and biome for each year are detailed in Table 7 and Table 8. For the case of FireCCI51, the metrics per biome are observed from 2017 to 2021 (Figure 14) and for FireCCIS311 from 2019 to 2022 (Figure 15).

For the FireCCI51 product, 2017 showed the highest accuracy in biomes such as Boreal forest, Mediterranean and Temperate savanna. 2018 was the year with the best accuracies for the Temperate forest, Tropical savannah and Tundra biomes. On the other hand, 2020 and 2021 showed the lowest accuracies in all biomes compared to previous years' metrics. The year 2021 had the highest Commission error values of all the series analysed (Figure 14).

For the FireCCIS311 product, 2019 had the best accuracies in Deserts & xeric shrublands, Temperate savanna, Tropical forest, Tropical savanna and Tundra biomes. 2020 was most accurate in Boreal forest, Mediterranean forest and Temperate forest. The years 2021 and 2022 were the years with the lowest accuracies and the highest omission and commission values compared to 2019 and 2020 (Figure 15).

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).

Table 7. Biome accuracy estimates (%) for each validation year (2017-2019) for FireCCI51 and FireCCIS311. Standard error (±) is shown in SE.

Biome	Metrics	2017		2018		2019			
		FireCCI51		FireCCI51		FireCCI51		FireCCIS311	
		value	SE	value	SE	value	SE	value	SE
Boreal Forest	DC	83.70	2.40	78.30	2.30	79.00	1.60	78.89	2.44
	Ce	17.40	0.10	19.50	1.80	23.20	3.30	21.38	3.40
	Oe	15.20	5.00	23.80	4.20	18.80	1.90	20.85	5.94
	relB	2.70	6.10	-5.30	5.90	5.70	5.90	0.68	10.64
Deserts & Xeric shrublands	DC	57.50	7.90	61.10	6.20	52.50	7.70	76.66	2.60
	Ce	39.80	12.80	13.50	2.60	20.10	0.90	15.13	0.79
	Oe	44.90	7.80	52.70	7.50	60.90	8.40	30.11	4.76
	relB	-8.50	1.20	-45.30	9.20	-51.00	10.10	-17.65	6.26
Mediterranean	DC	83.60	2.00	24.90	15.20	69.50	6.20	83.37	4.65
	Ce	19.30	1.70	32.10	3.50	23.70	6.90	17.04	1.83
	Oe	13.30	2.50	84.40	11.50	36.20	5.60	16.23	7.63
	relB	7.40	0.90	-77.60	17.50	-16.30	0.70	0.98	7.16
Temperate forest	DC	54.00	7.60	84.50	3.10	59.60	9.80	67.47	10.44
	Ce	51.60	11.80	7.70	1.20	22.30	10.40	21.45	12.59
	Oe	38.90	3.40	22.00	4.60	51.70	9.00	40.87	9.02
	relB	26.30	31.20	-15.50	4.20	-37.80	4.70	-24.73	2.94
Temperate savanna	DC	86.80	2.90	80.00	5.00	62.00	1.90	67.58	1.99
	Ce	14.10	3.30	11.60	2.40	19.60	1.60	17.88	0.88
	Oe	12.30	3.30	26.90	6.70	49.60	2.40	42.59	2.61
	relB	2.10	3.80	-17.40	5.40	-37.30	3.00	-30.08	2.85
Tropical forest	DC	46.70	5.10	39.20	3.50	50.70	8.00	57.77	6.60
	Ce	22.00	2.70	34.00	3.90	24.10	6.10	27.64	5.36
	Oe	66.60	4.80	72.20	3.10	61.90	7.70	51.93	7.08
	relB	-57.20	5.10	-57.90	4.20	-49.80	7.30	-33.57	6.31
Tropical savanna	DC	70.00	2.10	73.80	2.10	68.90	2.90	72.03	2.50
	Ce	20.60	2.60	15.10	1.00	19.40	1.70	15.50	1.34
	Oe	37.40	2.80	34.80	3.00	39.90	3.80	37.24	3.22
	relB	-21.20	4.00	-23.20	3.20	-25.40	4.00	-25.73	3.08
Tundra	DC	48.80	12.50	79.60	0.20	69.90	9.30	81.81	0.93
	Ce	39.90	16.30	18.60	1.30	26.90	12.70	11.85	5.73
	Oe	59.00	10.10	22.20	0.80	33.10	6.40	23.67	3.07
	relB	-31.80	1.60	-4.40	2.50	-8.60	7.10	-13.42	9.02

Table 8. Biome accuracy estimates (%) for each validation year (2020-2022) for FireCCI51 and FireCCIS311. Standard error (±) is shown in SE.

Biome	Metrics	2020				2021				2022	
		FireCCI51		FireCCIS311		FireCCI51		FireCCIS311		FireCCIS311	
		value	SE	value	SE	value	SE	value	SE	value	SE
Boreal Forest	DC	64.47	5.90	83.33	2.69	48.20	12.78	75.94	6.00	32.68	12.42
	Ce	7.63	0.91	12.60	1.27	48.30	15.24	19.07	3.50	34.54	5.23
	Oe	50.49	6.75	20.37	4.00	54.87	15.48	28.48	8.89	78.22	11.40
	relB	-46.40	6.88	-8.89	3.56	-12.71	31.76	-11.62	9.44	-66.73	19.28
Deserts & Xeric shrublands	DC	16.31	2.19	61.60	16.82	59.99	7.63	72.07	11.07	75.56	2.63
	Ce	26.58	4.53	20.43	6.77	16.15	3.26	13.92	2.07	12.99	3.77
	Oe	90.83	1.40	49.75	19.73	53.30	8.39	38.01	15.51	33.22	2.77
	relB	-87.51	2.22	-36.85	19.53	-44.31	8.32	-28.00	16.67	-23.26	3.72
Mediterranean	DC	59.79	15.27	85.17	3.40	16.93	4.91	24.18	7.02	40.36	16.03
	Ce	34.44	10.30	12.41	4.10	75.70	9.95	43.50	1.52	63.98	19.72
	Oe	45.04	18.57	17.12	2.95	87.01	8.63	84.62	5.57	54.10	14.03
	relB	-16.16	15.15	-5.38	2.09	-46.53	57.40	-72.78	9.12	27.43	58.16
Temperate forest	DC	71.99	8.57	74.75	10.87	30.24	18.94	44.47	11.59	29.40	6.71
	Ce	32.44	0.77	21.64	5.10	52.52	28.06	27.78	9.23	54.63	6.84
	Oe	22.95	20.16	28.54	15.78	77.82	14.91	67.88	10.33	78.25	6.53
	relB	14.04	30.58	-8.81	14.57	-53.28	16.51	-55.52	8.97	-52.06	13.09
Temperate savanna	DC	56.82	7.36	60.12	7.49	25.11	14.60	42.52	9.53	47.25	9.90
	Ce	24.19	7.52	19.99	6.19	36.73	5.34	26.81	5.69	27.38	3.31
	Oe	54.56	6.84	51.85	7.42	84.34	11.37	70.03	8.55	64.99	11.04
	relB	-40.07	4.05	-39.83	4.90	-75.25	18.16	-59.05	8.72	-51.78	15.91
Tropical forest	DC	21.84	6.75	34.99	9.06	27.51	3.88	42.30	1.53	40.95	7.05
	Ce	27.44	4.32	28.57	4.34	41.20	2.84	36.83	1.59	35.00	5.14
	Oe	87.14	4.64	76.83	7.91	82.05	3.08	68.20	1.46	70.11	7.01
	relB	-82.28	6.17	-67.56	11.16	-69.46	4.08	-49.66	1.86	-54.01	9.91
Tropical savanna	DC	60.53	3.13	64.75	2.78	61.34	3.18	66.56	2.45	67.50	2.07
	Ce	18.54	1.72	15.39	1.39	14.93	1.62	13.69	1.21	12.50	1.40
	Oe	51.84	3.65	47.56	3.21	52.04	3.59	45.83	2.94	45.05	2.52
	relB	-40.89	4.07	-38.02	3.03	-43.63	3.74	-37.24	2.97	-37.21	2.72
Tundra	DC	51.51	24.22	71.46	5.09	47.43	0.91	80.95	0.00	70.46	9.26
	Ce	43.14	0.00	22.24	5.28	46.47	2.31	11.20	0.00	32.22	13.84
	Oe	52.92	40.46	33.90	12.52	57.43	0.00	25.63	0.00	26.64	5.87
	relB	-17.20	71.16	-15.00	21.88	-20.48	3.43	-16.26	0.00	8.24	18.75

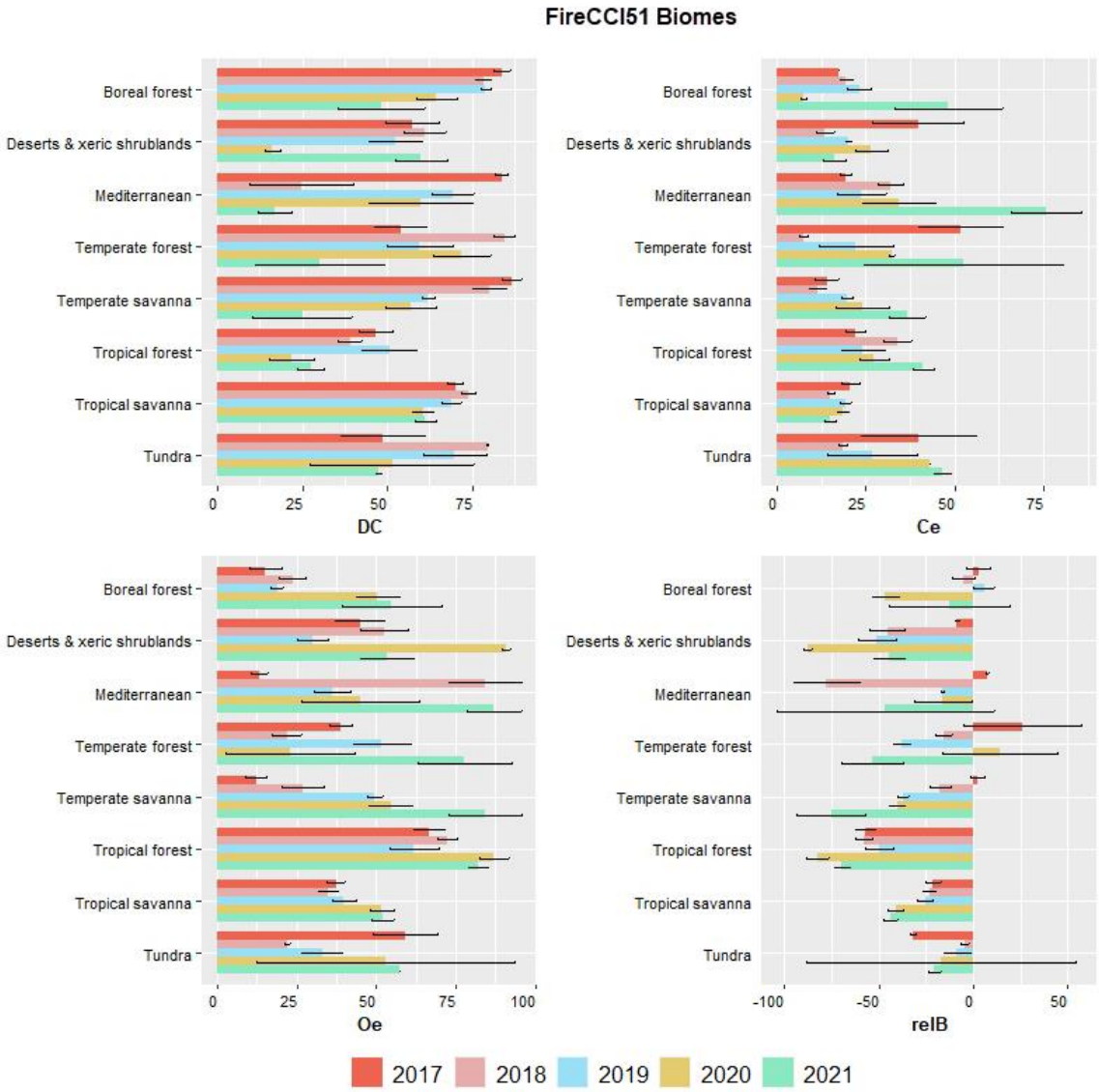

Figure 14. Estimated accuracy (%) of FireCCI51 product for the years 2017- 2021 by biome. 95% confidence intervals are shown with its respective standard errors as segments (black lines). The following metrics are shown: Dice coefficient (DC) (top left), Commission error (Oe) (top right), omission error (Oe) (bottom left) and relative bias (relB) (bottom right).

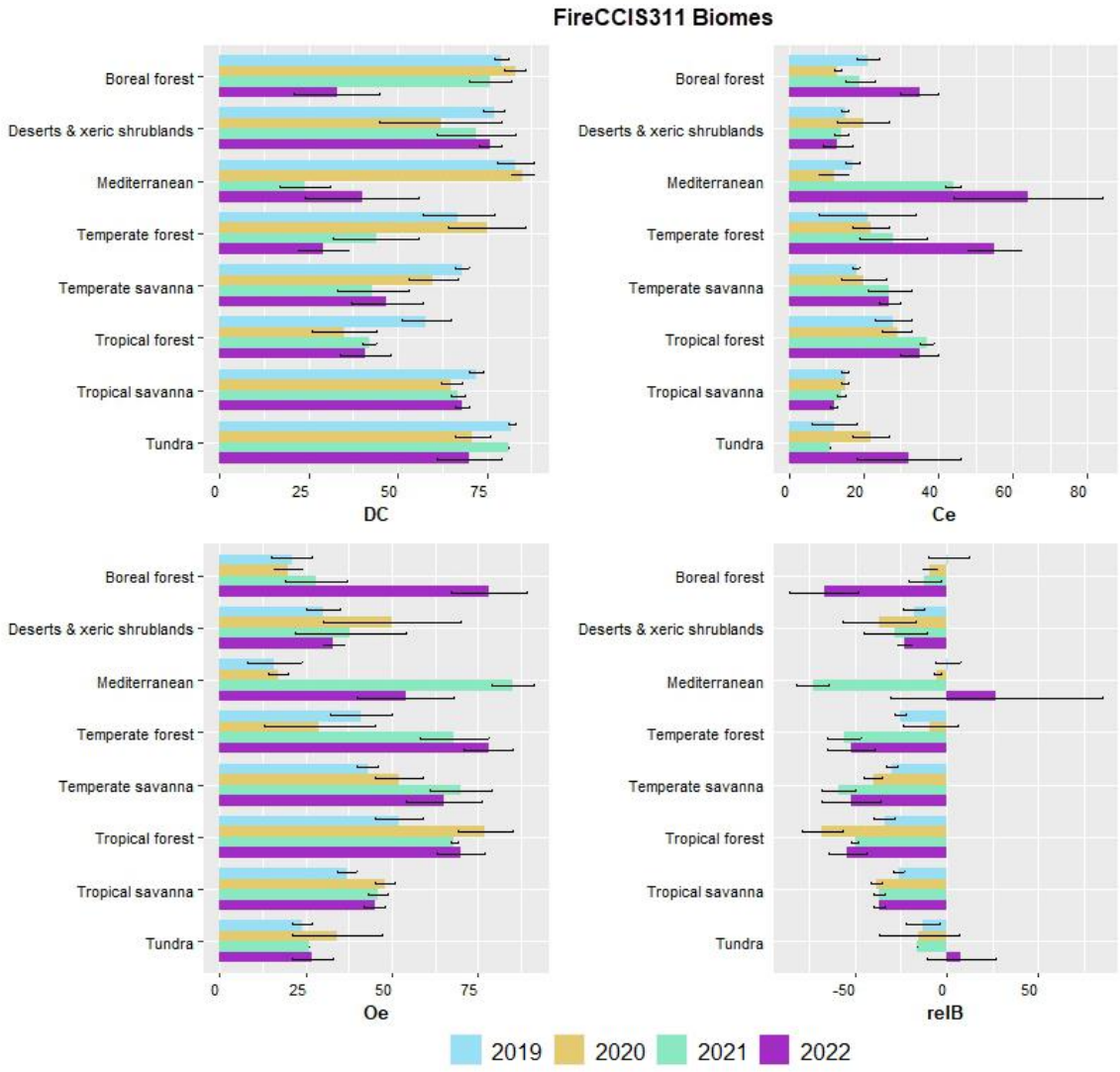

Figure 15. Estimated accuracy (%) of FireCCIS311 product for the years 2019- 2022 by biome. 95% confidence intervals are shown with its respective standard errors as segments (black lines). The following metrics are shown: Dice coefficient (DC) (top left), Commission error (Oe) (top right), omission error (Oe) (bottom left) and relative bias (relB) (bottom right).

3.1.9 Results of accuracy metrics 2021 for FireCCI51 and FireCCIS311

The population estimates of precision for the year 2021 (which is the last year for which data are available to compare both BA products) are presented in Figure 16, where FireCCIS311 has the highest DC value of the two products among all biomes. The highest accuracies in the FireCCIS311 product were achieved in the Tundra (81%) and Boreal Forest (76%) biomes. In the case of FireCCI51, the highest accuracies were in Tropical savanna and Xeric deserts and shrublands, with values of 60% and 61%, respectively.

The 2021 comparison between the FireCCI51 and FireCCIS311 products shows that Ce values were similar in both products in biomes such as Tropical savanna (14.9% in FireCCI51 and 13.69% in FireCCIS311) and Tropical forest (41.36% in FireCCI51 and 36.83% in FireCCIS311). For FireCCI51, the highest commission value was in the Mediterranean biome, with 75.7%, and the lowest in the Tropical savanna, with 14.9%.

For FireCCIS311, the highest commission value was in the Mediterranean biome, with 43.5%, and the lowest in the tundra, with 11%. As for the omission values, the Mediterranean and tropical savanna biomes presented similar values in the two global products.

In terms of relative bias (RelB), underestimation of BA (negative values) were observed in all biomes for both products.

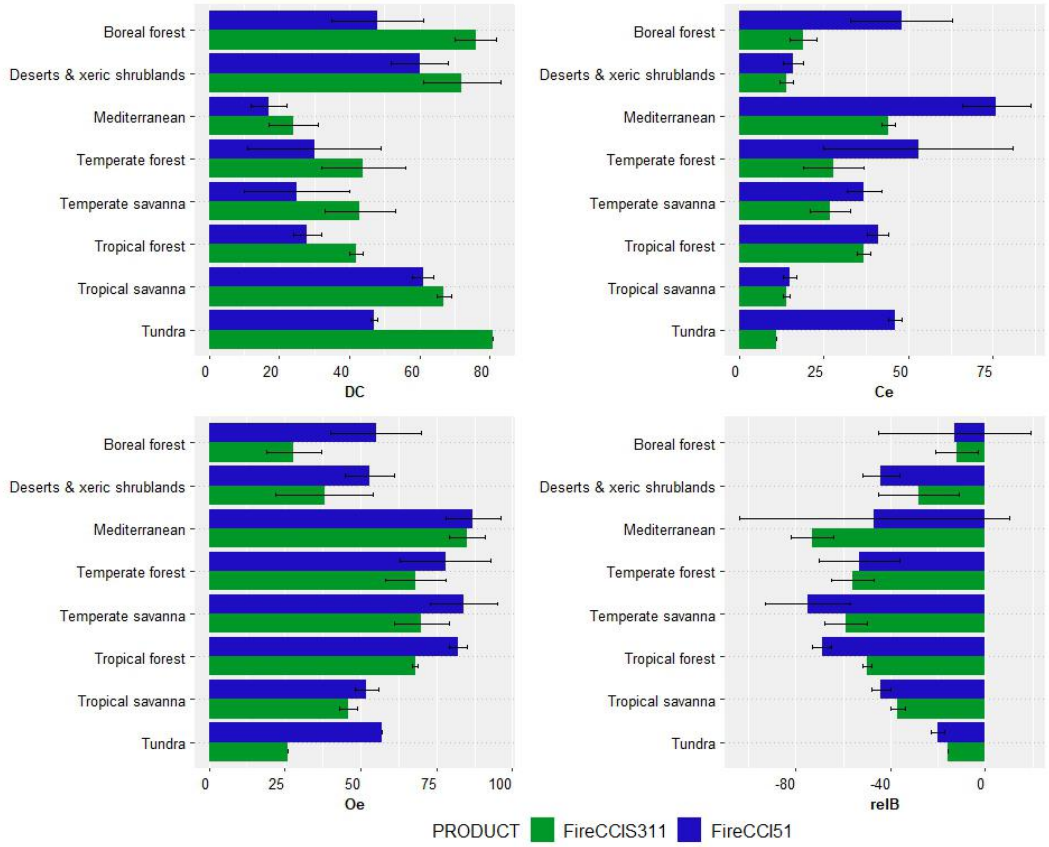

Figure 16. Estimated accuracy (%) of FireCCI51 and FireCCIS311 by biome for the year 2021. 95% confidence intervals are shown with its respective standard errors as segments (black lines). Dice coefficient (DC) (top left), Commission error (Oe) (top right), omission error (Oe) (bottom left) and relative bias (relB) (bottom right).

3.2 Validation of regional BA products

The validation of regional products focused on the assessment of the FireCCISFDL BA product over the three Regions Of Interest (ROIs): Africa-Sahel (AF), South America-Amazonia (SA) and Russia-Siberia (SI) (Figure 17). These datasets of areas affected by fires are derived from Landsat data (Khairoun and Solano, 2023) and they cover a multi-annual period (1990-2019).

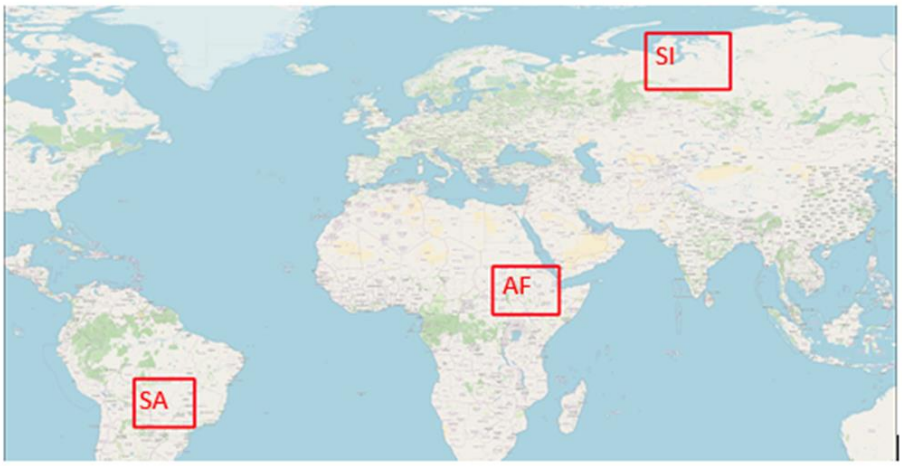

Figure 17. The three ROIs where the multi annual FireCCISFDL product has been generated and validated.

Fire reference perimeters were derived by combining Sentinel-2 data (2019), Landsat 8 (2013-2018) and Landsat 5 (1990-2012) to cover the entire period of the FireCCISFDL BA product and validation units were selected by stratified random sampling. Only Level 2 surface reflectance data products were considered hence source imagery relied on data availability within the Google Earth Engine catalogue of satellite images (https://developers.google.com/earth-engine/datasets). We decided not to rely on Landsat 7 ETM+ data due to the failure of the Scan Line Corrector (SLC) after May 31, 2003 that produced data gaps in the images and significantly reduced the frame area covered by reliable image data.

For the stratified random sampling we considered the biomes within the ROIs extracted from the global Ecoregions 2017 map (Dinerstein et al., 2017) that was aggregated into eight major biomes according to Franquesa et al. (2022) (Figure 3.).

The generation of the fire reference perimeters followed the approach of using validation long units (Franquesa et al., 2020; Franquesa et al., 2022, Stroppiana et al., 2022b) that can reduce the bias in accuracy metrics due to date mismatch between burn detection in the product and reference.

3.2.1 Spatial definition of regional validation units

Validation units represent the region where the BA product and the fire reference perimeters are compared to derive the confusion matrix and the accuracy metrics (Table 4 and Table 5). Each unit covers an area of ~10 000 km2 and is selected by stratified random sampling.

Source data in the validation of the regional FireCCISFDL BA product are mainly provided by Landsat missions, hence the sampling units were spatially defined from the Worldwide Reference System (WRS) of path/row Landsat frames. Since validation units are fixed areas throughout the years regardless the source imagery used for generating fire reference perimeters, only WRS-2 frames fully included in a single Sentinel-2 orbit were retained (*i.e.*, frames encompassing different Sentinel-2 orbits were discarded) as shown in Figure 18 for the AF ROI.

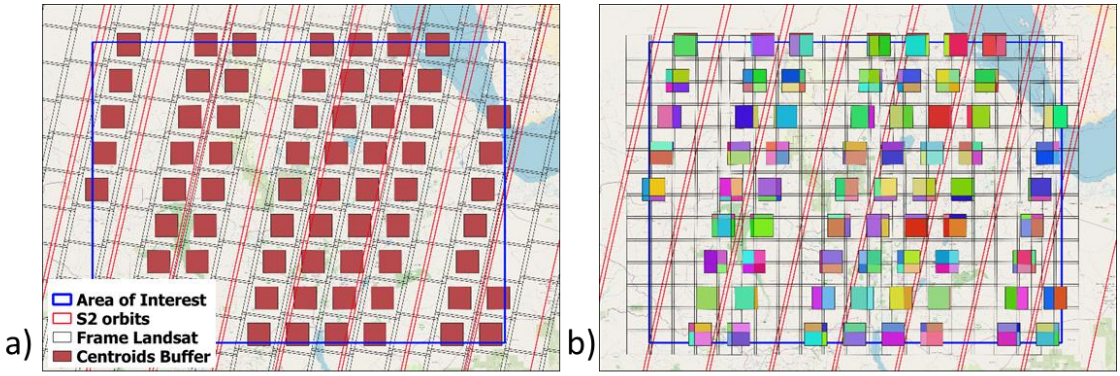

Figure 18. The validation units for the Africa-Sahel (AF) ROI (*Region of Interest*) that are coincident with RWS-2 Landsat frames (dark red squares) (a) and the corresponding S2 tiles mosaicked to cover each validation units for the year 2019 (b).

The area of the validation unit was identified with a buffer of 100 km x 100 km around the centroids of Landsat frames. In the case of S2 data (only year 2019), different tiles were mosaicked to cover the validation unit; this step involved also mosaicking the Scene Classification Layer (SCL) to compute cloud cover percentage for the generation of the time series. Figure 18 shows sampling units available for the Africa-Sahel region (a) and the corresponding coverage from S2 tiles highlighted by the different colours (b); when S2 images are mosaicked, data is provided by the same S2 orbit hence the mosaic preserves the same date of acquisition.

This spatial definition led to a total of 387 validation units distributed in the three ROIs as follows: 68 AF, 50 SA, 269 SI. These units identified the areas suitable for product-to-reference comparison, which are fixed across the years. Over these units, long units were defined as described below.

3.2.2 Temporal definition of regional validation units

In the case of validation of regional products, reference long units are composed of time series of Landsat (L8, L5) and S2 images for each validation unit defined based on threshold applied to the following parameters:

1. **Cloud cover**: scene cloud cover (CC<10%) and cumulated cloud cover (CC_{cum}<30%)
2. **Time interval** between consecutive images (Δt<16 days)
3. **Length** of the time series (L>60 days)

Maximum cloud cover and **maximum cumulated cloud cover** thresholds were set as for the global validation (Section 3.1.2). The **time interval** between consecutive images was set to **16 days and 32 days** for AF/SA and SI, respectively. The greater time interval for SI ROI was decided due to the reduced clear sky image availability (greater cloud cover) and supported by the expected more persistent burned area signal in boreal forests. After preliminary analysis of image availability over the three ROIs, the minimum **length** of the long unit was set to **60 days**.

By applying the above thresholds over the validation unit's area, we identify a total of 1516 long temporal reference units (AF=806, SA=376, SI=334) distributed across the aggregated biomes as summarized in Table 9.

Table 9. The number of long reference validation units distributed across the biomes for the three ROIs.

Biome	N reference long units
Boreal_Forest	144
Deserts&Shrublands	102
Temperate_Savanna	98
Tropical Forest	118
Tropical Savanna	864
Tundra	190
Total	**1516**

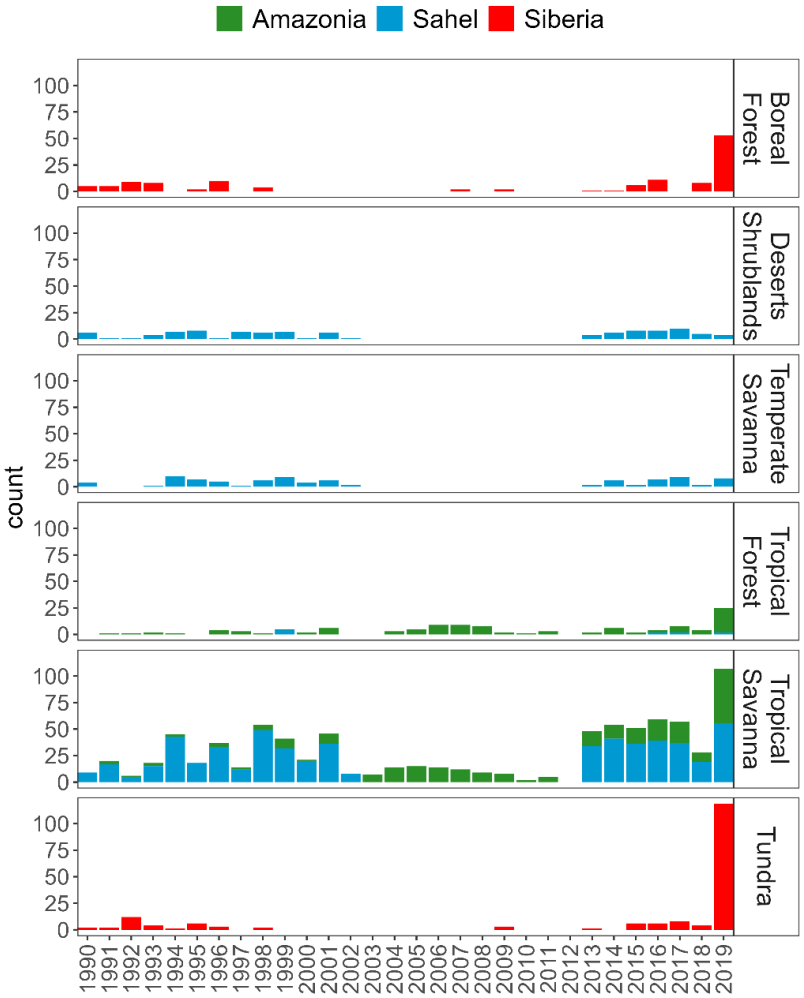

Figure 19. The number of available validation units for the three ROIs and the biomes over the years of the period of interest.

Figure 19 shows how the long units are distributed in the three ROIs and biomes. Boreal forest and Tundra are present only in the SI ROI; very few units are available due to cloud cover of the area. Deserts and temperate savanna are located only in AF ROI (Sahel); in this ROI, there is a significant reduced image acquisition and availability in the archives

in the 2000-2010 epoch. Indeed, the amount of Landsat data in the archives changes from year to year, geographically, or among sensors, due to changes in data reception capabilities and critical technical issues in Landsat missions (such as in the case of Landsat 7) (Kovalskyy and Roy, 2013). Gaps in Landsat 5 image archives covering Africa for the first decade of 2000's is due to several factors such as gaps in the distribution of ground receiving stations, limited on-board recording capacity, and/or missing transmission of data from local ground stations to the USGS archives (Wulder et al., 2016) (Figure 20). These factors and the failure of Landsat 7 produced a significant gap in image time series availability for validation.

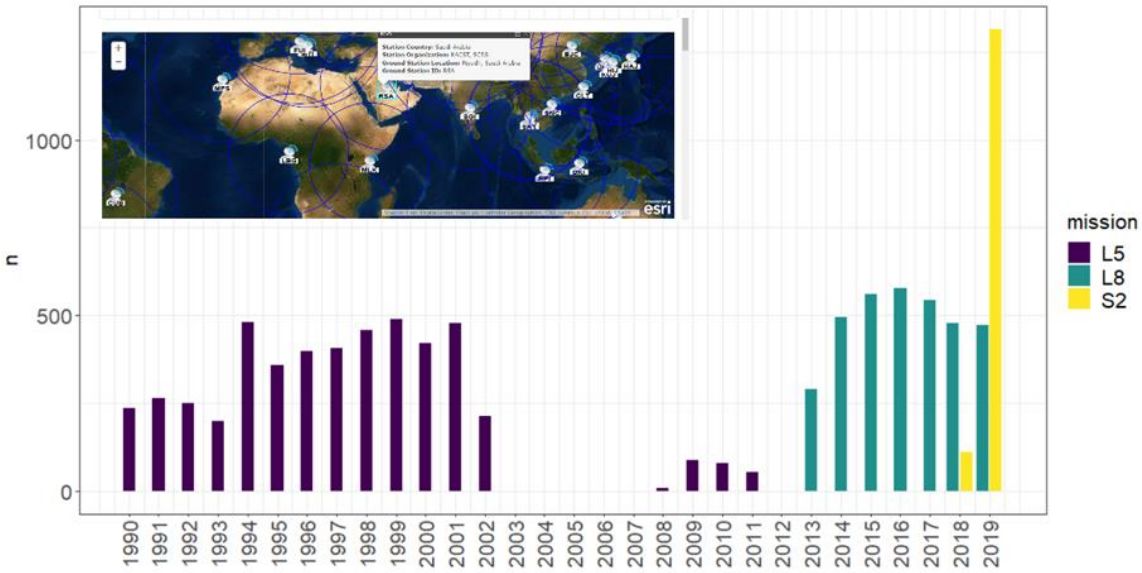

Figure 20. Number of images available for the AF ROI (Africa-Sahel) over the period of interest and for the L5, L8 and S2 missions. In the insert, the location of ground receiving stations.

These figures also highlight the increased availability of validation units for the year 2019 for which we used S2 data as source imagery for generating fire reference perimeters. The greater frequency of observation of S2, compared to Landsat missions, increases the likelihood of satisfying criteria for the definition of the time series to generate cloud free image time series.

3.2.3 Stratification of regional units

As for the validation of global product (Section 3.1.5), stratified random sampling of the validation units was carried out based on two strata; **biomes** derived from the global Ecoregions 2017 map (Dinerstein et al., 2017) (Figure 21) and **fire activity**. High/low fire activity labels were assigned to each unit by applying the same method described for the global validation (Section 3.1.3) with the only difference of the total burned area values used by biome. TotBA for the long units was extracted from the FireCCI51 BA product and for the pre-MODIS era we assumed annual burned area distribution coincident with the multi-annual FireCCI51 average.

In previous FireCCI phases, TotBA was the annual total burned area rather than the total burned area over the long unit time interval; since fires are seasonal phenomena and often concentrated during short season, the use of annual TotBA generated a larger number of

units in the low fire activity class with small or negligible amount of burned surfaces; the new conditions on the contrary produce a more balanced split into high/low fire activity. Table 10 shows the threshold value used for each ROI and biome for the stratification in high or low fire activity.

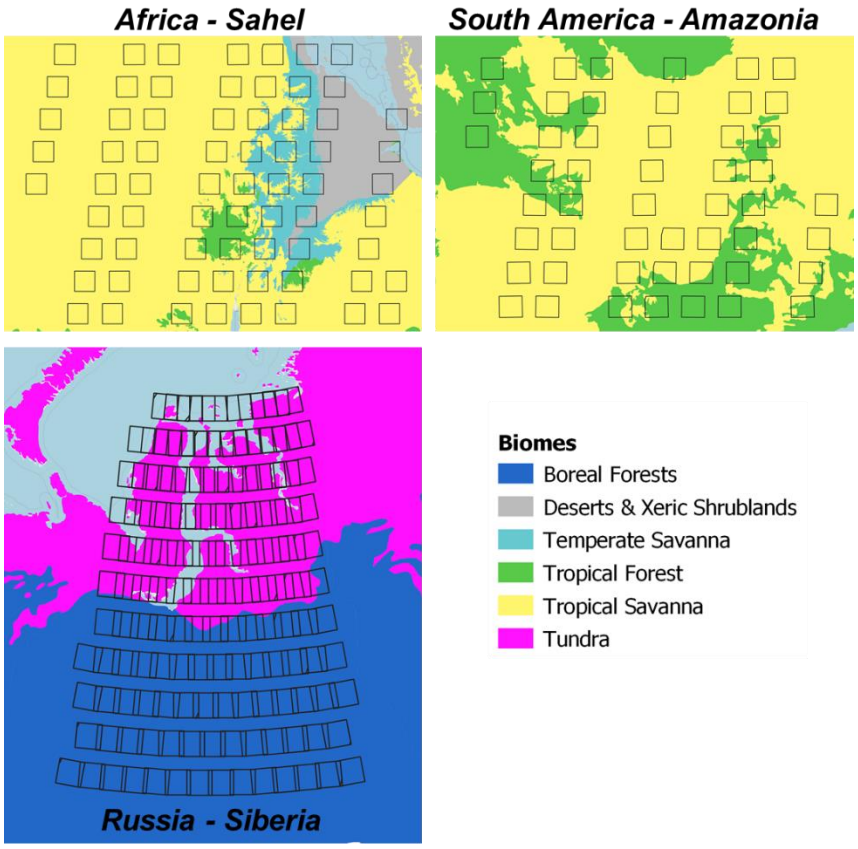

Figure 21. The biome classes covering the three ROIs (biomes are from Dinerstein et al. (2017).

Table 10. The threshold values for each biome used to split into high and low fire activity.

Biome	Th value AF [km2]	Th value SA [km2]	Th value SI [km2]
Boreal Forest	-	-	342
Deserts and Shrublands	5.89	-	-
Temperate Savanna	4381	-	-
Tropical Forest	13770	2543	-
Tropical Savanna	22653	3721	-
Tundra	-	-	164

3.2.4 Sampling validation units

A total of **93 validation** units (Sahel=38, Amazonia=29, Siberia=26) were randomly sampled from the suitable population identified according to criteria listed in Section 3.2.2 (CC<10%, CCcum<30%, Δtmax=16/32 days, Lmin=60 days). An initial set of 25 validation units was later increased to better represent the variability in space and time

(over the period 1990-2019). Reference long unit were distributed among strata in each ROI based on Eq. 1. Cardinality of the sampled units together with the size of the total population is reported in Table 11; notice that all years are pooled together, hence random sampling is applied spatially and temporally.

Table 11. The number of validation units available (top row) and randomly selected (bottom row) in each ROI (numbers are detailed by biome and by Fire Intensity class).

Biome	AF (N_{high} + N_{low})	SA (N_{high} + N_{low})	SI (N_{high} + N_{low})
Boreal Forest	-	-	144 (29+115) **19 (8+11)**
Deserts and Shrublands	102 (8+94) **6 (3+3)**	-	-
Temperate Savanna	98 (20+78) **6 (3+3)**	-	-
Tropical Forest	11 (3+8) **8 (3+5)**	107 (79+28) **9 (5+4)**	-
Tropical Savanna	595 (99+496) **18 (9+9)**	269 (211+58) **20 (10+10)**	-
Tundra	-	-	190 (3+187) **7 (2+5)**
Total	806 (130+676) **38 (18+20)**	376 (86+290) **29 (15+4)**	334 (32+302) **26 (10+16)**

Figure 22 shows the distribution of validation units randomly sampled over the years and ROIs that, as highlighted by the distribution of available long units (Figure 19), has the greatest proportion in the year 2019 when S2 data are available. Indeed, the greater frequency of observation of S2, compared to Landsat missions, increases the likelihood of satisfying criteria for the definition of the time series. Figure 23, Figure 24 and Figure 25 show the spatial distribution of the reference long units within the ROIs.

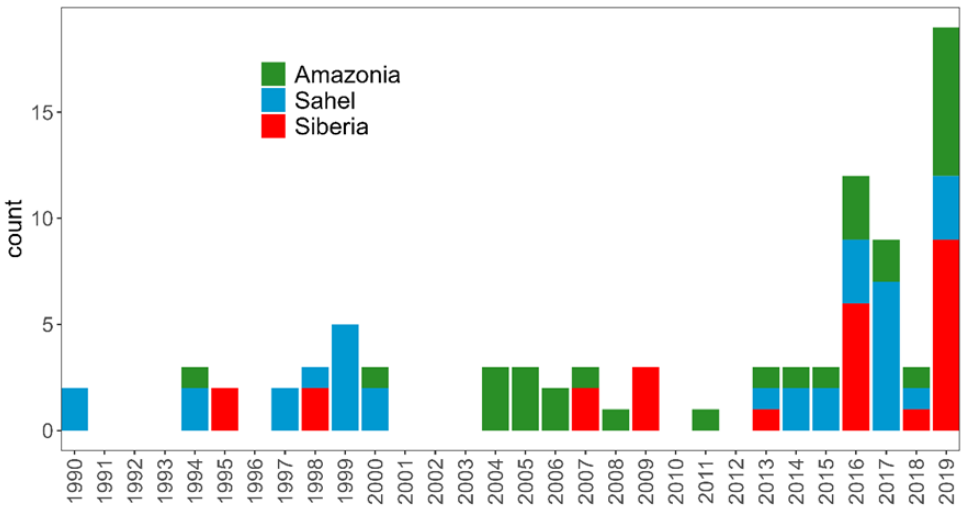

Figure 22. Distribution of the validation units sampled randomly over the three ROIs for each stratum (biome/fire intensity) and over the years.

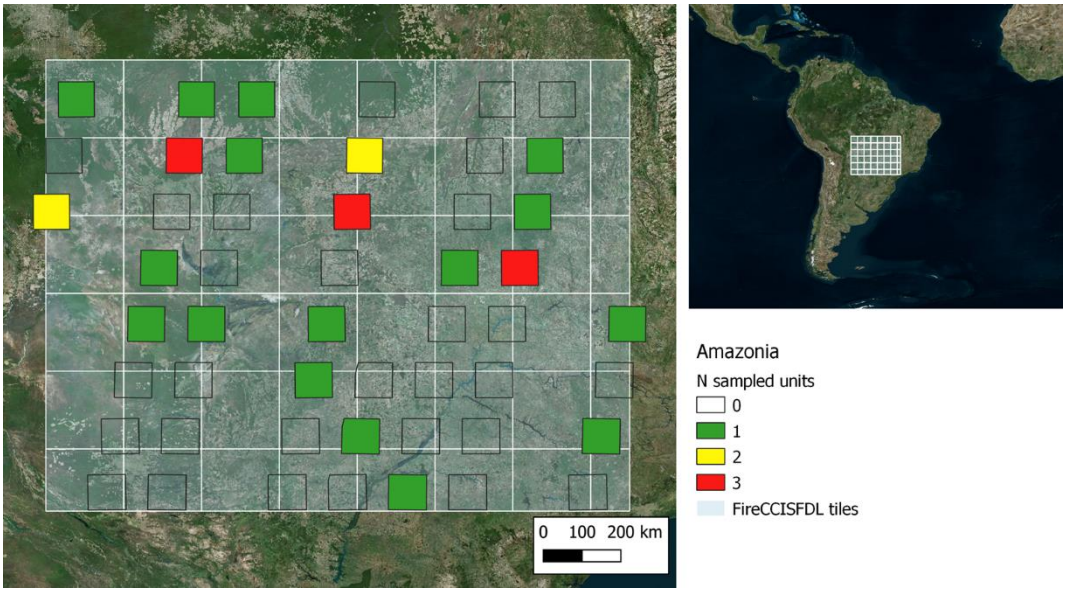

Figure 23. Distribution of the multi-annual validation units in Amazonia ROI, colors represent the number of long reference units within the validation area.

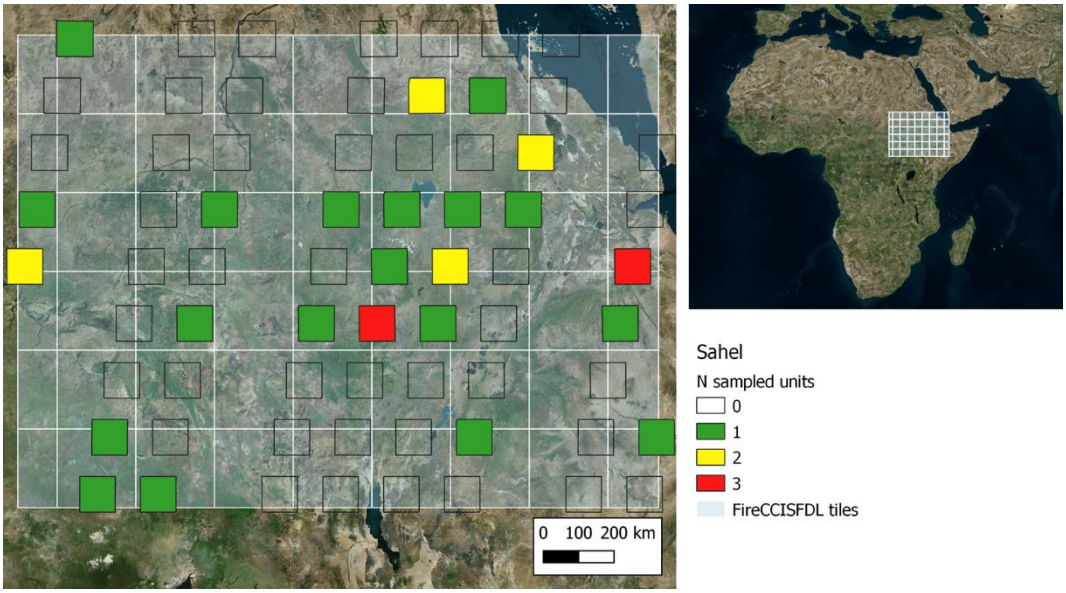

Figure 24. Distribution of the multi-annual validation units in Sahel ROI, colors represent the number of long reference units within the validation area.

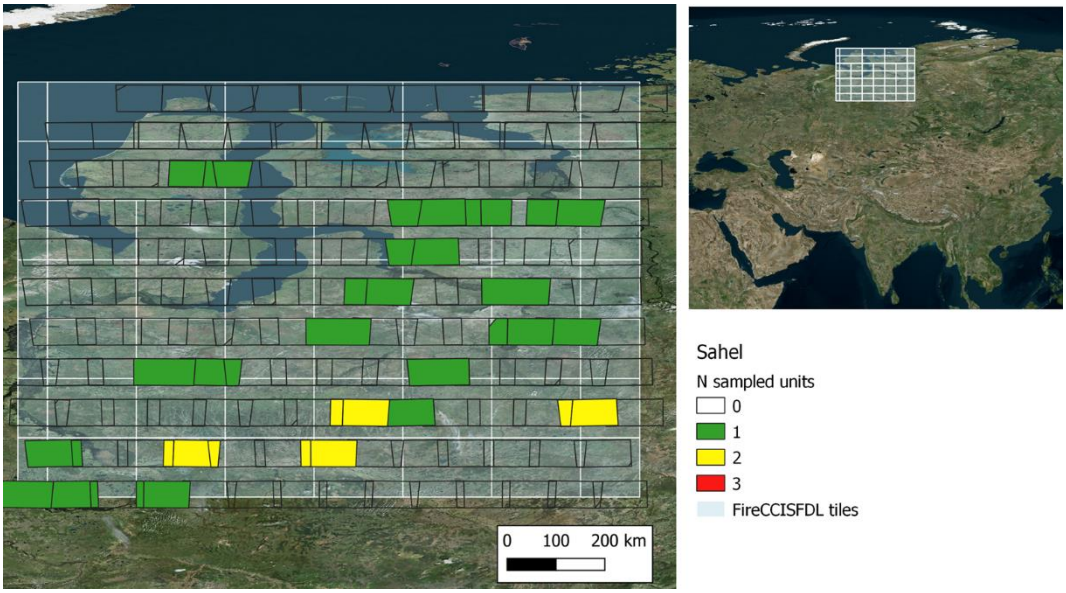

Figure 25. Distribution of the multi-annual validation units in Siberia ROI, colors represent the number of long reference units within the validation area.

3.2.5 Generation of the regional fire reference perimeters

Fire perimeters for the sampled validation units were extracted for short units to map areas burned between the two dates of consecutive Landsat/S2 images (t1, t2). The processing steps are the same as described for the global validation, illustrated in Figure 10 and implemented in GEE. The post-fire date (t2) from each short unit is assigned to each burned detected polygon as burn detection date (Figure 1).

The output of GEE was the short unit classification composed of the following information layers exported as vector shapefile format:

- Burned areas in vector KML format;
- The validation region of 100 km x 100 km;
- Cloud Mask in vector format;
- Training polygons as vector shapefiles;

3.2.6 Comparison of fire reference perimeters to FireCCISFDL BA

The comparison of reference fire perimeters with the FireCCISFDL BA products was carried out for each ROI and validation unit following the scheme depicted in Figure 26.

Key pre-processing steps of the reference to product comparison are: subset of the FireCCISFDL BA product over the extent of the validation unit, extraction of the FireCCISFDL burned pixels for the time period covered by the validation unit (first to last date), applying the non-burnable mask (derived from the CCI High Resolution Land Cover HRLC product, https://climate.esa.int/en/projects/high-resolution-land-cover/) to the reference (Figure 27), converting reference vector file to 30 m x 30 m raster, applying 3x3 filtering to reference.

After these steps were applied, pixel by pixel comparison was carried out to estimate accuracy metrics from the confusion matrix (omission and commission errors, Dice coefficient, Relative bias) (Table 4 and Table 5).

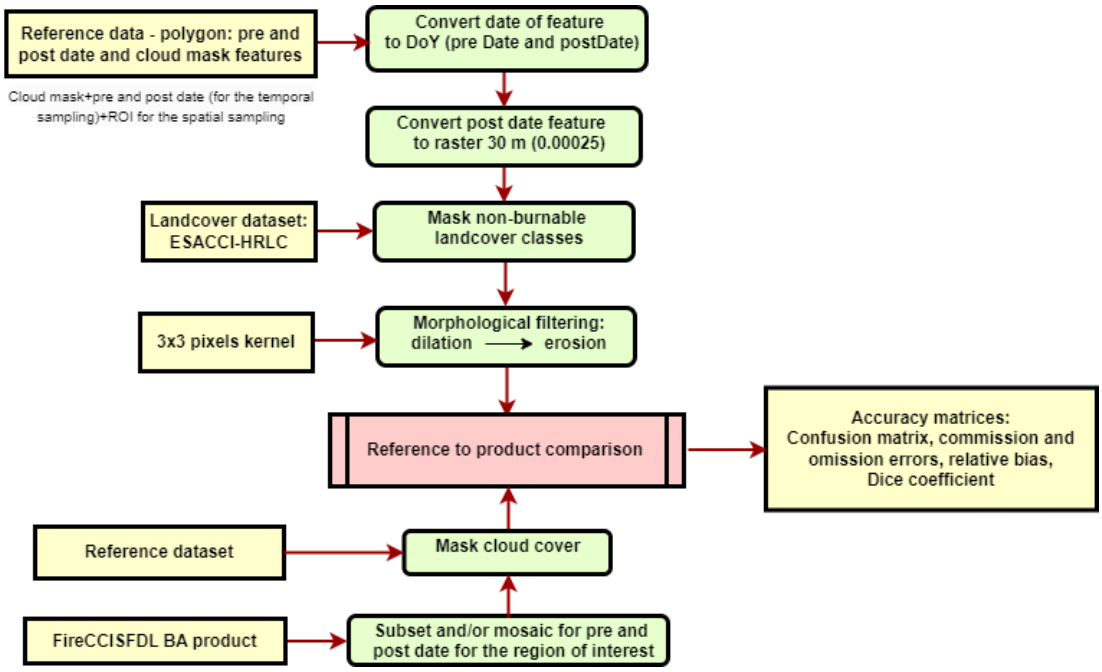

Figure 26. Generation of the fire reference perimeters form the classification of the long units.

Value	LC type description	Burnable
0	No data	
10	Tree cover evergreen broadleaf	x
20	Tree cover evergreen needleleaf	x
30	Tree cover deciduous broadleaf	x
40	Tree cover deciduous needleleaf	x
50	Shrub cover evergreen	x
60	Shrub cover deciduous	x
70	Grasslands	x
80	Croplands	x
90	Woody vegetation aquatic or regularly flooded	x
100	Grassland vegetation aquatic or regularly flooded	x
110	Lichens and mosses	x
120	Bare areas	
130	Built-up	
140	Open water	
141	Open water seasonal	
142	Open water permanent	
150	Permanent snow and/or ice	

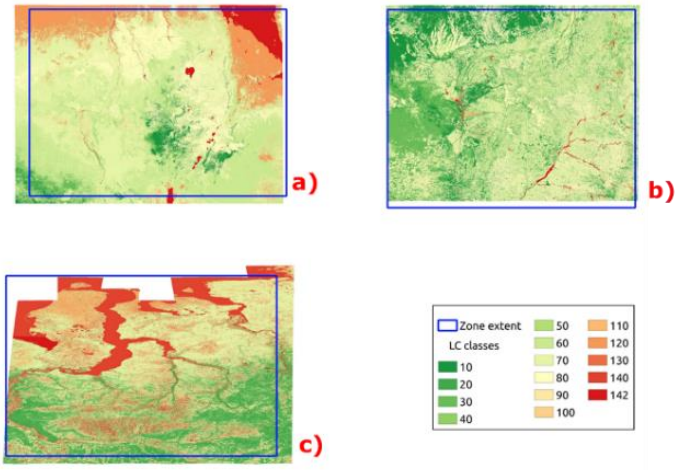

Figure 27. The CCI High Resolution Land Cover product (HRLC) for the three ROIs: a) Sahel, b) Amazonia and c) Siberia. The table shows the classes retained as burnable (all vegetated classes marked as 'x') and the non-burnable classes that were masked out when reference and product were compared.

Besides the confusion matrix and accuracy metrics, the agreement between reference and the FireCCISFDL product was quantified also using 'agreement maps' (showing spatial distribution of the agreement and disagreement areas) and regression analysis of the reference and product proportion of burned area within 5 km x 5 km grid cells (Boschetti et al., 2019). Both these analyses were carried out for each validation unit.

3.2.7 Computation of accuracy metrics for regional FireCCISFDL products

Table 12 shows the overall accuracy metrics estimated from the confusion matrices of the validation units for the three ROIs; values highlight the significantly greater accuracy for

the FireCCISFDL BA product over Amazonia with commission and omission errors both below 10% and DC> 90%. Over Siberia, on the contrary, we obtained the worst accuracy with a significant greater omission error (~89%) and DC<20%. Several factors could be pointed out to explain these results. First, this region of interest is little affected by fires, indeed the proportion of the validation units covered by burned areas according to the reference dataset is below 0.5% (~ 1 km2); among the 26 sampled units, 11 had no burns in the reference data and seven of these units had no burned areas in the FireCCISFDL product either (perfect agreement between reference and product on unburned areas). Moreover, persistent cloud cover can impact the detection capability as well as bias the burn detection date in both reference and product, thus inflating error metrics that in the aim of this validation activity should only depict spatial agreement between reference and product. In Sahel ROI, the metrics obtained were in between the SA and SI ROIs: CE=19.58%, OE=34.82% and DC=72.00%. In all ROIs, omission is greater than commission and relative bias < 0 showing underestimation of burned area.

Table 12. The accuracy metrics (CE=commission error, OE=omission error, DC=Dice coefficient, relB=relative bias) estimated for the three ROIs from the comparison of the reference to FireCIISFDL burned areas. These values are overall values for the entire ROI. N is the number of validation units sampled and processed for each ROI

ROI	CE	OE	DC	relB
Amazonia (N=29)	6.00%	8.36%	92.81%	-2.51%
Siberia (N=26)	54.29%	89.21%	17.45%	-76.40%
Sahel (N=33)	19.58%	34.82%	72.00%	-18.59%

A high variability was, however, observed, in the accuracy metrics for the single validation units as shown in the examples reported in Figure 28, Figure 29 and Figure 30; in these figures results for three example validation units in each ROI are reported as agreement maps and accuracy metrics.

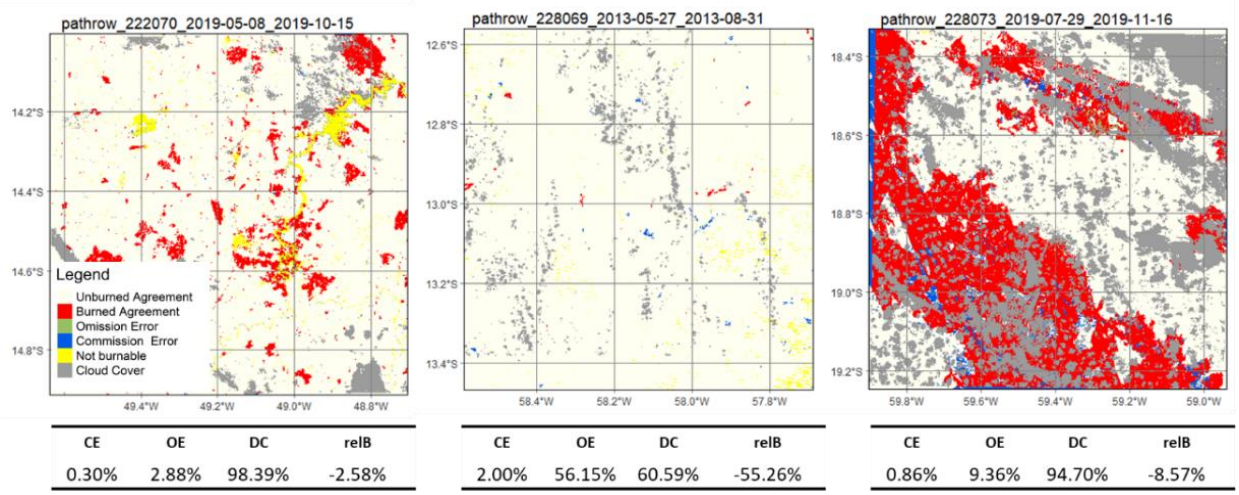

Figure 28. The agreement maps for three example validation unit in Amazonia ROI together with the corresponding estimated accuracy metrics from the confusion matrix.

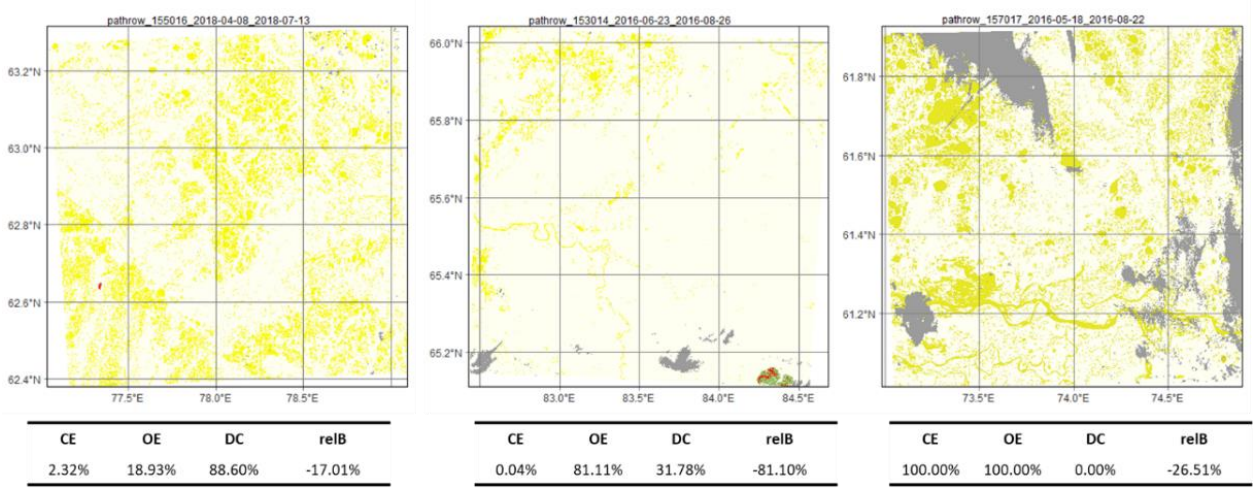

Figure 29. The agreement maps for three example validation unit in Siberia ROI together with the corresponding estimated accuracy metrics from the confusion matrix.

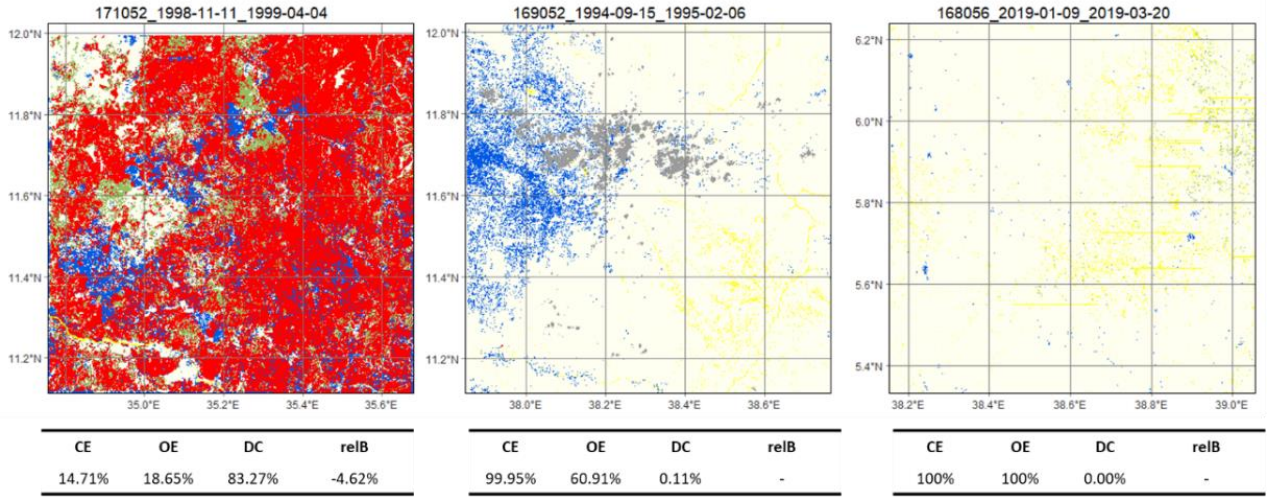

Figure 30. The agreement maps for three example validation unit in Sahel ROI together with the corresponding estimated accuracy metrics from the confusion matrix.

Figure 31 shows the results for the grid regression analysis of the proportion of burned area in 5 km x 5 km cells. The greater agreement for the Amazonia ROI is confirmed with R2=96.6% and RMSE=2.788. The opposite for Siberia, where all cells have a low proportion of area burned and several are below the 1:1 line confirming underestimation of the area burned (R2=32.3%, slope=2.365).

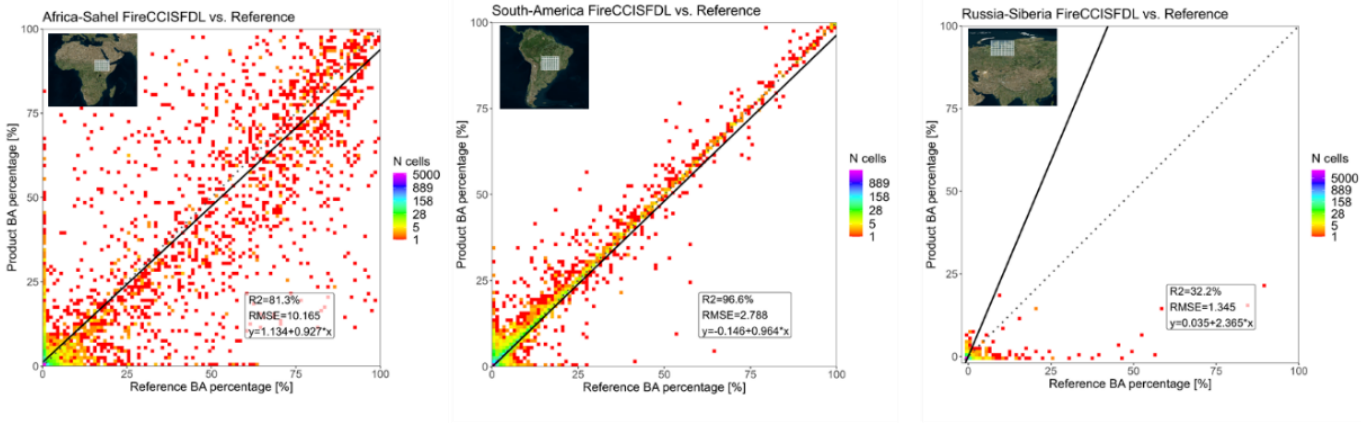

Figure 31. Scatter plots of the proportion of area burned in the reference (x-axis) and in the FireCCISFDL product (y axis) for the three ROIs: a) Sahel, b) Amazonia and c) Siberia. Parameters of the linear regression model are reported in each graph. Scatter plots are produced with a 10% bin size and the marker colors represent the number of points in each bin (number of 5 km x 5 km grid cells).

4 Inter-comparison of global BA products

This section presents a comparison between four global burned area products for the period 2017-2022. The products compared were FireCCI51, FireCCIS311, MCD64A1 c6, and C3SBA11. It is clarified that FireCCI51 has data from 2017 to 2021 and FireCCIS311 has data from 2019 to 2022. The characteristics of each of these products are summarized in Table 13.

Table 13. Summary of BA products characteristics considered in the inter-comparison analysis carried out at global scale.

BA product	Input Data Sensor		Coverage	Spatial resolution		Time Series analysis	Reference
	Surface Reflectance	Active Fires		Pixel	Grid		
FireCCI51	MODIS	MODIS	Global	250 m	0.25 deg	2017-2021	Lizundia-Loiola et al. (2020)
FireCCIS311	Sentinel-3 SYN	VIIRS	Global	300 m	0.25 deg	2019-2022	Lizundia-Loiola et al. (2022)
C3SBA11	Sentinel-3 OLCI	MODIS	Global	300 m	0.25 deg	2017-2022	Lizundia-Loiola et al. (2021)
MCD64A1 C6	MODIS	MODIS	Global	500 m	-	2017-2022	Giglio et al. (2018)

4.1 Burned area detection from global BA products

Figure 32 shows monthly global BA since 2017 of the four BA products analysed; annual totals are summarised in Table 14. Considering the four global medium resolution products (250 to 500 m), the average total BA (Table 14) is highest for FireCCIS311 with 4.7 Mkm^2, followed by FireCCI51 (4.0 Mkm^2), C3SBA11 (3.6 Mkm^2) and MCD64 c6 (3.5 Mkm^2). FireCCIS311 detects 15% more burned area compared to the next highest product, FireCCI51.

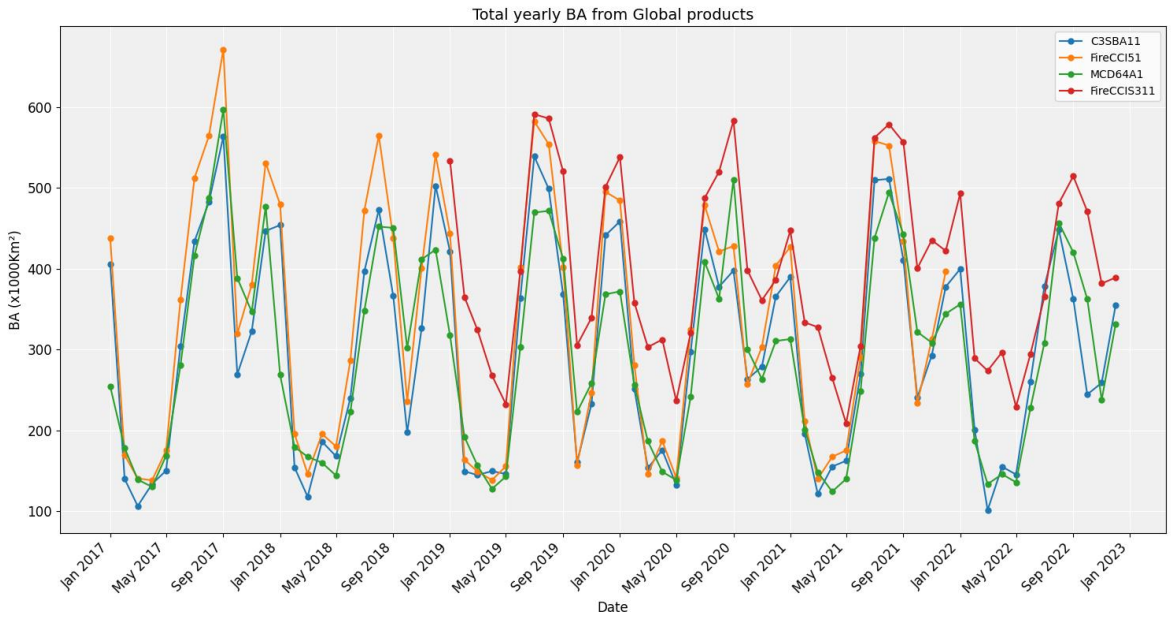

Figure 32. Annual burned area of different global products

Table 14. Total annual BA for each product (in Mkm²).

Year	FireCCI51	FireCCIS311	C3SBA11	MCD64A1 C6
2017	4.4	-	3.8	3.8
2018	4.1	-	3.6	3.5
2019	3.9	4.9	3.6	3.4
2020	3.9	4.8	3.6	3.5
2021	3.9	4.8	3.6	3.5
2022	-	4.4	3.3	3.3
Average	4.0	4.7	3.6	3.5

Figure 33 shows the total annual BA estimated by the four products in the eight major biomes; as expected, the biomes with the largest amount of area burned are Tropical savannas, followed by Tropical forests. While the four products analysed show similar BA in the biomes that have few fires (i.e., Tundra and Mediterranean), some differences are observed in the remaining biomes. As highlighted in Table 14, FireCCIS311 is confirmed to provide the greatest estimates for all years except in Boreal Forests and in 2021 in the Tundra. The most significant difference is in the Tropical Forest biome, where FireCCIS311 detects more burned area (0.53 Mkm^2) than the average of the other three products (0.33 Mkm^2), which present similar values (average calculated for all years where information is available for each product).

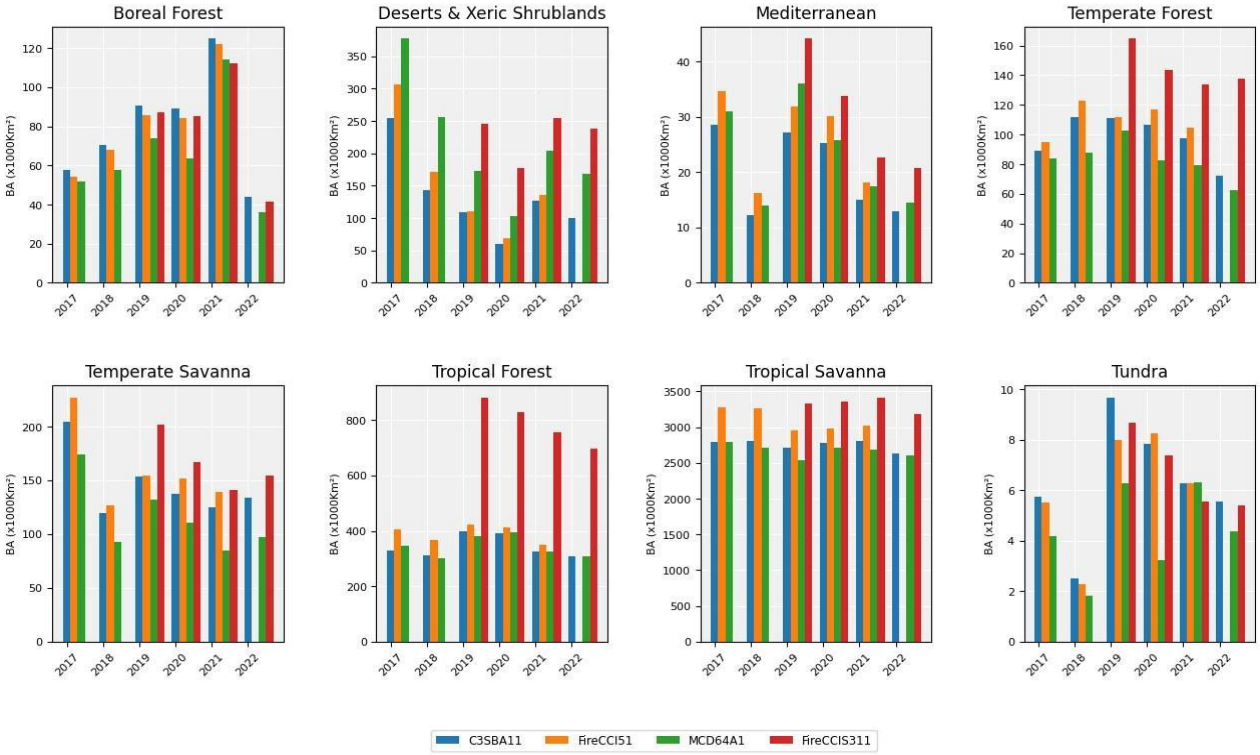

Figure 33. Total annual burned area for years 2017-2022 stratified per biome in different global products.

4.2 Spatial accuracy assessment for global BA products

Figure 34 and Table 15 present a comparison of the accuracy metrics derived for the year-by-year validation of the four global burned area products FireCCI51, FireCCIS311, C3BA11 and MCD64A1 c6. Validation results for 2017 to 2019 are taken from Franquesa et al. (2022) and Lizundia-Loiola et al. (2022). FireCCIS311 has the highest Dice Coefficient in all years where the product is available. Its highest accuracy was achieved in 2019 with a DC of 68.1%, followed by 2022 with 59.5%. This product also has the lowest error of omission of the various products, with 41.2% in 2019 followed by 53% for 2022. FireCCI51 had a systematic higher DC than C3SBA11 and MCD64A1 c6, reaching a value of 69.2% in 2018, while MCD64A1 C6 has the lowest commission error for the years 2017, 2019, 2020 and 2021. C3SBA11 has the lowest commission errors for the year 2018. The negative relB value for all BA products indicates that the products underestimate the area burned as estimated by the reference dataset.

Figure 34. Accuracy metrics (DC=dice coefficient, Ce=commission error, Oe=omission error, relB=relative bias) for each calendar year for the four global products analysed FireCCI51 (yellow), FireCCIS311 (red), C3SBA11 (blue) and MCD64A1 C6 (green).

Table 15. Accuracy metrics (DC=dice coefficient, Ce=commission error, Oe=omission error, relB=relative bias) estimates (%) for the four products with standard error SE. Data for FireCCI51, C3SBA11 and MCD64A1 C6 for the years 2017 to 2019 were extracted from Table 5 of Franquesa et al. (2022), while data for FireCCIS311 for 2019 were extracted from Table 1 of Lizundia-Loiola et al. (2022). In bold the product showing the highest accuracy in each specific metric.

Year	Metrics	FireCCI51		FireCCIS311		MCD64A1 C6		C3SBA11	
		value	SE	value	SE	value	SE	value	SE
2017	**DC**	**66.90**	2.30	-	-	62.20	2.70	62.30	2.60
	CE	21.40	2.20	-	-	21.40	3.10	**19.50**	2.10
	OE	**41.80**	3.10	-	-	48.50	3.30	49.20	3.20
	relB	**-26.00**	4.10	-	-	-34.40	4.10	-36.90	4.00
2018	**DC**	**69.20**	2.70	-	-	65.10	3.50	64.80	2.90
	CE	15.70	1.40	-	-	16.60	1.70	**13.10**	1.30
	OE	**41.30**	3.40	-	-	46.70	4.30	48.30	3.40
	relB	**-30.40**	3.30	-	-	-36.00	4.60	-40.50	3.50
2019	**DC**	63.90	2.80	**68.10**	2.40	59.80	3.20	61.70	2.90
	CE	20.80	1.70	19.20	1.60	**17.50**	1.40	18.60	1.70
	OE	46.50	3.40	**41.20**	2.90	53.10	3.60	50.30	3.40

Year	Metrics	FireCCI51		FireCCIS311		MCD64A1 C6		C3SBA11	
		value	SE	value	SE	value	SE	value	SE
	relB	-32.50	3.40	**-27.10**	2.70	-43.10	3.80	-39.00	3.50
2020	**DC**	47.90	5.00	**56.00**	4.90	45.70	5.40	46.20	5.20
	CE	21.70	1.70	19.10	1.60	**17.10**	1.00	19.40	1.70
	OE	65.50	5.20	**57.20**	5.60	68.50	5.10	67.60	5.20
	relB	-55.90	6.70	**-47.10**	6.80	-62.00	6.10	-59.80	6.50
2021	**DC**	50.40	5.00	**58.90**	3.10	44.40	6.20	48.10	5.30
	CE	21.90	2.80	20.30	3.10	**14.30**	1.00	19.10	2.50
	OE	62.80	4.90	**53.20**	3.40	70.00	5.60	65.80	5.00
	relB	-52.40	4.80	**-41.30**	2.40	-65.00	6.40	-57.70	5.00
2022	**DC**	-	-	**59.50**	3.00	52.30	4.20	56.17	3.70
	CE	-	-	18.80	2.40	25.30	8.00	**17.00**	2.34
	OE	-	-	**53.00**	3.30	59.80	4.10	57.55	9.94
	relB	-	-	**-42.10**	3.50	-46.20	7.50	-48.86	4.37

Figure 35 shows a comparison only for the years 2019 to 2021, which is the period where we have information of the 4 products to compare, highlighting that the FireCCIS311 product has the best accuracies in terms of DC, the minimum errors of omission and the smallest relB, while MCD61A1 c6 has the minimum errors of commission.

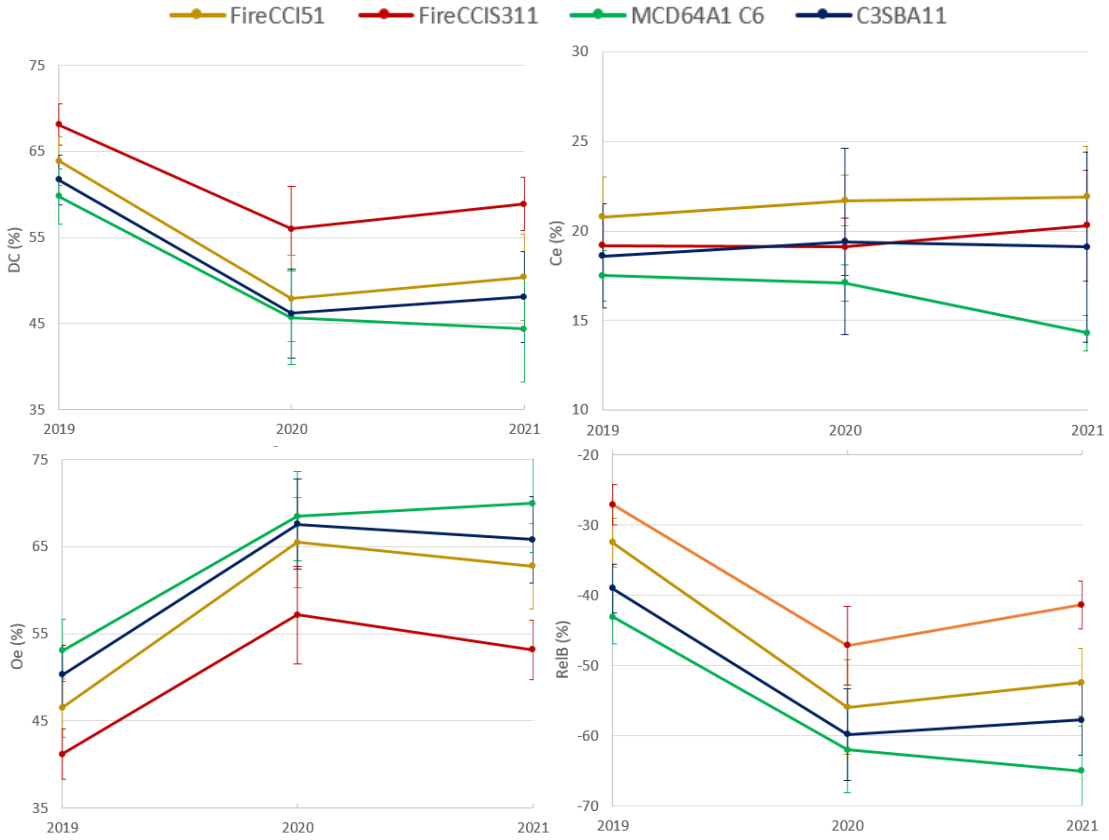

Figure 35. Accuracy estimates for each calendar year in the period 2019-2021 for FireCCI51, FireCCIS311, C3SBA11 and MCD64A1.

4.3 Global product intercomparison of 2022

For the year 2022, the FireCCIS311 product was recently validated. Table 16 and Figure 36 show the results of the metrics of the CCI product and the comparison with MCD64A1 C6 and C3SBA11 products that were also validated for this year.

Overall accuracy results show that FireCCIS311 was the most accurate product in terms of estimated Dice coefficient (DC), omission error (Oe) and relative bias (relB). The DC was 59.5% for FireCCIS311 and similar between MCD64A1 c6 and C3SBA11, with values of 52.3% and 56.17%, respectively. C3SBA11 had the lowest Ce value of 17%, followed by FireCCIS311 at 18.8% and MCD64A1 c6 at 25.3%. FireCCIS311 showed an Oe of 53.0%, C3SBA11 of 57.5% and the highest Oe was in MCD64A1 c6 with 59.8%. Oe was higher than Ce in all three products, reflecting an underestimation of the area burned. This relationship is also evident in the negative relB values that ranged from -42.1% for FireCCIS311 to -46.86% for C3SBA11.

Table 16. Overall error estimates (%) for FireCCIS311, MCD64A1 C6 and C3SBA11 for the year 2022 with standard error (SE). In bold the product showing the highest accuracy in each specific metric.

Year	Metrics	FireCCIS311		MCD64A1 C6		C3SBA11	
		value	SE	value	SE	value	SE
2022	**DC**	**59.50**	3.00	52.30	4.20	56.17	3.70
	CE	18.80	2.40	25.30	8.00	**17.00**	2.34
	OE	**53.00**	3.30	59.80	4.10	57.55	9.94
	relB	**-42.10**	3.50	-46.20	7.50	-48.86	4.37

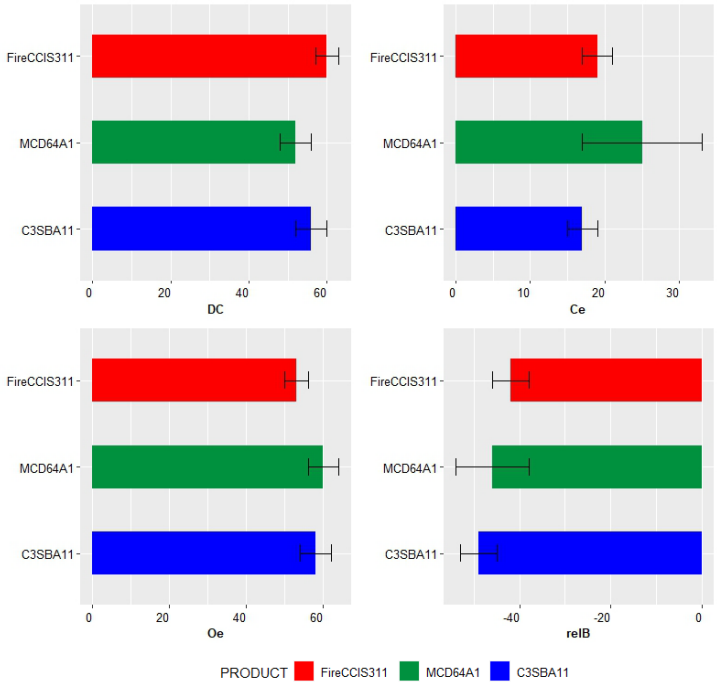

Figure 36. Estimated accuracy of each product for the year 2022. 95% confidence intervals are shown with its respective standard errors as segments (black lines). Each quadrant shows: Dice coefficient (DC) (top left), Commission error (Oe) (top right), omission error (Oe) (bottom left) and relative bias (relB) (bottom right).

Population estimates of precision by biome are presented in Table 17 and Figure 37 for the three BA products. As Figure 37 shows, FireCCIS311 has the highest DC value of the three products in biomes such as Deserts and xeric shrubland (75.6%), Tropical forest (40.9%), Tropical savanna (67.5%) and Tundra (70.5%). In the case of C3SBA11 the highest accuracies were in Boreal forest with 68.9%, Mediterranean with 41.9% and Temperate savanna with 62.1%. On the other hand, MCD64A1 C6 presented intermediate values with respect to the other products in most biomes.

Regarding the commission error values, FireCCIS311 presented the lowest values in the biomes Deserts and xeric shrubland (12.9%), Mediterranean (63.9%), Tropical savanna (12.5%) and Tundra (32.2%). In terms of omission errors, MCD64A1 C6 presented the highest values in most biomes, obtaining the highest value globally in this parameter. The Temperate forest, Tropical forest and Tropical savanna biomes presented similar values in the three global products.

In terms of relative bias (relB), underestimates (negative values) were observed in all biomes except Tundra, which showed overestimates in all products, and the Mediterranean, which showed overestimates for the products FireCCIS311 and C3SBA11.

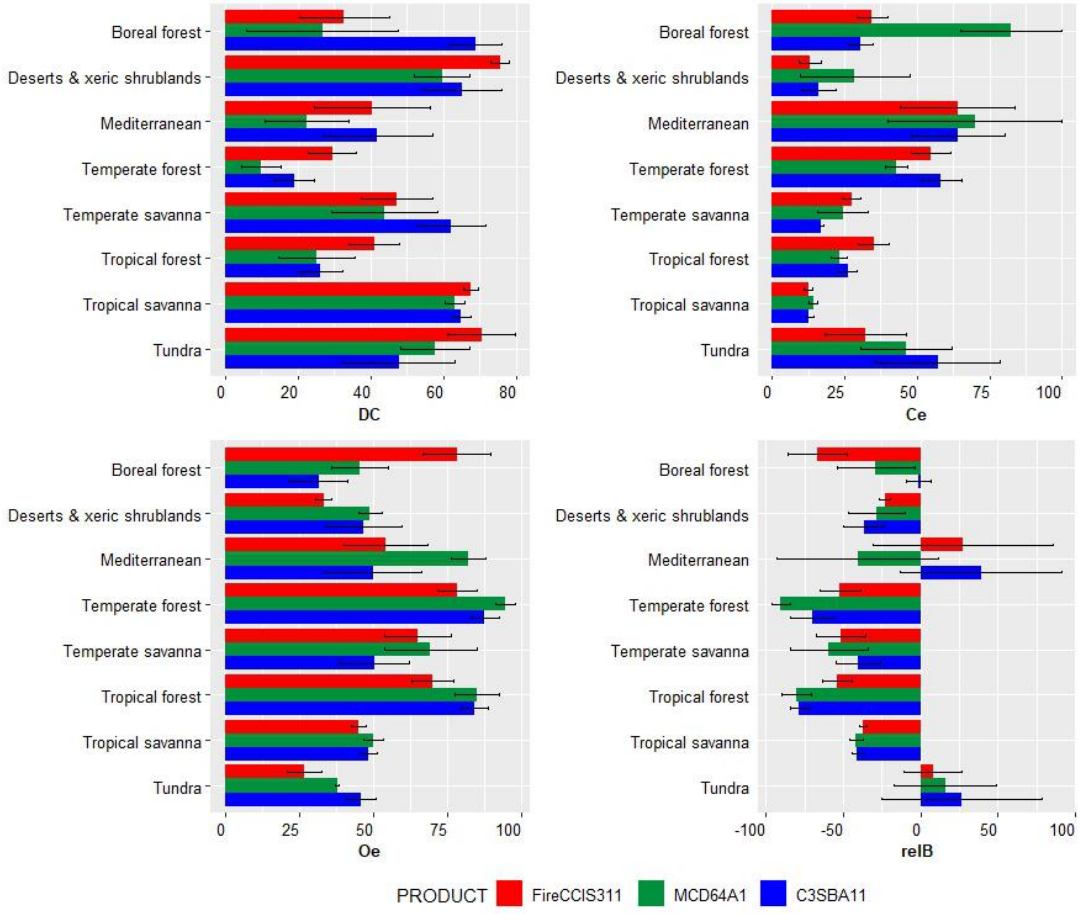

Figure 37. Estimated accuracy (%) of each product by biome for the year 2022. 95% confidence intervals are shown with its respective standard errors as segments (black lines). Dice coefficient (DC) (top left), Commission error (Oe) (top right), omission error (Oe) (bottom left) and relative bias (relB) (bottom right).

Table 17. Biome accuracy estimates (%) of validation year 2022 for FireCCIS311, MCD64A1 c6 and C3SBA11. Standard error (±) is shown in SE.

Biome	Metrics	2022					
		FireCCIS311		MCD64A1 C6		C3SBA11	
		value	SE	value	SE	value	SE
Boreal Forest	DC	32.68	12.42	26.69	20.99	68.86	7.04
	Ce	34.54	5.23	82.34	17.41	30.59	4.24
	Oe	78.22	11.40	45.41	9.47	31.68	9.76
	relB	-66.73	19.28	-29.00	25.00	-1.57	8.05
Deserts & Xeric shrublands	DC	75.56	2.63	59.62	7.52	65.13	10.82
	Ce	12.99	3.77	28.42	18.93	15.99	5.88
	Oe	33.22	2.77	48.92	3.82	46.83	12.93
	relB	-23.26	3.72	-28.64	18.58	-36.70	13.31
Mediterranean	DC	40.36	16.03	22.42	11.45	41.88	15.08
	Ce	63.98	19.72	69.92	30.15	64.04	16.24
	Oe	54.10	14.03	82.13	5.91	49.88	16.49
	relB	27.43	58.16	-40.59	52.41	39.35	52.32
Temperate forest	DC	29.40	6.71	10.00	5.49	19.07	5.27
	Ce	54.63	6.84	42.92	3.93	58.36	7.10
	Oe	78.25	6.53	94.52	3.29	87.63	4.64
	relB	-52.06	13.09	-90.40	5.72	-70.30	13.91
Temperate savanna	DC	47.25	9.90	43.78	14.48	62.14	9.32
	Ce	27.38	3.31	24.38	8.59	17.03	0.74
	Oe	64.99	11.04	69.19	15.63	50.32	11.93
	relB	-51.78	15.91	-59.25	24.96	-40.13	14.44
Tropical forest	DC	40.95	7.05	25.25	10.46	26.21	6.10
	Ce	35.00	5.14	23.15	2.75	26.09	3.28
	Oe	70.11	7.01	84.89	7.47	84.07	4.46
	relB	-54.01	9.91	-80.34	9.63	-78.45	5.81
Tropical savanna	DC	67.50	2.07	63.19	2.77	64.86	2.54
	Ce	12.50	1.40	14.22	1.57	12.75	1.43
	Oe	45.05	2.52	49.98	3.44	48.39	3.01
	relB	-37.21	2.72	-41.68	4.17	-40.85	3.24
Tundra	DC	70.46	9.26	57.66	9.45	47.83	15.41
	Ce	32.22	13.84	46.29	15.88	57.18	21.57
	Oe	26.64	5.87	37.77	0.75	45.85	5.18
	relB	8.24	18.75	15.86	32.96	26.47	52.13

4.4 Comparison of fire patch sizes distribution in the years 2019 -2022

A comparison of the sample polygon sizes for the years 2019, 2020, 2021 and 2022 was performed to establish whether the increased omission of the last two years could be related to the size of the reference BA patches obtained, as it is well known that small burn patches are more difficult to detect using coarse resolution sensors (Chuvieco et al.,

2022). Table 18 and Figure 38 show the number of reference BA patches and their equivalent in percentage (%), classified in sizes (hectares) for each year.

Table 18. Number of reference BA patches and their equivalent in percentage (%), classified in sizes (hectares) for 2019-2022

Hectares (ha)	2019		2020		2021		2022	
	Number of BA patches	%	Number of BA patches	%	Number of BA patches	%	Number of BA patches	%
<100	601771	97.47	784385	98.02	718984	97.53	441035	97.02
100-250	8950	1.45	9335	1.17	10598	1.44	7745	1.7
250-500	3298	0.53	3304	0.41	3876	0.53	2904	0.64
500-1000	1669	0.27	1637	0.2	1941	0.26	1468	0.32
>1000	1726	0.28	1596	0.2	1799	0.24	1413	0.31
Total	**617414**	**100**	**800257**	**100**	**737198**	**100**	**454565**	**100**

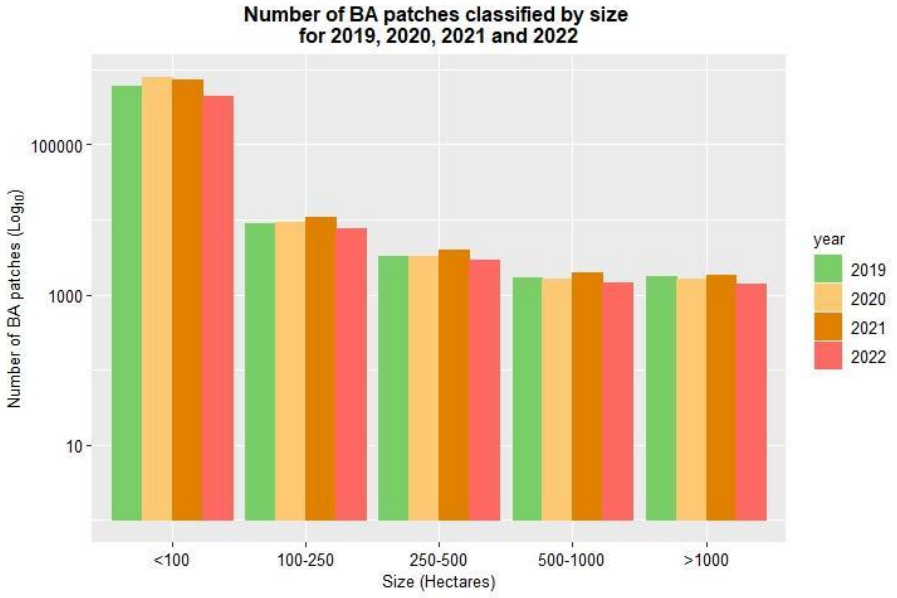

Figure 38. Number of reference BA patches (in logarithmic scale) classified by size for 2019-2022.

In the 2021 dataset, there was a minimal increase in percentage of fires of less than 100 ha compared to 2019 (only 0.06%), while, relative to 2020, 0.49% fewer surface plots burned in 2021. In 2022, the number of small fires was significantly reduced (39% compared to 2021, 44% compared to 2020 and 27% compared to 2019). This reduction in small polygons is related to the decrease in commission errors, coinciding with a similarity in the 2020-2022 metrics, showing in 2022 a lower Oe value than 2021. It is also observed that although 2022 has a lower number of small polygons than 2019, it does not reach the accuracy of that year. Although the methodology was the same for all years, the quality of the sensors used might have been higher in 2019 compared to 2022. Additionally, biomes such as Boreal Forest, Mediterranean and Temperate Forest presented considerable increases in omission values in 2022 (Table 7 and Table 8), which affects the overall metrics.

4.5 Comparison of global fire products to reference

The fire reference product was used for comparison to coarse resolution BA products FireCCI51 (2002-2019) and MCD64A1 C6 (2000-2019) over the validation units of the three ROIs. This comparison is considered as validation at stage 1, since it is lacking a sampling design to generate statistically robust accuracy metrics; hence it can provide only insights on the products agreement.

Figure 39 shows the comparison of omission and commission errors for the three products FireCCISFDL, FireCCI51 and MCD64A1 C6 computed for the validation units for each ROI (green Amazonia, blue Sahel, orange Siberia) compared to omission and commission errors for each ROI and product (derived by pooling together all units). Finally, overall metrics were computed (all units polled together regardless the ROI) and plotted in the graphs (red star marker). These graphs confirm the high variability of the accuracy for the validation units for the FireCCISFDL product and, in particular, for the Sahel ROI. The graphs for the coarse resolution products highlight a greater commission error for Sahel while the opposite occurs for Amazonia (i.e., omission is lower than commission errors as highlighted by the blue points below the 1:1 line) that is the expected behaviour of a coarse resolution product (omission due to undetected small burns).

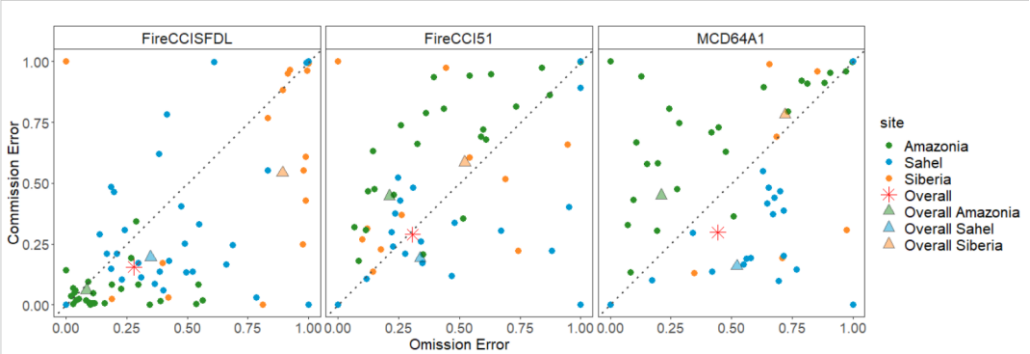

Figure 39. Comparison of omission and commission errors in the FireCCISFDL, FireCCI51 and MCD64A1 C6 BA products for the validation units and overall for the three sites.

Figure 40, Figure 41 and Figure 42 show the scatter plot of the proportion of area burned in the reference (x axis) and in the products (y axis) for the 5 km x 5km grid analysis. Metrics of the regression analysis are summarised in Table 19.

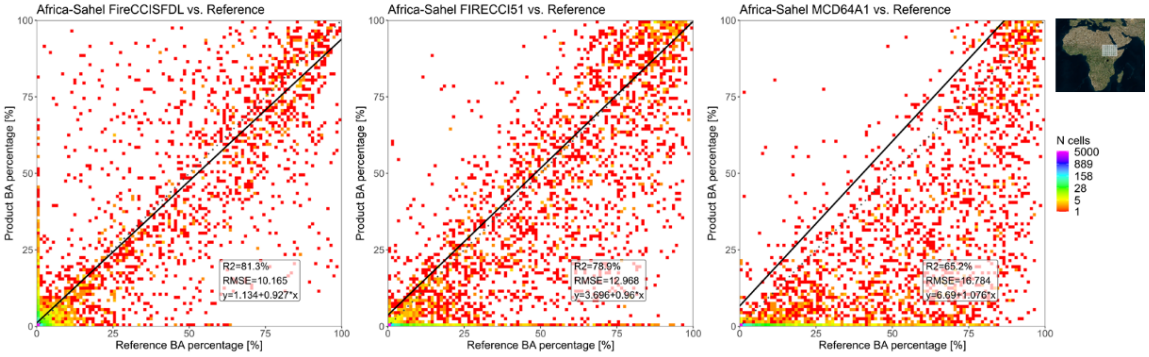

Figure 40. Scatter plots of the proportion of area burned in the reference (x-axis) and in the FireCCISFDL, FireCCI51 and MCD64A1 C6 BA products (y axis) for the Africa-Sahel ROI. Scatter plot are produced with a 10% bin size and the marker colors represent the number of points in each bin (number of 5 km x 5 km grid cells).

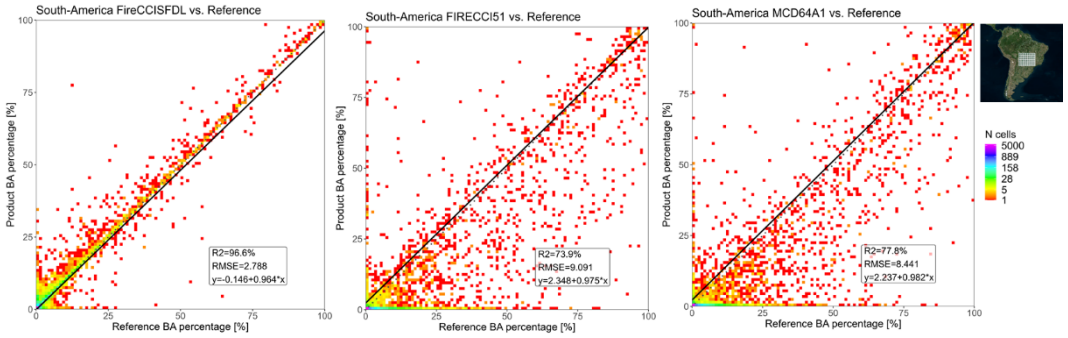

Figure 41. Scatter plots of the proportion of area burned in the reference (x-axis) and in the FireCCISFD, FireCCI51 and MCD64A1 C6 BA products (y axis) for the South America-Amazonia ROI. Scatter plot are produced with a 10% bin size and the marker colors represent the number of points in each bin (number of 5 km x 5 km grid cells).

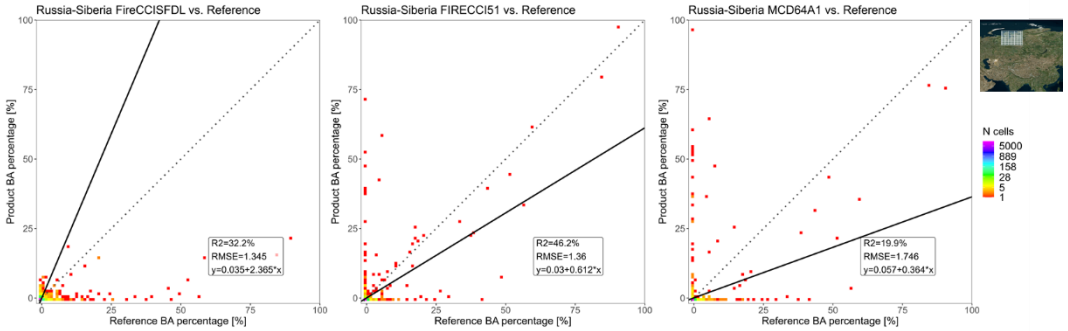

Figure 42. Scatter plots of the proportion of area burned in the reference (x-axis) and in the FireCCISFD, FireCCI51 and MCD64A1 C6 BA products (y axis) for the Russia-Siberia ROI. Scatter plot are produced with a 10% bin size and the marker colors represent the number of points in each bin (number of 5 km x 5 km grid cells).

Table 19. R2, RMSE, intercept and slope of the linear regression model between the proportion of burned area in 5 km x 5km grid cells in the reference (independent variable) and the BA products (dependent variable).

Amazonia (N=29)	**R2**	**RMSE**
FireCCISFDL	96.6%	2.788
FireCCI51	73.9%	9.091
MCD64A1 C6	77.8%	8.441
Siberia (N=26)	**R2**	**RMSE**
FireCCISFDL	32.2%	1.345
FireCCI51	46.2%	1.36
MCD64A1 C6	19.9%	1.746
Sahel (N=33)	**R2**	**RMSE**
FireCCISFDL	81.3%	10.165
FireCCI51	78.9%	12.968
MCD64A1 C6	65.2%	16.784

5 References

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., E. Roteta, M. Sali, D. Stroppiana, M. Boettcher, G. Kirches, T. Storm, A. Khairoun, M. L. Pettinari, M. Franquesa, C. Albergel (2022). Building a small fire database for Sub-Saharan Africa from Sentinel-2 high-resolution images, Science of The Total Environment, 845, 157139, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2022.157139.

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

Dinerstein, E., D. Olson, A. Joshi, C. Vynne, N.D. Burgess, E. Wikramanayake, N. Hahn, S. Palminteri, P. Hedao, R. Noss, M. Hansen, H. Locke, E.C. Ellis, B. Jones, C.V. Barber, R. Hayes, C. Kormos, V. Martin, E. Crist, W. Sechrest, L. Price, J.E.M. Baillie, D. Weeden, K. Suckling, C. Davis, N. Sizer, R. Moore, D. Thau, T. Birch, P. Potapov, S. Turubanova, A. Tyukavina, N. de Souza, L. Pintea, J.C. Brito, O.A. Llewellyn, A.G. Miller, A. Patzelt, S.A. Ghazanfar, J. Timberlake, H. Klöser, Y. Shennan-Farpón, R. Kindt, J.-P.B. Lillesø, P. van Breugel, L. Graudal, M. Voge, K.F. Al-Shammari, M. Saleem (2017). An ecoregion-based approach to protecting half the terrestrial realm, BioScience, 67, 534-545

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

Khairoun, A., Solano, E. (2023) ESA CCI ECV Fire Disturbance: D2.2.2 Algorithm Theoretical Basis Document-SFDL, version 1.2. Available at: https://climate.esa.int/en/projects/fire/key-documents/

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.

Kovalskyy, V, & D.P. Roy, (2013). The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30m Landsat data product generation, Remote Sensing of Environment, 130, 280-293, https://doi.org/10.1016/j.rse.2012.12.003.

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

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

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., 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

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.

Sali M, Piaser E, Boschetti M, Brivio PA, Sona G, Bordogna G, Stroppiana D. 2021. A Burned Area Mapping Algorithm for Sentinel-2 Data Based on Approximate Reasoning and Region Growing. Remote Sens. 13:2214, https://doi.org/10.3390/rs13112214.

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 fire perimeters to assess accuracy of Burned Area products over Sub-Saharan Africa for the year 2019, ISPRS Journal of Photogrammetry and Remote Sensing, 191, 223-234, ISSN 0924-2716, https://doi.org/10.1016/j.isprsjprs.2022.07.015.

Wulder, M.A., White J.C., Loveland T.R., Woodcock C.E., Belward A.S., Cohen W.B., Fosnight E.A., Shaw J., Masek J.G., Roy D.P., 2016, The global Landsat archive: Status, consolidation, and direction, Remote Sensing of Environment, 185, 271-283, https://doi.org/10.1016/j.rse.2015.11.032.

6 Annex 1: Acronyms and abbreviations

AF	Africa – Sahel ROI
BA	Burned Area
C3SBA11	C3S Burned Area C3S Fire-BA v1.1
CCI	Climate Change Initiative
Ce	Commission error
DC	Dice Coefficient
DOY	Day of the Year
ECV	Essential Climate Variables
EO	Earth Observation
ESA	European Space Agency
FireCCI51	MODIS Fire_cci v5.1
FireCCIS311	Sentinel-3 FireCCI v1.1 BA product
FireCCISFDL	Landsat-based SFD Long term
GEE	Google Earth Engine
HR	High Resolution
KML	Keyhole Markup Language
L	Validation unit Length
L5	Landsat 5
L8	Landsat 8
L9	Landsat 9
MCD64A1 c6	MODIS Burned Area product collection 6

NBR	Normalized Burned Ratio
NBR2	Normalized Burned Ratio 2
NDVI	Normalized Difference Vegetation Index
NIR	Near InfraRed
Oe	Omission error
PVIR	Product Validation and Intercomparison Report
R2	Coefficient of Determination
RF	Random Forest
RGB	Red-Green-Blue composite
RMSE	Root Mean Square Error
ROI	Region of Interest
S2	Sentinel-2
SA	South America – Amazonia ROI
SCL	Scene Classification Layer
SE	Standard Error
SFD	Small Fire Dataset
SI	Russia – Siberia ROI
SWIR	Short Wave InfraRed
TSA	Thiesen Scene Area
USGS	United States Geological Survey
WRS	Worldwide Reference System