ESA Climate Change Initiative – Fire_cci
D1.3 Product Validation Plan (PVP)

Project Name	ECV Fire Disturbance: Fire_cci
Contract Nº	4000126706/19/I-NB
Issue Date	24/03/2021
Version	2.0
Author	Daniela Stroppiana, Lorenzo Busetto, Mirco Boschetti, Luigi Ranghetti, Matteo Sali
Document Ref.	Fire_cci_ D1.3_PVP_v2.0
Document type	Public

To be cited as: Stroppiana D., Busetto L., Boschetti M., Ranghetti L., Sali M. (2021) ESA CCI ECV Fire Disturbance: D1.3 Product Validation Plan, version 2.0. Available at: https://climate.esa.int/en/projects/fire/key-documents/

Project Partners

Prime Contractor/ Scientific Lead & Project Management	UAH – University of Alcala (Spain)
Earth Observation Team	UAH – University of Alcala (Spain) UPM – Universidad Politécnica de Madrid (Spain) CNR-IREA - National Research Council of Italy – Institute for Electromagnetic Sensing of the Environment (Italy)
System Engineering	BC – Brockmann Consult (Germany)
Climate Modelling Group	MPIM – Max Planck Institute for Meteorology (Germany) CNRS - National Centre for Scientific 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) Joshua Lizundia (UAH) Gonzalo Otón (UAH) Mihai Tanase (UAH) Consuelo Gonzalo (UPM) Dionisio Rodríguez Esparragón (UPM) Daniela Stroppiana (IREA) Mirco Boschetti (IREA) Thomas Storm (BC) Angelika Heil (MPIM) Idir Bouarar (MPIM) Florent Mouillot (CNRS) Philippe Ciais (CNRS)	emilio.chuvieco@uah.es mlucrecia.pettinari@uah.es joshua.lizundia@uah.es gonzalo.oton@uah.es mihai.tanase@uah.es consuelo.gonzalo@upm.es dionisio.rodriguez@ulpgc.es stroppiana.d@irea.cnr.it boschetti.m@irea.cnr.it thomas.storm@brockmann-consult.de angelika.heil@mpimet.mpg.de idir.bouarar@mpimet.mpg.de florent.mouillot@cefe.cnrs.fr philippe.ciais@lsce.ipsl.fr	electronic copy

Summary

This Product Validation Plan (PVP) describes the approaches and methods used to assess the quality of global and regional BA products obtained from the Fire_cci algorithms.

	Affiliation/Function	Name	Date
Prepared	CNR	Daniela Stroppiana, Lorenzo Busetto, Mirco Boschetti, Luigi Langhetti, Matteo Sali	18/03/2021
Reviewed	UAH – Project Manager	Lucrecia Pettinari	24/03/2021
Authorized	UAH - Science Leader	Emilio Chuvieco	24/03/2021
Accepted	ESA - Technical Officer	Clément Albergel	24/03/2021

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

Document Status Sheet

Issue	Date	Details
1.0	01/07/2019	First Version of the document
1.1	20/12/2019	Addressing comments of ESA-CCI-EOPS-GLCR-MEM-19-0312
1.2	08/03/2021	Review and update of the document
2.0	24/03/2021	Addressing comments of Fire_cci+_D3.1.PVP_v1.2_RID

Document Change Record

Issue	Date	Request	Location	Details
1.1	20/12/2019	ESA UAH CNR	Sections 1, 3, 4, 4.2.1, 4,2,2, 4.3, 5, 5.4 Section 2 Sections 3.2, 4.2, 4,2,3, 5.2, Annex 2 Section 5.5 Section 6	Text updated Validation years updated Minor updates to the text New section added References updated
1.2	08/03/2021	CNR	Sections 1, 3, 5.5, Annexes 2 and 3 Sections 2, 3.1, 4.1, 5, 5.1, 5.2, 5.3, 5.4 Sections 4, 4.2 (and sub-sections), 4.3 Section 6	Small changes in the text Text partially updated Section re-written References updated
2.0	24/03/2021	ESA UAH CNR	Table 1 Section 3 Figures 2, 6, 7, 11 All document	Table updated Text reviewed Captions reviewed Minor text corrections

Table of Contents

1 Executive Summary 6
2 Introduction and objectives......6
3 Review of existing validation methods for BA products 6
3.1 State of the art of validation methods from international initiatives 7
3.2 Background from previous Fire cci phases......8
3.2.1 Validation protocol......8
3.2.2 Specific validation protocol for the FireCCISFD11 product......10
4 Validation criteria to follow in the Fire_cci project......11
4.1 Reference fire perimeters and EO data 11
4.2 Criteria to select validation units and sampling scheme......12
4.2.1 Spatial definition of validation units 13
4.2.2 Temporal definition of validation units......13
4.2.3 Sampling cardinality......14
4.2.4 Sampling design 15
4.3 Protocol for creating and documenting BA reference files 19
5 Description of validation activities 21
5.1 Sampling scheme implementation and reference EO data 21
5.2 Processing steps for creating reference BA data......22
5.3 Processing steps for Fire_cci BA product extraction......22
5.4 Cross-tabulation and accuracy metrics 22
5.5 Error analysis 23
6 References......24
Annex 1 Acronyms and abbreviations......27
Annex 2 Accuracy metrics 28
Annex 3 Data structure and naming convention 29

List of Tables

Table 1. BA products foreseen in the Fire_cci project, time period covered (in brackets), source data (Mission & sensor) and the years that are validated...... 6
Table 2. 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). 28
Table 3. Accuracy metrics computed from the error matrix 28
Table 4. Satellite-sensor codes naming convention...... 29
Table 5. Example of attribute table for BA reference fire perimeter shapefile 30

List of Figures

Figure 1: Thiessen scene areas (TSAs) with at least one unit selected in the sample and biome stratification based on a reclassification of the 14 Olson biomes (Olson et al. 2001). 9

Figure 2: Example of reference burn perimeters (yellow polygons) over Miombo woodlands (Zambia, Africa) derived from supervised classification of Landsat image pair: August 14th (a) and August 30th (b), 2018. 10

Figure 3: a) Spatial distribution of reference data availability for long sampling units shown as percentage of time; b) Thiessen scene areas (TSAs) with at least one unit selected in the sample and biome stratification based on a reclassification of the 14 Olson biomes (Olson et al. 2001). 10

Figure 4: Illustration of short (a) and long (b) sampling units for a Thiessen scene area (TSA) on a three-dimensional space. Each sampling unit is delimited spatially by a TSA (two-dimensions) and temporally (the third dimension) by the time between two or more consecutive Landsat images. Images are displayed as false colour composites with SWIR, NIR and red bands in the red, green and blue channels respectively. 11

Figure 5. Flowchart of the sampling scheme for L8 and S2 global and regional/continental validation. Validation area is the geographic location of the validation units identified as TSAs and S2 tiles for global and regional scheme, respectively, ΔT [days] is the time step between consecutive L8/S2 scenes, CC [%] is the scene level cloud cover percentage, L is the length of the “long sampling unit” [days]. 12

Figure 6: a) example S2 tiles not suitable (yellow) as validation areas for sampling since they cover different orbits (Blue lines); b) example S2 tiles suitable (green) as validation areas for sampling since they cover the same orbit (blue lines); c) all S2 tiles suitable for sampling (after first filtering) in the example area centred over Senegal. 16

Figure 7: a) example S2 tiles suitable (green/blue) and not suitable (yellow) as validation areas for sampling due to overlap of adjacent UTM zones; b) S2 tiles suitable (orange) for sampling using a random selection in the example area centred over Senegal. 16

Figure 8: a) S2 tiles available for sampling after applying the filtering criteria; b) total burned area for each S2 [m2]; c) Major Olson biome for each S2 tile. 17

Figure 9: S2 tiles suitable for sampling for each stratum (intersection between biome and high/low fire intensity). 17

Figure 10. Example of S2 tile (36KWE, Tropical savanna) for the year 2019: single scene cloud cover (dotted line), S2 scenes with cloud cover < 10% (red dots) and date of occurrence of maximum cumulated cloud cover of 30% from first suitable scene (vertical orange line). 18

Figure 11: Location of the areas of interest for SAR-O algorithm development showed over the total burned area for year 2016 (pixel level, FireCCI51 product) and the main Olson biomes (Olson et al., 2001) (a), and daily mean cloud cover at tile level (Military Grid Reference System - MGRS) (b). 19

Figure 12. Flowchart showing steps for extraction of fire reference perimeters over sampled validation units. 20

1 Executive Summary

This document is the *Product Validation Plan* (PVP) that outlines the approach to follow for the validation of the global and regional BA products. The validation protocol defined and implemented in the Fire_cci builds on standard methods published in the literature (Padilla et al. 2017, Boschetti et al., 2016) and developed during previous Fire_cci phases (Fire_cci background). To this aim, this document reviews validation methods and protocols from the Fire_cci background and the literature from the EO (Earth Observation) research community (Section 3). Sections 4 and 5 describe the protocol that is implemented for validation and the activities carried out, respectively.

2 Introduction and objectives

The PVP describes the approaches and methods that are used to assess the quality of burned area (BA) products obtained by applying the Fire_cci algorithms (Table 1).

The products object of validation cover global and regional (Africa), one year (2019) or multi-annual (2018-2020) (Table 1). Furthermore, validation will be done on BA products delivered over the Africa test sites where the Sentinel-1 and Sentinel-2 algorithms have been developed.

Table 1. BA products foreseen in the Fire_cci project, time period covered (in brackets), source data (Mission & sensor) and the years that are validated.

Products	Sensor	Area	Years to be validated
FireCCI51 (2018-2019)	Terra-MODIS	**Global**	**2018-2019**
FireCCIS311 (2019-2020)	S-3 SYN (OLCI+SLSTR)	**Global**	**2019-2020**
FireCCI60 (1982-2019)	Merged AVHRR-MODIS-SYN BA	**Global**	**2018-2019**
FireCCISFD20 (2019)	S2 MSI	**Sub-Saharan Africa**	**2019**
FireCCIS1S2AF10 (2019)	S1 SAR and S-2 MSI	**Selected sites in Africa**	**2019**

In the case of the FireCCI51, and FireCCI60 products, which cover a longer time period, validation will be focused on years 2018 to 2019. For the other products, validation will cover 2019 (and extend to 2020 for FireCCIS311).

3 Review of existing validation methods for BA products

Validation is a critical and necessary task in any EO project, as it provides a quantitative assessment of the accuracy of geo-information delivered by the product; indeed, accuracy is relevant information for both scientists and end-users (Congalton and Green 1999). With increasing availability of regional and global burned area products delivered at different scales, agreement on protocols and data to be used as reference has to be achieved. For large-area (global and/or continental) EO products, validation is even more challenging because of the great variety of ecosystem and climatic conditions that could affect mapping accuracy (Chuvieco et al., 2008).

Several issues arise when addressing the task of assessing accuracy of EO-derived maps and among them: statistically robust sampling of geographic areas worldwide, reference data to be assumed as having no or least error rates and metrics to be used as a measure of accuracy.

The need for the definition of standard approaches and protocols for validation of EO products led to the creation of international working groups such as the Committee on Earth Observation Satellites' Land Product Validation Subgroup (CEOS-LPVS) (http://lpvs.gsfc.nasa.gov/, accessed January 2021). CEOS-LPVS defines validation as: "The process of assessing, by independent means, the quality of the data products derived from the system outputs" (European Space Agency, 2007; Morisette et al. 2006).

For burned area assessment globally or regionally, the use of *in-situ* reference field data is not feasible or very expensive in time and effort. Therefore, remote sensing validation projects rely on reference burn perimeters derived from medium spatial resolution satellite images (10-30 m) extracted from image pairs for multitemporal comparison (Franquesa et al, 2020). A major data source used for this purpose are Landsat missions with imagery acquired by TM/ETM+ and OLI sensors (Padilla et al., 2015; Vanderhoof et al., 2017; Roteta et al., 2019) since they provide historical long-term archives for validation of multi-annual BA products. Sentinel-2 MSI imagery is also a suitable source of data for validation purposes, although only covering recent years. Since missing data and/or atmospheric disturbance (clouds, cloud shadows, smoke plumes) could affect multitemporal image series, thus reducing data availability for building burn reference perimeters, denser series could be obtained with techniques of data fusion in a multisource approach combining, for example, Landsat and Sentinel-2 images.

Reference images should be acquired simultaneously as to portray the same ground conditions as the input images from which the validating product is generated. Methods on the generation of BA reference data used in previous Fire cci are fully detailed in Padilla et al. (2014c).

Accuracy is characterized through cross-tabulation, by accounting for the spatio-temporal coincidences and disagreements on estimates of location and timing of burns between a reference map and the target map (Padilla et al. 2017; Padilla et al. 2014a; Padilla et al. 2015). In this framework, accuracy is measured in terms of global agreement of the common overlapping area of reference and classified products.

In this document, "accuracy" refers to the closeness to the ground truth or the condition of the surface to be assumed as ground truth. Whereas, "uncertainty" generally refers to a probability that a pixel is burned based on sources of error within the input data and algorithm.

In remote sensing, the assessment of uncertainty is the quantification of the error associated to all steps of data acquisition and processing although sources of uncertainty are often unknown and difficult to be quantified.

This document and the validation procedure described herein, only refers to thematic accuracy assessment. The estimation of uncertainty is out of the scope of this document and of the validation and it is fully addressed in Lizundia-Loiola et al. (2020b).

3.1 State of the art of validation methods from international initiatives

Through the first decade of the 2000s, large-scale BA products were typically subjected to a first stage validation. Padilla et al. (2014c) summarizes validation efforts for validating BA products in international projects. At global scale, some examples are Globcarbon (Plummer et al., 2007), or L3JRC (Tansey et al., 2008), which were validated with a set of 72 Landsat scenes mostly from the year 2000. At the regional scale, the

MODIS-MCD45 in Southern Africa (Roy et al., 2008; Roy and Boschetti 2009) was validated with 11 Landsat scenes. Over Latin America, Chuvieco et al. (2008) validated a regional product obtained from MODIS reflectance data with 19 Landsat scenes and 9 China–Brazil Earth Resources Satellite (CBERS) scenes. A first evaluation of MODIS burned area products (Roy et al. 2008) was produced using as reference data the MODIS active fire product, globally from July 2001 to June 2002, with linear regressions for the assessment. Several other publications are available comparing global BA datasets and assessing their relative accuracy (e.g. Padilla et al, 2015).

In order to achieve higher stage validation, a probabilistic sampling design in time and space is necessary. This issue was addressed by Padilla et al. (2014a; 2015) who implemented a stratified random sampling for validating and comparing global BA products. Boschetti et al. (2016) improved the sampling by specifically including the temporal dimension at the sampling units, and recently a methodological approach for stratification and sample allocation of reference burned area data has been proposed (Padilla et al 2017). This approach was applied during Fire_cci Phase 2.

During 2020, a group of validation datasets at global and regional level were made publicly available (Franquesa et al, 2020), including a description of their characteristics and the methodology used to obtain them, to be useful for the research community.

3.2 Background from previous Fire cci phases

In previous phases of Fire_cci, a large effort was placed in validation activities to address a two-fold objective:

1. to define a robust validation protocol based on probabilistic sampling;
2. to cover several years thus assuring consistent accuracy metrics through the years.

Indeed, temporal variability of algorithm performance is one of the key validation aspects to be assessed according to end-user requirements (Heil et al., 2017). Validation then should provide a measure of whether results include temporal trends or not.

In this framework, in Fire_cci previous phases, reference fire perimeters were generated to cover twelve years, (multi-annual validation, from 2003 to 2014) following a probability sampling scheme; this way validation of global BA products achieved CEOS-LPV validation stage 4.

In previous Fire_cci phases, samples of reference data were also specifically generated over Africa (2016) to validate the Small Fire Database (SFD, FireCCISFD11) derived from Sentinel-2 data. This separate sample used consecutive images pairs, over a time lag of at least 100 days, ensuring long temporal overlaps with SFD BA estimates. In fact, the temporal resolution of the SFDs (Sentinel-2 observations of the Earth are acquired every 5-10 days) might generate some discrepancies in the detection of burn date. However, dating errors are mitigated by increasing the temporal time span covered by the reference datasets compared to using single image pairs used for validation of global BA products.

3.2.1 Validation protocol

For global BA products, the validation protocol defined and implemented for previous Fire cci Phases is fully documented and described in projects reporting documents (Padilla et al. 2014c; Padilla et al. 2018) and in the literature (Padilla et al. 2014a; Padilla et al. 2014b; Padilla et al. 2015; Padilla et al. 2017).

The validation protocol for global BA products was composed of the following steps:

- Sampling scheme design: sampling units were designed based on Thiessen scene areas (TSAs) constructed by Cohen et al. (2010) and Kennedy et al. (2010) specifically for use with Landsat WRS-2 frames. Sampling units were selected by stratified random sampling in each calendar year (2003-2014) taking into account the major Olson biomes (Olson et al. 2001) and with special focus on regions with high and low fire activity as depicted by the MODIS-MCD64A1 Collection 5 product (Giglio et al. 2009) (Figure 1). Reference data was subsampled using a 30 km wide by 20 km height window located in the centre of the scene. An upper limit of 100 sampling units (i.e. image pairs) per year was set a priori.
- Reference fire perimeters generation: reference fire perimeters were generated from two consecutive images acquired at the same TSA, i.e. from multi-temporal comparison of medium resolution satellite imagery (Landsat) (Figure 2). Reference data was subsampled using a 30 km wide by 20 km height window located in the centre of the scene. High accuracy perimeters were obtained with a semi-automatic mapping of burns: a machine learning algorithm (i.e. Random Forest classifier) (Breiman, 2001) followed by a systematic quality control performed through visual inspection and refinement.
- Accuracy metrics computation: the following accuracy metrics are computed from the error matrix (Congalton and Green 1999; Latifovic and Olthof 2004): commission error ratio, omission error ratio, Dice Coefficient (*DC*) (Dice 1945), bias and relative bias (Table 3). Temporal stability of the accuracy was also computed for multi-annual products.

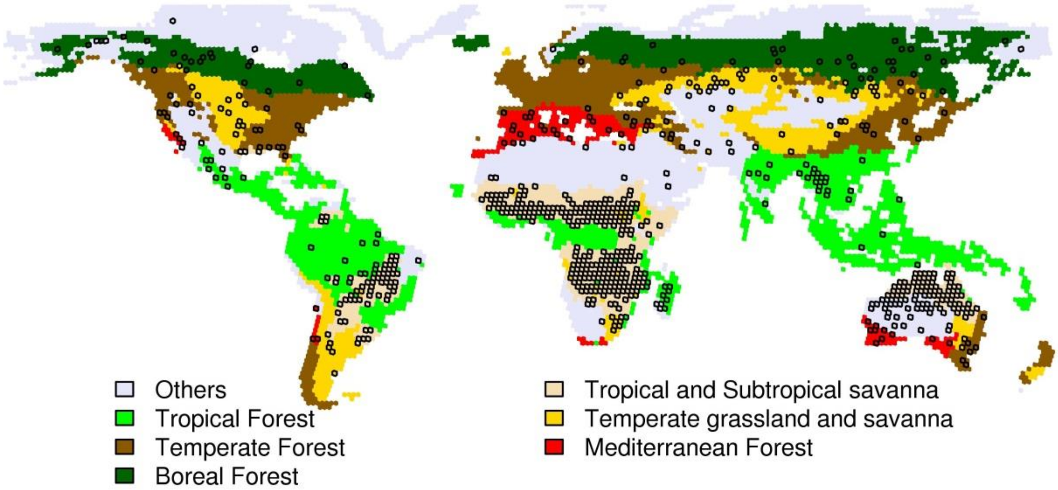

Figure 1: Thiessen scene areas (TSAs) with at least one unit selected in the sample and biome stratification based on a reclassification of the 14 Olson biomes (Olson et al. 2001).

Specifically, to assess accuracy of multi-annual global BA products, about 1200 pairs of Landsat images were processed for the period 2003-2014; Figure 2 shows an example for image pairs over Canada.

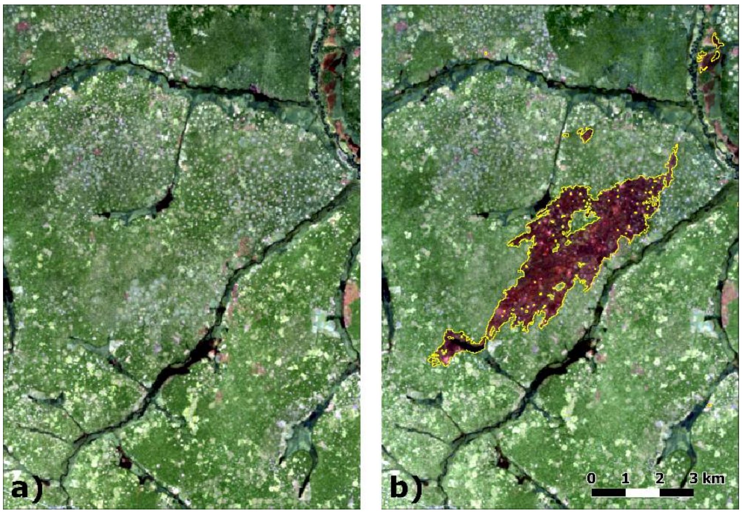

Figure 2: Example of reference burn perimeters (yellow polygons) over Miombo woodlands (Zambia, Africa) derived from supervised classification of Landsat image pair: August 14th (a) and August 30th (b), 2018.

3.2.2 Specific validation protocol for the FireCCISFD11 product

The **Sentinel-2 Small Fire Database (SFD)** (FireCCISFD11) (Roteta et al., 2019) was produced for the year 2016 over the Sub-Saharan African continent; the product was derived from processing of all S2 images acquired during a single year. In the previous Fire_cci phase, a sample design was specifically generated to validate the FireCCISFD11 product. Indeed, in order to mitigate the effect of temporal errors in the detection date, due to the lower temporal resolution of the product, consecutive pairs of reference images were used instead of the single image pair (as done for the global products). In this sampling design, consecutive Landsat image pairs separated by 16 days, or less, and covering at least 100 days were selected. Such a long coverage (>= 100 days) was set to ensure a good overlap with products generated with S2 imagery, which do not observe the surface on a near daily basis. Figure 3 shows the spatial distribution of data availability expressed as percentage of time on Thiessen scene areas covered by consecutive pairs (a) and the selected sampling units for Africa 2016 (b).

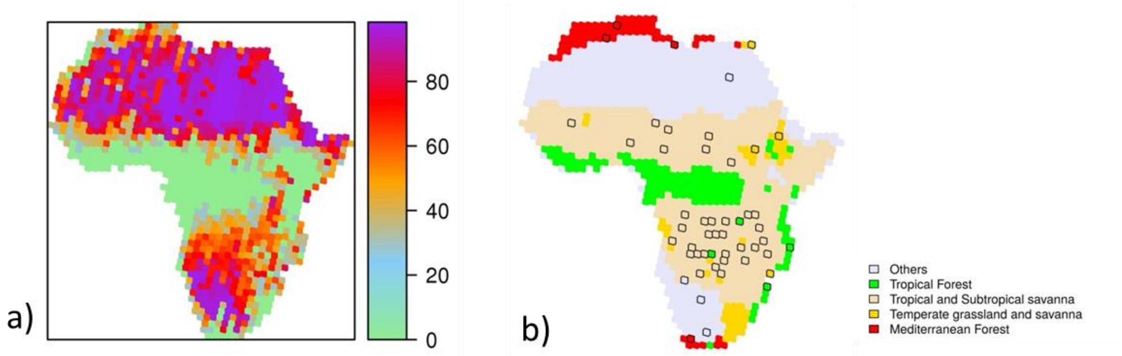

Figure 3: a) Spatial distribution of reference data availability for long sampling units shown as percentage of time; b) Thiessen scene areas (TSAs) with at least one unit selected in the sample and biome stratification based on a reclassification of the 14 Olson biomes (Olson et al. 2001).

This criterion for the selection of reference images within the sampled TSAs was named “long sampling unit”. Contrarily, the unit defined by a single pair of consecutive images is referred as “short” (Padilla et al., 2018). Hence, a long sampling unit is temporally bounded by the acquisition dates of the first and last image (Figure 4).

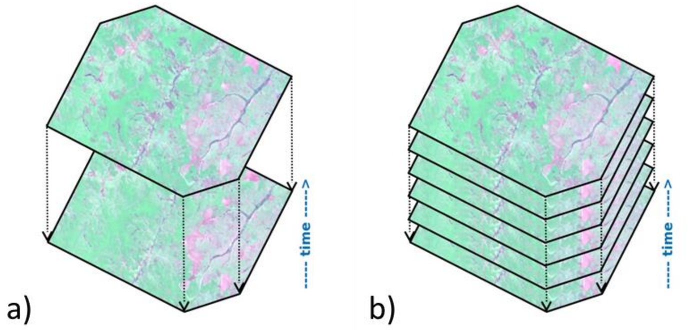

Figure 4: Illustration of short (a) and long (b) sampling units for a Thiessen scene area (TSA) on a three-dimensional space. Each sampling unit is delimited spatially by a TSA (two-dimensions) and temporally (the third dimension) by the time between two or more consecutive Landsat images. Images are displayed as false colour composites with SWIR, NIR and red bands in the red, green and blue channels respectively.

4 Validation criteria to follow in the Fire_cci project

In the current Fire_cci project, validation is carried out for the three major categories of products (Table 1): **global BA** (FireCCI51, FireCCILT11[1], FireCCIS311, and FireCCI60), **regional BA** for Africa (FireCCISFD20) and **local BA** products for three test sites in Africa derived from the combined use of S1 and S2 (FireCCIS1S2AF10).

4.1 Reference fire perimeters and EO data

A key requirement for reference datasets is to be highly **accurate** and generated **independently** so that it can be assumed as **ground truth**. Since systematic collection of reliable and representative ground/*in situ* fire data is hardly feasible to be achieved over large areas (global/continental), the use of medium/high resolution remotely sensed data is widely accepted (Morisette et al., 2006).

High/very high resolution (HR/VHR) remotely sensed data are characterized by low revisiting time and limited geographical coverage thus not assuring systematic sampling in time and space for the estimation of statistically robust and unbiased accuracy metrics. Therefore, this type of EO data is not suitable for achieving validation stage 4 (Morisette et al., 2006).

Moreover, reference fire perimeters are required to be as accurate as possible to be assumed as ground truth; this objective can be achieved by processing pairs of consecutive source images with operator intervention and supervised classification algorithm (e.g. random Forest).

Finally, independence, which is a critical issue in validation, assures that estimated accuracy metrics are unbiased. Independence requires that source datasets used for validation be not used during the design of the BA algorithms, either for calibration or “tuning” processes. It is acceptable that EO source data (i.e. imagery) used for deriving BA products (to be validated) is also used for deriving the reference datasets as long as processing is separated and independent.

[1] The FireCCILT11 has been validated with a different methodology to the other global products. Its much coarser spatial resolution does not recommend to use Landsat or S-2 data for validation. We instead use FireCCI51 for assessing outputs of FireCCI51 at grid level (0.25-degree resolution), as indicated in Otón et al. (2021, in review)

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

For what concerns BA products over the three test sites in Africa, HR/VHR Planetscope images are exploited as source of EO data for the generation of fire reference perimeters. Despite that this dataset does not guarantee a systematic acquisition, it represents a higher spatial resolution data source useful to assess decametric BA products derived from Sentinel data. Under this condition, Level 1 validation stage can be achieved.

4.2 Criteria to select validation units and sampling scheme

Sampling is defined as the temporal and spatial selection of **validation units** for the estimation of accuracy metrics and it is a critical step in any validation activity to optimize resources dedicated to the generation of the reference data. A sampling scheme should be based on random selection of representative areas (sites) to assure unbiased collection of reference data and estimation of accuracy metrics. Random sampling is stratified based on biome and fire characteristics/conditions to guarantee covering major factors of variability.

Figure 5 presents a flowchart of the sampling scheme for the global and regional BA products, and each step is further described in the following sections.

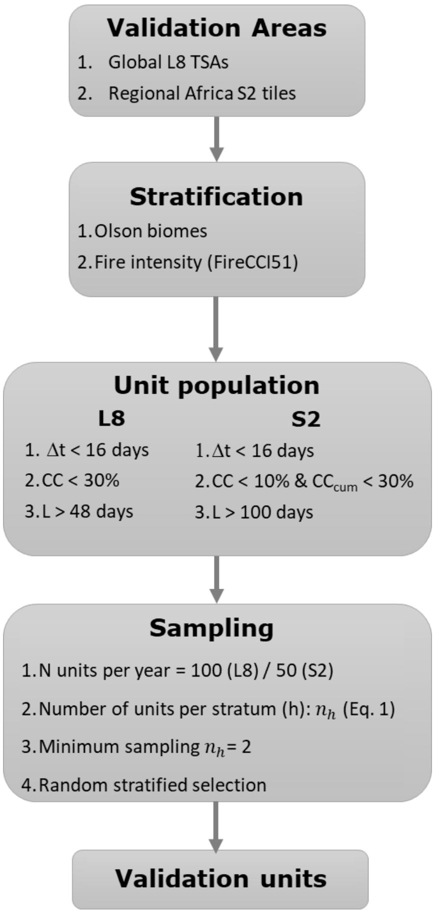

Figure 5. Flowchart of the sampling scheme for L8 and S2 global and regional/continental validation. Validation area is the geographic location of the validation units identified as TSAs and S2 tiles for global and regional scheme, respectively, ΔT [days] is the time step between consecutive L8/S2 scenes, CC [%] is the scene level cloud cover percentage, L is the length of the "long sampling unit" [days].

4.2.1 Spatial definition of validation units

In Fire_cci, the spatial dimension of the validation is defined by contiguous non-overlapping validation areas while the temporal dimension is defined by the dates of (consecutive) pair(s) of remotely sensed source images (i.e. L8/S2). The validation area of each unit covers approximately 100 km by 100 km, where fire reference perimeters are extracted by comparison of pairs of EO images. With respect to previous Fire cci validation efforts, the area was increased to represent a larger portion of the surface and to allow a more accurate analysis of fire size and patches. As explained in the following specific section, validation area is a subset of the L8 frame while it coincides with tile extent for S2.

Since validation areas should be selected to properly represent the variety of conditions that affect the accuracy of BA cartography, a stratified random sampling scheme is adopted. Stratification is based on i) major Olson biomes (Olson et al. 2001) and ii) fire intensity defined hereafter from the annual amount of area burned by creating strata of high and low fire activity within each biome.

For each validation area, fire activity is estimated from the annual global FireCCI51 (Chuvieco et al., 2018; Lizundia-Loiola et al., 2020a) BA product. From year to year, fire area and fire activity might significantly change, thus requiring a source of information specific for each year that is object of validation. With respect to previous Fire_cci phases, the source of information for classifying into high/low fire intensity was updated from lower spatial resolution MCD64A1 Version 6 Burned Area data product to the higher resolution FireCCI51 BA product.

High/low fire intensity classes are assigned for each unit based on the total annual burned area (TotBA, m^2) and by applying a threshold estimated for each biome. Each validation area is assigned to either the high (TotBA > threshold) or low (TotBA <= threshold) fire intensity class based on a threshold value calculated for each biome as the value of the 20th percentile of the normalized cumulated burned areas (Padilla et al. 2017, Boschetti et al., 2016). To compute this value all validation areas from the same biome are sorted in increasing order based on the normalized TotBA to extract the TotBA value of the 20th percentile. Normalized values are computed with respect to the biome's maximum value of TotBA. These steps are applied to both the global and regional/continental sampling schemes and threshold values derived for year 2018 (global) and 2019 (Africa continental) are reported in the Product Validation and Inter-comparison Report (PVIR) (Stroppiana et al, 2020).

For validation of both global and Africa continental BA products, the validation areas in the biome/fire intensity strata are used for building the sampling population only if, for a given validation area, time series of L8 and S2 images, respectively, satisfy requirements for generating a "long validation unit" (Section 3.2).

4.2.2 Temporal definition of validation units

A "long validation unit" is a time series of clear-sky (according to a maximum cloud cover percentage allowed for each scene) L8/S2 consecutive images from which fire reference perimeters are extracted. Each pair of L8/S2 images within the long unit constitutes a "short unit" (i.e. two images: pre- and post-fire images). By using "long validation units", the time period covered by the reference fire perimeters will be sufficiently long to provide a more accurate reporting of the fire date (i.e. date of observation of the burned area). Indeed, this represents an improvement over previous Fire cci phases where the use of "short units" for generating reference fire perimeters had an impact on the accuracy assessment of burn time reporting from moderate resolution product. The following parameters identify a "long validation unit":

1. **Cloud cover** percentage on each scene (CC)
2. **Time step** between consecutive L8/S2 images (Δt)
3. **Length** of the time series (L)

In order to build time series of clear sky conditions, a maximum percentage of cloud cover is allowed for each L8/S2 scene. For the L8 time series the threshold on **maximum cloud cover percentage is set to 30%** (CC_{max}=30%) consistent with previous Fire_cci phases. In the case of S2 time series, the greater frequency of observation available with both A&B Sentinel sensors allowed lowering the maximum percentage threshold value on single scenes to 10%; details on the specific S2 conditions for building "long validation units" are given in Section 4.2.2. Information on cloud coverage percentage is extracted from L8/S2 metadata hence this value refers to the entire scene for L8 and for the S2 tile.

The requirement on image maximum cloud cover (CC_{max}) determines the availability of time series from each source sensor (L8/S2) and for each analysed year depending on local weather conditions (over the validation area).

A maximum time step (Δt_{max}) between consecutive images is set to assure that the burned area signal is detectable even in those ecosystems where persistence of the signal is lower (e.g. savannas and grasslands). A **maximum time step of 16 days (**Δt_{max}=16 days) is generally preferred but it could be increase in very cloudy regions up to **32 days.**

A minimum length of the validation unit (L_{min}**)** is set to guarantee accurate estimation of the date of detection of burns and a good representation of burned polygons. Requirements on Δt_{max} and L_{min} are correlated and dependent on cloud cover and too restrictive conditions could significantly limit the number of validation area available for sampling in each stratum. As presented in the PVIR (Stroppiana et al., 2020), the analysis of L8 and S2 archives of images for the year 2018 and 2019, respectively, highlighted that the choice of a L_{min} threshold value could significantly bias sampling towards less cloudy regions within each biome (i.e. outside the tropical areas) by reducing at the same time the size of the population of suitable validation units.

The specificity of sampling design for global and regional products is detailed in the following sections.

4.2.3 Sampling cardinality

The total number of validation units to be sampled to represent all conditions of biomes and fire intensity for each year is distributed among strata based on Eq. 1.

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

where n_h is the number of units to be sampled for each stratum h, $\overline{BA}_h$ is the average total annual burned area for stratum h and N_h is the total amount of units available for sampling for each stratum h. N_h depends on the parameters set for the definition of the "long validation units" outlined in Section 4.2.2. Particularly cloudy conditions over a certain stratum and for a given year could significantly decrease N_h from the potential population of available validation areas: hence, N_h changes from year to year. For smaller strata a minimum of n_h= 2 is assigned. As explained above, strata are given by the high/low fire intensity partition of each biome (Padilla et al. 2014a; 2015) estimated from annual global FireCCI51 BA product (Chuvieco et al., 2018; Lizundia-Loiola et al., 2020a).

4.2.4 Sampling design

4.2.4.1 Global BA products

For the global BA products, the spatial dimension of validation units will be based on **Thiessen Scene Areas (TSAs)** constructed over the Landsat WRS-2 by Cohen et al. (2010) and Kennedy et al. (2010) and exploited in previous Fire_cci phases (Padilla et al. 2014a; 2015). The number of selected units for each year will be consistent with previous Fire_cci and equal to **100 TSAs y^{-1}** over the period 2018-2020.

The population of TSAs available for sampling for each stratum is given by those validation units that satisfy the requirements for building times series of consecutive L8 images outlined in Section 4.2.2:

1. **Cloud cover:** scene cloud cover ≤ 30% (CC_{max}=30%)
2. **Time step:** time step between pairs ≤ 16 days (Δt_{max}=16 days)
3. **Length:** time series length ≥ 48 days (L_{min}=48 days)

The minimum length of the unit was defined based on the analysis of L8 2018 archive and chosen to assure the least possible biased sampling over biomes significantly affected by cloud cover.

For each sampled TSA, a time series of consecutive pairs of Landsat images are selected to create a "long sampling unit". Cloud cover is cumulated over the long unit since regions covered by clouds in any of the "short units" is assumed as not overserved, hence discarded. In case of persistent cloud coverage over a significant portion of the TSA, the threshold on maximum time step between consecutive clear sky L8 images (Δt_{max}) could be increased (Padilla et al., 2014a).

4.2.4.2 Regional Africa BA products

For the regional Africa BA products, a specific sampling scheme was designed based on the Sentinel-2 imagery tiling system following an approach similar to the one implemented for global sampling (Padilla et al. 2014a; 2015). **Validation units** are identified by spatio-temporal partition of the S2 2019 archive over sub-Saharan Africa (latitude range 25°N- 35°S, Figure 6). Since validation areas cover an area of 100 km x 100 km and S2 tiles are 10,000 km^2 ortho-images in UTM/WGS84 projection, the S2 tiling grid is directly used for spatial partition. However, in order to provide a robust statistical sampling and avoid overlapping between units that could determine an area with higher probability of been selected, two major issues were first addressed and solved.

First, S2 tiles whose footprint crosses different orbits, as shown by yellow polygons in Figure 6a, are discarded. This is done because in overlapping tiles only a portion of the total 100 x100 km area is acquired on a given date; since fires are a dynamic phenomenon, reference perimeters should be extracted from images acquired on the same date. Hence, only S2 tiles fully covered by a single orbit (i.e. the surface observed on the same date) are retained as shown by green polygons in Figure 6b. According to this criterion the valid S2 tiles are keep (Figure 6c).

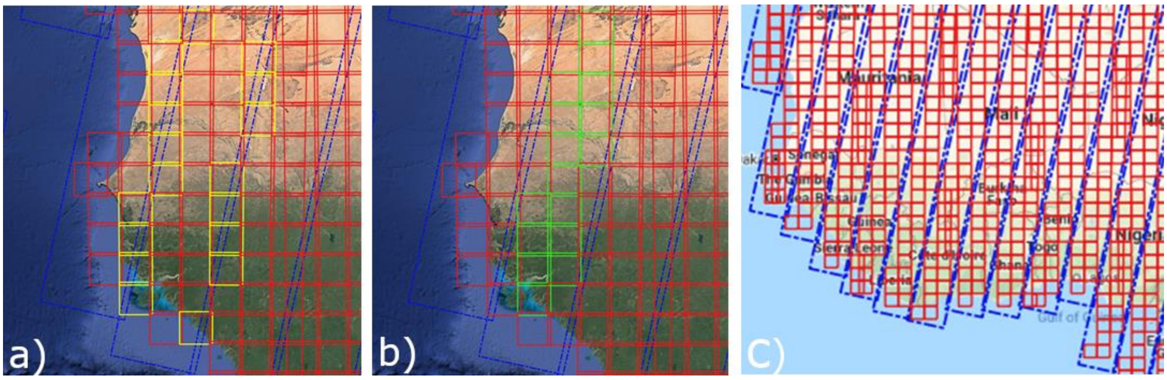

Figure 6: a) example S2 tiles not suitable (yellow) as validation areas for sampling since they cover different orbits (Blue lines); b) example S2 tiles suitable (green) as validation areas for sampling since they cover the same orbit (blue lines); c) all S2 tiles suitable for sampling (after first filtering) in the example area centred over Senegal.

The second issue is generated by S2 tiles overlapping due to adjacent zones of the UTM coordinate projection system (yellow polygons in Figure 7a). In these conditions, keeping both overlapping tiles would increase sampling probability of the common land area; hence, only one S2 tile is randomly selected and retained (Figure 7b).

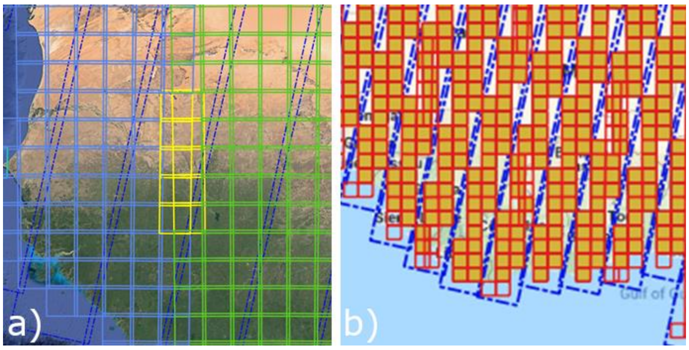

Figure 7: a) example S2 tiles suitable (green/blue) and not suitable (yellow) as validation areas for sampling due to overlap of adjacent UTM zones; b) S2 tiles suitable (orange) for sampling using a random selection in the example area centred over Senegal.

Once the two previous filtering conditions are applied to the S2 tiling grid system, the population of tiles suitable for sampling is composed by the red polygons shown in Figure 8a. In the case of S2 sampling design, filtering applied to S2 tiling system guarantees that validation units are non-overlapping polygons, but polygons are not contiguous as in the global design (Section 4.2.4).

For each tile, i) the total annual burned area from the FireCCI51 BA product Figure 8b and ii) the major Olson biome Figure 8c are computed; in Figure 9 an example is given for the year 2019. The annual FireCCI51 BA product is exploited to split each Olson biome into sub-strata of high and low fire intensity by applying a threshold derived as in section 4.2.4. The threshold values derived for year 2019 are reported in the PVIR (Stroppiana et al, 2020).

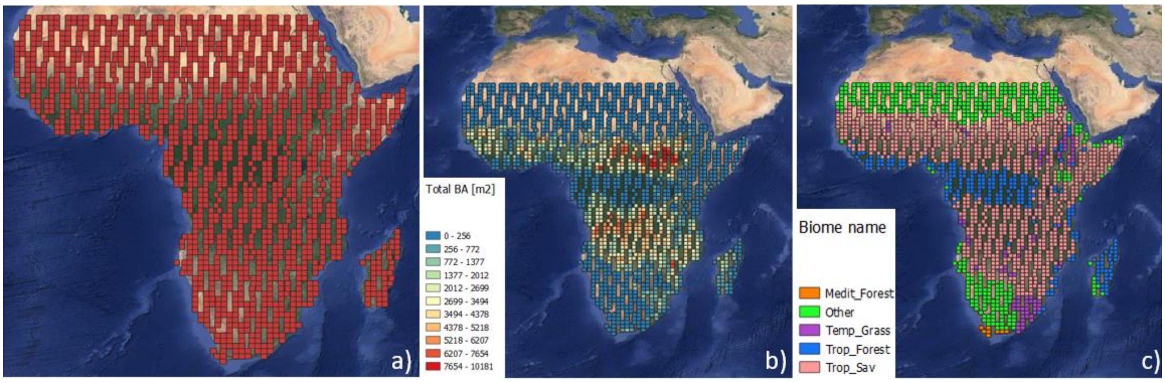

Figure 8: a) S2 tiles available for sampling after applying the filtering criteria; b) total burned area for each S2 [m2]; c) Major Olson biome for each S2 tile.

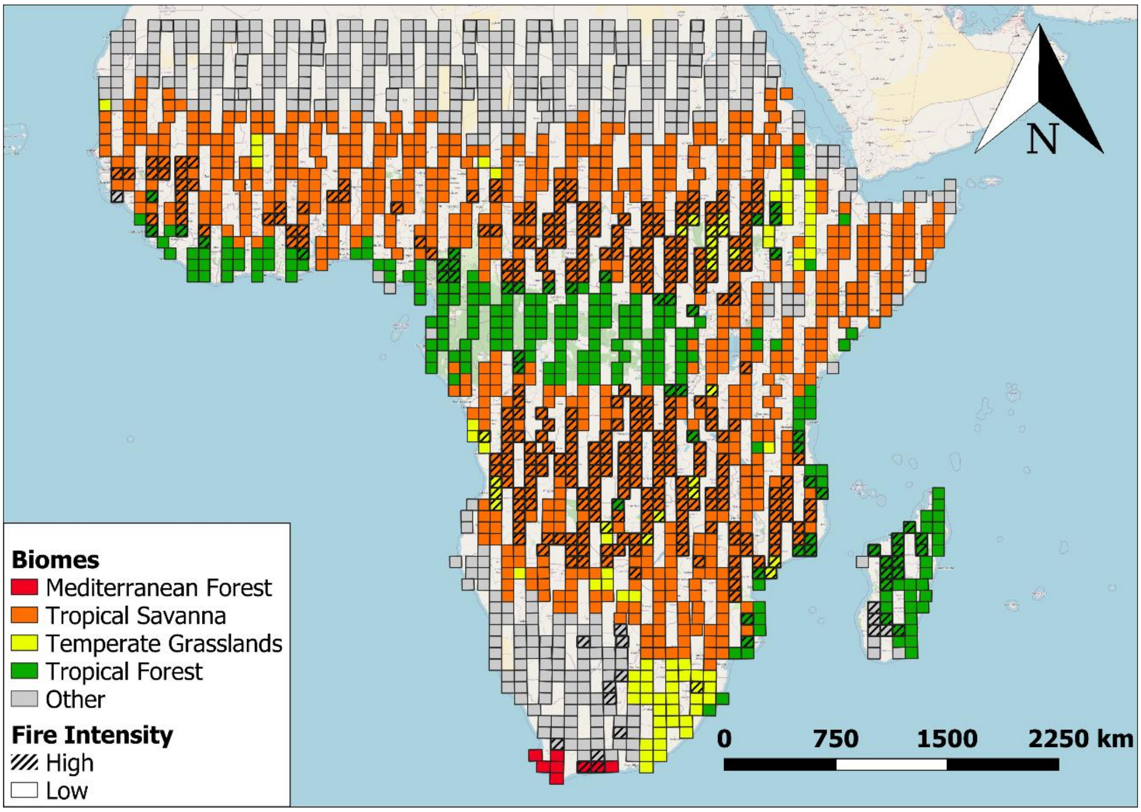

Figure 9: S2 tiles suitable for sampling for each stratum (intersection between biome and high/low fire intensity).

As in the case of the Global BA validation (Section 4.2.4), the combination of biome and fire intensity information defines the strata for a stratified random sampling (Figure 9). The population of S2 tiles available for sampling for each stratum is given by those validation units that satisfy the requirements for building time series of consecutive S2 images.

With respect to the criteria outlined in Section 4.2.2 and applied for building the L8 time series for each validation unit, conditions on cloud cover and minimum length of the validation unit were modified to take full advantage of higher S2 frequency of acquisition (combined A&B sensors). Hence, a S2 "long validation unit" is built so that:

1. **Cloud cover:** scene cloud cover ≤ 10% & cumulated cloud cover ≤ 30% (CC_{max}=10% & $CC_{cum,max}$ = 30%)
2. **Time step:** time step between pairs ≤ 16 dyas (Δt_{max}=16 days)
3. **Length:** time series length ≥ 100 days (L_{min}=100 days)

The conditions set on maximum cloud cover percentage of each single S2 scene and maximum cumulated cloud cover over the S2 time series are necessary to reduce the impact of not observable surface over the long validation unit produced by cumulated cloud cover. Moreover, the frequency of S2 acquisitions also allowed us to set a greater threshold on the minimum length of the validation units (L_{min}) without compromising the spatial distribution of units available for sampling (i.e. bias of geographic distribution towards regions less affected by cloud cover) (Figure 10).

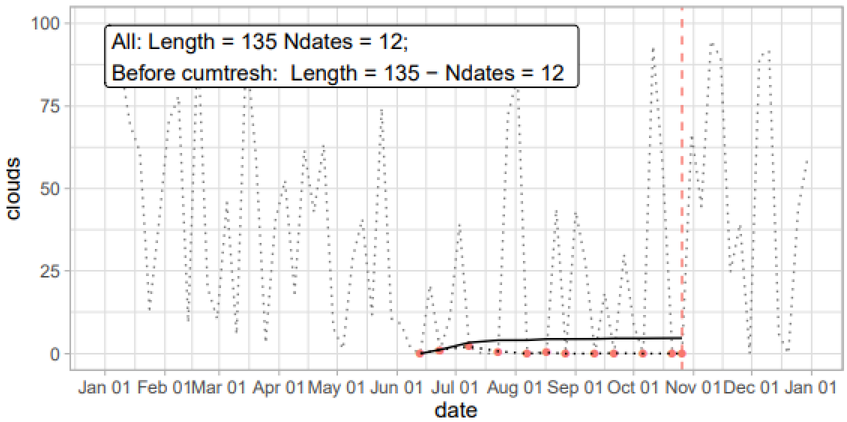

Figure 10. Example of S2 tile (36KWE, Tropical savanna) for the year 2019: single scene cloud cover (dotted line), S2 scenes with cloud cover < 10% (red dots) and date of occurrence of maximum cumulated cloud cover of 30% from first suitable scene (vertical orange line).

S2 tiles that satisfy for the year 2019 the above listed requirements for building a "long validation unit" represent the population of available units for sampling. A total of **50 S2 tiles** are sampled for the year 2019 ("golden year") and distributed among the strata based on Eq. 1. The total number of S2 tiles was decided to balance effort/resources and representativeness of the validation units; notice that "long validation units" extracted from S2 time series are composed of a greater number of "short units" thus requiring greater effort/time for extracting reference perimeters.

4.2.4.3 Local BA products

Over the three study sites selected to test the BA mapping algorithm based on the integration of S1 and S2 (Figure 11), also called SAR-Optical (SAR-O), no stratified sampling design is applicable (stage 1 validation); hence, within each test area, sites for validation are selected to cover different biomes and fire regimes. These sites include an area with persistent cloud cover (where optical data is expected to perform worse than radar), an area with low cloud cover (where the opposite is expected) and a transitional zone. The location of the three test sites is shown in Figure 11, along with information on major biomes of Africa, mean daily cloud cover and annual burned area as estimated by FireCCI51 2016 product. The primary source of data for extracting fire reference perimeters is the Planetscope mission (Planet, 2017).

PlanetScope is a constellation composed by more than 120 optical satellites (also named Doves) operated by Planet from 2016. Each Dove satellite is a CubeSat 3U form factor (10 cm by 10 cm by 30 cm) (https://earth.esa.int/eogateway/missions/planetscope, accessed January 2021). The sensor mounted on this platform is characterized by four bands: three in the visible (b1: 455-515 nm; b2: 500-590 nm; b3: 590-670 nm) and one in the NIR wavelengths (b4: 780-860 nm). Planetscope images have a swath of about 25 km and a spatial resolution of 3 m for all bands. Imagery is captured as a continuous strip of single frame known as "scenes" (Lemajic et al., 2018).

The Planetscope archive (https://www.planet.com, last access March 2021) was searched to identify locations, within the three sites, where temporally consecutive Planet images were available with least cloud cover.

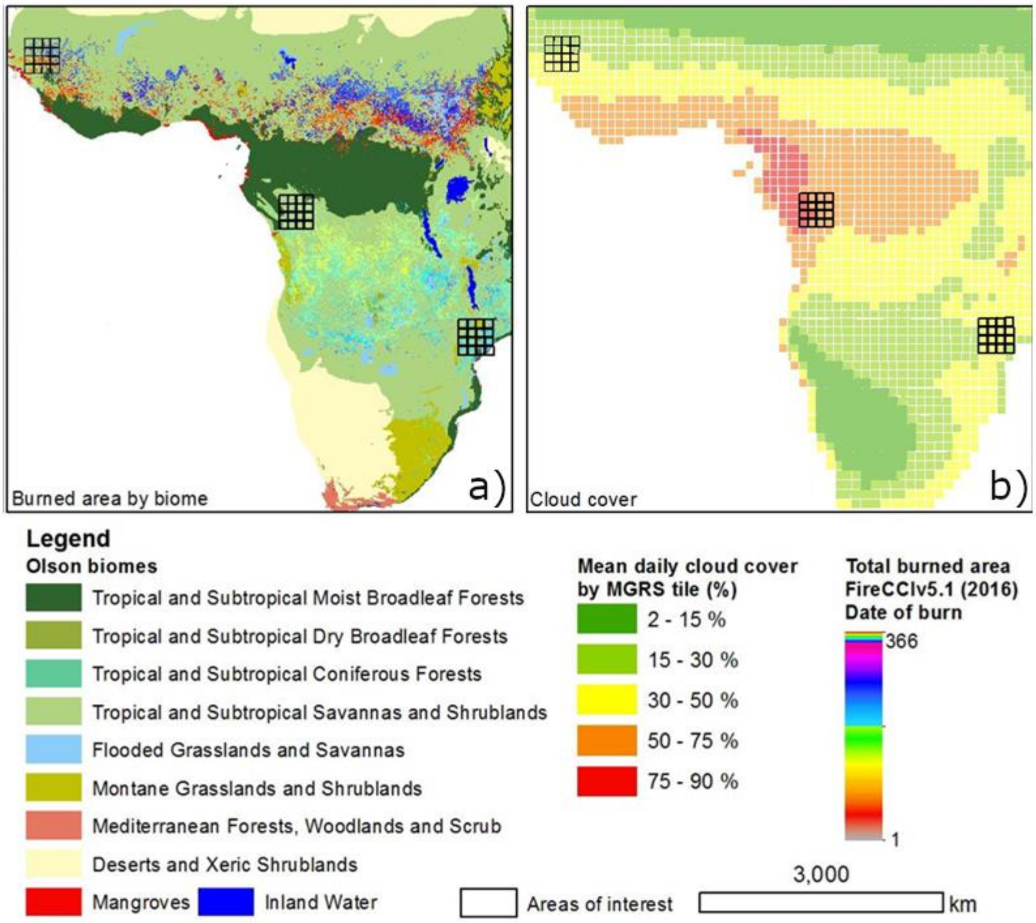

Figure 11: Location of the areas of interest for SAR-O algorithm development showed over the total burned area for year 2016 (pixel level, FireCCI51 product) and the main Olson biomes (Olson et al., 2001) (a), and daily mean cloud cover at tile level (Military Grid Reference System - MGRS) (b).

4.3 Protocol for creating and documenting BA reference files

Figure 12 outlines the flowchart followed for the extraction of fire reference perimeters from the classification of L8/S2 time series over each validation unit selected by applying the sampling schemes described in the above sections (**Step 1**).

A validation area of 100 km x 100 km is drawn centered on L8 frame and S2 tiles (**Step 2**). A supervised classification Random Forest (RF) algorithm (Breiman, 2001) is applied to each pair of L8/S2 images ("short unit") (**Step 3**) to map areas burned between the two dates (t1, t2). Input to the RF classifier are spectral indices NBR, NBR2 and NDVI and their temporal differences (t1, t2) and training polygons for burned/unburned areas selected by photointerpretation of RGB false colour composites.

The use of consecutive L8/S2 images selected with a maximum step Δt_{max} guarantees the greatest separability between burned and unburned surfaces. Training of the RF algorithm is performed by independent interpreters and iteratively to assure the least possible bias and highest accuracy of the output perimeters.

The classification of the short units is then converted to shapefile format and processed to extract fire reference perimeters, clouds and burn date (**Step 4**). All short units over the same validation unit are combined to derive fire perimeters over the "long validation unit" (**Step 5**) and to build attribute table (**Step 6**).

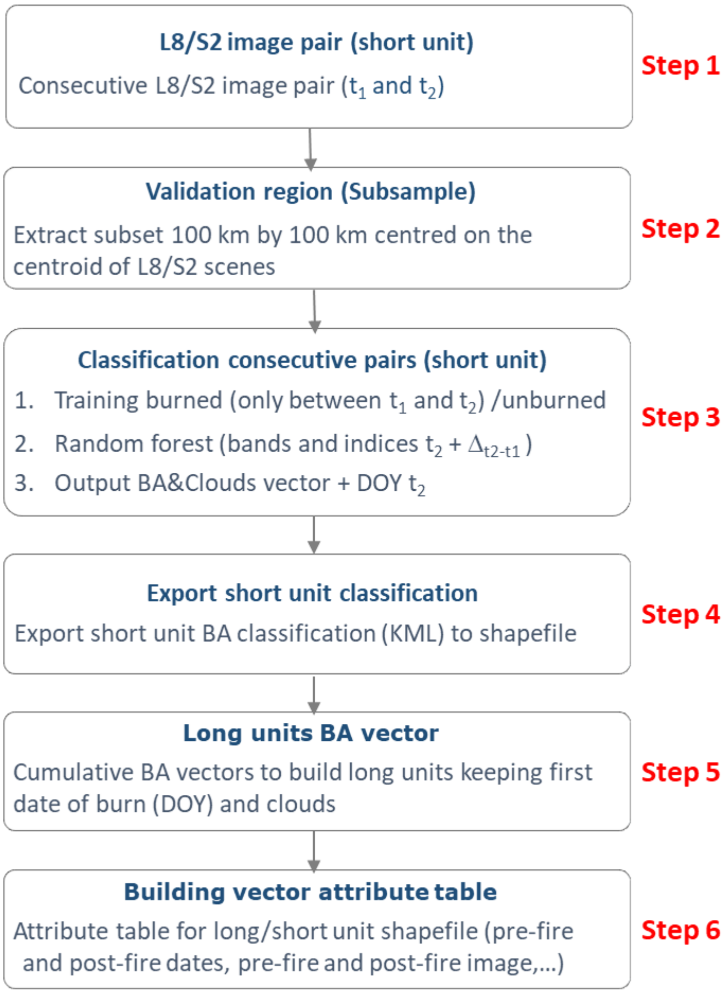

Figure 12. Flowchart showing steps for extraction of fire reference perimeters over sampled validation units.

BA reference files are delivered as ESRI© shapefiles and contain, besides the burned and unburned categories, unobserved areas or not valid pixels (i.e. regions where the surface conditions could not be observed and assigned to neither of the two burned and unburned categories). These areas are generally related to the presence of clouds and cloud shadows, and other factors affecting atmospheric conditions (e.g. very thick smoke plumes). Not valid pixels in the reference image datasets are identified from mask and quality flags provided with the L8/S2 source data. Data structure and naming convention for reference files are described in details in Annex 3.

Steps 1 to 4 of the workflow in Figure 12 are implemented in the Earth Engine (EE) Code Editor (code.earthengine.google.com, last access March 2021) for processing L8/S2 short units (RF classification). Steps 5 and 6 are implemented with a Python script coded in ArcGIS (Arcpy).

For the three sites in Africa selected for the SAR-O algorithm, reference fire perimeters are generated from multiple consecutive pairs of VHR Planet images with the same approach followed for L8 and S2. The RF algorithm is in this case implemented in an R script.

5 Description of validation activities

Validation activities are the processing steps necessary for the implementation of the protocol described above. For pixel-based Fire cci BA products (namely FIRECCI51, FireCCI60, FIRECCIS311, and FIRECCISFD20, Table 1), major activities are:

1. Identification of statistical representative validation areas (spatio-temporal sampling);
2. Creation of reference fire perimeters;
3. Processing of Fire_cci BA products for building “comparable” dataset;
4. Performing cross-tabulation (i.e. confusion matrix);
5. Conducting error analysis and multi-temporal product performance analysis (only for 2017-2019 period).

5.1 Sampling scheme implementation and reference EO data

Reference fire perimeters for validation of **global Fire_cci BA products** are derived from Landsat 8 data. The sampling criteria and scheme are detailed in section 4 and are applied to select, for each year, a total of 100 TSAs for building the “long sampling units”. Each sampling unit is composed of consecutive pairs of pre- and post-fire images separated by 16 days. Exceptionally, in case of lack of pairs satisfying the above criteria, it could be considered to extend the time lag or to shorten the length of the “long sampling units”. Still, preliminary analysis of the L8 archive for the year 2018 showed that a maximum step of 16 days can be maintained in the “long units” when the requirement on the minimum length of the unit is set to 48 days.

For the **African regional BA product** (FireCCISFD20), fire reference perimeters are created for 50 validation units by using Sentinel-2 data and the specific sampling design outlined in section 4.2.4.2. Although BA fire perimeters are derived from the same dataset used for deriving Fire_cci BA product, the validation protocol assures the extraction of independent reference perimeters (different algorithms) with the highest accuracy possible (operator supervision), hence ensuring to be as close as possible to the ground truth.

For the SAR-O **BA product obtained over the three study sites in Africa** and focusing on testing the **integration of Sentinel-1 and Sentinel-2** data, no stratified sampling can be implemented since VHR Planetscope data do not assure systematic acquisitions. Since sites are selected only for algorithm testing purposes, validation is carried out at stage 1 (Morisette et al. 2006). As for validation of global and regional products, reference fire perimeters are extracted from time series of consecutive pairs of Planet images representing pre- and post-fire conditions.

5.2 Processing steps for creating reference BA data

A supervised semi-automatic classification algorithm is applied to all the above listed product categories to consecutive pairs of reference images (L8, S2 and Planet) to identify areas burned between pre- and post-fire dates, as described in Section 4.3. Training data is extracted for image pair(s) by photointerpretation of RGB false colour composites.

The semi-automatic classification is applied at the pixel level (20-30 m spatial resolution for L8/S2 and 3 m for Planet). Output raster are converted to ESRI© shapefile. In the case of VHR Planet data, cleaning of small polygons is carried out to reduce the impact of isolated pixels, geo-location errors and different spatial resolutions.

ESRI© shapefiles of reference fire perimeters are uploaded in a GIS environment to fulfil the task of final visual check by photointerpretation and comparison with false colour composites of EO image data bands. Shapefiles of consecutive image pairs are summed up over the time period covered by first and last date of the time series of reference EO source data: the "long sampling unit" for L8/S2, and first and last date for each selected site where Planetscope images are available. These files represent the fire reference perimeters for each validation unit/site and they are input for the cross-tabulation with the Fire_cci BA product to be validated.

ESRI© shapefiles of reference fire perimeters are accompanied by metadata including information about geographic projection, year of detection, information on the original input images (first and last dates/images), name of interpreter, institution, etc. according to Fire_cci standards.

5.3 Processing steps for Fire_cci BA product extraction

The Fire_cci BA products to be validated undergo the following processing steps before being compared to reference fire perimeters for cross-tabulation:

- Projection conversion: Fire_cci BA products are re-projected to UTM, WGS84, with the UTM zone coincident with the one covered by the major part of the reference L8/S2/Planet images assigned to each validation area/site;
- Clipping to the sub-sampled window size (100 km by 100 km) centred over the validation unit(s), if necessary;
- Temporal sampling of Fire_cci BA between pre- and post-fire dates of the reference data (e.g. first and last date of the "long sampling unit" in the case of L8/S2);

5.4 Cross-tabulation and accuracy metrics

Reference and Fire_cci BA products are overlaid and intersected to extract the confusion matrices; each cell of the confusion matrix e_{ij} (Table 2) expresses the proportion of area of agreement or disagreement for all pixels of the validation area between the Fire_cci BA product (map) and the reference classes (Padilla et al. 2014a; Padilla et al., 2017).

Confusion matrices are summed up over all of the validation units and, if applicable (i.e. multi-years BA products) for each year, to compute the accuracy metrics listed in Table 2.

5.5 Error analysis

Part of the validation activities is the analysis of the accuracy metrics as a function of those factors that are affecting the performance of BA detection: e.g. land cover, vegetation density, climate-ecosystem zones and burned patch size.

When BA products encompass multiple years, the algorithm should perform consistently and detect fire-affected areas for fire seasons of high, medium and low occurrence across the years. In the specific case of multi-year Fire_cci products, accuracy metrics over the 2018-2020 period will be presented and discussed in the Product Validation and Inter-comparison Report (PVIR). Moreover, consistency with accuracy estimates from previous Fire_cci phases will be analysed.

6 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.

Breiman, L. (2001). Random Forests. Machine Learning, 45, 5-32.

Chuvieco, E., Opazo, S., Sione, W., Valle, H. d., Anaya, J., Bella, C. D., Cruz, I., Manzo, L., López, G., Mari, N., González-Alonso, F., Morelli, F., Setzer, A., Csiszar, I., Kanpandegi, J. A., Bastarrika, A. and Libonati, R. (2008), Global burned-land estimation in Latin America using MODIS composite data. Ecological Applications, 18: 64-79. doi:10.1890/06-2148.1.

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.

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

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.

European Space Agency (2007). Earth Observation Satellites: A Roadmap For Calibration And Validation. ScienceDaily. ScienceDaily, 29 October 2007. Available at: www.sciencedaily.com/releases/2007/10/071022120154.htm.

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 System Science Data 12, 3229-3246, https://doi.org/10.5194/essd-12-3229-2020

Giglio, L., Loboda, T., Roy, D.P., Quayle, B., & Justice, C.O. (2009). An active-fire based burned area mapping algorithm for the MODIS sensor. Remote Sensing of Environment, 113, 408-420.

Heil, A., Yue, C., Mouillot, F., & Kaiser, J.W. (2017). ESA CCI ECV Fire Disturbance: D1.1 User requirements document, version 5.2. Available at: https://www.esa-fire-cci.org/documents.

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

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.

Lemajic, S., Vajsová, B. and Aastrand, P., 2018. New sensors benchmark report on PlanetScope: Geometric benchmarking test for Common Agricultural Policy

(CAP) purposes, EUR 29319 EN, Publications Office of the European Union, Luxembourg, 2018, ISBN 978-92-79-92833-8, doi:10.2760/178918, JRC111221.

Lizundia-Loiola, J., Otón, G., Ramo, R. and Chuvieco, E. (2020a). 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

Lizundia-Loiola J., Otón G., Belenguer-Plomer M.A., Tanase M.A., Pettinari M.L., Chuvieco E. (2020b) ESA CCI ECV Fire Disturbance: D2.2 End to End ECV Uncertainty Budget, version 1.1. Available at: https://www.esa-firecci.org/documents.

Morisette, J.T., Baret, F. Privette, J.L. et al. (2006). Validation of global moderate-resolution LAI products: a framework proposed within the CEOS land product validation subgroup. IEEE Transactions on Geoscience and Remote Sensing, 44(7). 1804-1817, doi: 10.1109/TGRS.2006.872529.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., & Kassem, K.R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. BioScience, 51, 933-938.

Otón, G., Lizundia-Loiola J., Pettinari, M.L., Chuvieco, E. (2021) Development of a consistent global long-term burned area product (1982-2018) International Journal of Applied Earth Observations and Geoinformation. In revision.

Padilla, M., Stehman, S.V., & Chuvieco, E. (2014a). Validation of the 2008 MODIS-MCD45 global burned area product using stratified random sampling. Remote Sensing of Environment, 144, 187-196.

Padilla, M., Stehman, S.V., Litago, J., & Chuvieco, E. (2014b). Assessing the temporal stability of the accuracy of a time series of burned area products. Remote Sensing, 6, 2050-2068.

Padilla, M., Chuvieco, E., Hantson, S., Theis, R., Sandow, C. (2014c) ESA CCI ECV Fire Disturbance: D2.1 Product Validation Plan, version 4.0. Available at: https://www.esa-fire-cci.org/documents.

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

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.

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

Planet Team. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. https://api.planet.com, 2017.

Plummer, S., Arino, O., Ranera, F., Tansey, K., Chen, J., Dedieu, G., Eva, H., Piccolini, I., Leigh, R., Borstlap, G., Beusen, B., Fierens, F., Heyns, W., Benedetti, R., Lacaze, R., Garrigues, S., Quaife, T., De Kauwe, M., Quegan, S., Raupach, M., Briggs, P., Poulter, B., Bondeau, A., Rayner, P., Schultz, M., & McCallum, I. (2007). An update on the GlobCarbon initiative: multi-sensor estimation of global biophysical products for global terrestrial carbon studies. In, Envisat Symposium 2007. Montreux, Switzerland.

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.

Roy, D.P., Boschetti, L., Justice, C.O., & Ju, J. (2008). The collection 5 MODIS burned area product - Global evaluation by comparison with the MODIS active fire product. Remote Sensing of Environment, 112, 3690-3707.

Roy, D.P., Boschetti, L. (2009). Southern Africa validation of the MODIS, L3JRC, and GlobCarbon burned-area products. IEEE Transactions on Geoscience and Remote Sensing, 47, 1032-1044.

Stroppiana D., Boschetti M., Busetto L., Luini C., Sali M., Franquesa M. Lizundia-Loiola J. (2020) ESA CCI ECV Fire Disturbance: D4.1 Product Validation and Intercomparison Report, version 1.1. Available at: https://climate.esa.int/en/projects/fire/key-documents/

Tansey, K., Grégoire, J.-M., Defourny, P., Leigh, R., Pekel, J.-F., Bogaert, E., & Bartholome, E. (2008). A new, global, multi-annual (2000-2007) burnt area product at 1 km resolution. Geophysical Research Letters, 35, L01401, doi:10.1029/2007GL03156.Vanderhoof, M.K., Fairaux, N., Beal, Y.-J. G., Hawbaker, T.J. (2017). Validation of the USGS Landsat Burned Area Essential Climate Variable (BAECV) across the conterminous United States, Remote Sensing of Environment, 198, 393-406, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2017.06.025.

Annex 1 Acronyms and abbreviations

BA	Burned Area
CBERS	China-Brazil Earth Resources Satellite
CC	Cloud Cover
Ce	Commission error ration
CEOS	Committee on Earth Observation Satellites
CSDGM	Content Standard for Digital Geospatial Metadata
DC	Dice Coefficient
ECV	Essential Climate Variables
EE	Earth Engine
EO	Earth Observation
ESA	European Space Agency
ESRI	Environmental Systems Research Institute
FI	Fire Intensity
HR	High Resolution
ISO	International Organization for Standardization
L	Length
L3JRC	Global Multi-year (2000-2007) Validated Burnt Area Product
L8	Landsat 8
LPVS	Land Product Validation Subgroup of CEOS
OLI	Operational Land Imager
MCD45	MODIS Collection 5 Burned Area product using the Roy et al. (2008) algorithm
MCD64	MODIS Collection 5 Burned Area product using the Giglio et al. (2009) algorithm
MODIS	Moderate Resolution Imaging Spectroradiometer
NBR	Normalized Burn Ratio
NDVI	Normalized Difference Vegetation Index
NIR	Near InfraRed
Oe	Omission error ration
OLI	Operational Land Imager
PVIR	Product Validation and Inter-comparison Report
relB	Relative bias
RF	Random Forest
S1	Sentinel-1
S2	Sentinel-2
SAR	Synthetic Aperture Radar
SAR-O	Combination of SAR and Optical data
SLC	Scan Line Corrector
SFD	Small Fire Database
SWIR	Short Wave InfraRed
TM	Thematic Mapper
TotBA	Total annual burned area
TSA	Thiessen Scene Area
UTM	Universal Transverse Mercator
VHR	Very High Resolution
WGS84	World Geodetic System 1984
WRS(-2)	Worldwide Reference System (version 2)
XML	eXtensible Markup Language

Annex 2 Accuracy metrics

Commonly in BA validation, accuracy estimates are based on the cross-tabulation approach (Congalton and Green 1999; Latifovic and Olthof 2004). The result of the cross tabulation can be represented by the error matrix (Table 2) which expresses the proportion of agreements or disagreements between product and reference classifications. The error matrix is derived for each validation unit (L8 TSAs, S2 tiles and Planet images); matrix entries e_{ij} express the agreements (i=j) or disagreements (i≠j) for all pixels of the validation unit between the BA product (map) class and the reference data summed up over the total number of validation units.

Table 2. 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}	

Accuracy metrics are computed from the global error matrix with the equations listed in Table 3.

Table 3. Accuracy metrics computed from the error matrix

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

Annex 3 Data structure and naming convention

Reference fire perimeters are delivered as ESRI shpeafiles © (.shp), along with the auxiliary files required (.dbf, .prj, shx, .sbn, .xml). The projection is UTM, WGS84, with the UTM zone being the zone that is covered by the major part of the scene.

The following attribute fields are included in the shape file (Table 5):

- PreDate. Acquisition date of the image taken before the occurrence of the fire: yyyymmdd (year, month, day).
- PostDate. Acquisition date of the satellite image taken after the fire: yyyymmdd (year, month, day).
- PreImg and PostImg. The pre- and post-fire image names, following this format: satellitecodePathRow (e.g. LE719905). The satellite codes are given in Table 4.
- Area (in square metres, m^2)
- Category (Observation category):
  - Burned area = 1. This area includes all polygons detected as burned.
  - No-Data = 2. This area includes all polygons that could not be interpreted or were not observed by the sensor, either by clouds and/or cloud shadows, topographic shadows, smoke, or sensor errors (for instance, those caused by SLC-off problems of ETM+)
  - Unburned = 3. This area includes all polygons observed as not burned within the limits of the area covered by the image.

Table 4. Satellite-sensor codes naming convention

Satellite-sensor	Mission Code (MMM)	Reference system	
Landsat-8 OLI	LC8	Path (ppp)	Row (rrr)
Sentinel-2A	S2A	Relative orbit Number ROOO	Tile Number field (Txxxxx)
Sentinel-2B	S2B	Relative orbit Number ROOO	Tile Number field (Txxxxx)

Table 5. Example of attribute table for BA reference fire perimeter shapefile

	category	preDate	postDate	preImg	postImg	path	row	year	area
1	2	2016-01-25	2016-02-02	LE71990502016025ASN00	LC81990502016...	199	50	2016	69300.00000000...
2	2	2016-01-25	2016-02-02	LE71990502016025ASN00	LC81990502016...	199	50	2016	544500.0000000...
3	2	2016-01-25	2016-02-02	LE71990502016025ASN00	LC81990502016...	199	50	2016	159300.0000000...
4	2	2016-01-25	2016-02-02	LE71990502016025ASN00	LC81990502016...	199	50	2016	525600.0000000...
5	2	2016-01-25	2016-02-02	LE71990502016025ASN00	LC81990502016...	199	50	2016	177300.0000000...
6	2	2016-01-25	2016-02-02	LE71990502016025ASN00	LC81990502016...	199	50	2016	506700.0000000...
7	2	2016-01-25	2016-02-02	LE71990502016025ASN00	LC81990502016...	199	50	2016	205200.0000000...
8	2	2016-01-25	2016-02-02	LE71990502016025ASN00	LC81990502016...	199	50	2016	485100.0000000...
9	2	2016-01-25	2016-02-02	LE71990502016025ASN00	LC81990502016...	199	50	2016	217800.0000000...
10	2	2016-01-25	2016-02-02	LE71990502016025ASN00	LC81990502016...	199	50	2016	465300.0000000...
11	2	2016-01-25	2016-02-02	LE71990502016025ASN00	LC81990502016...	199	50	2016	223200.0000000...

The name of the .shp and associated files is defined as follows:

PRO_RD_ppprrr _yyyymmdd_ yyyymmdd (Landsat)

PRO_RD_Txxxxx_yyyymmdd _ yyyymmdd (Sentinel)

where:

PRO: project where the reference data were generated. For the fire perimeters developed within the Fire_cci project, PRO=FireCCI.

RD: stands for Reference Data

ppprrr: represents the Landsat Worldwide Reference System (WRS) path and row of the scene (in the case where no Landsat imagery was used, the closest path-row is selected): ppp=path; rrr=row.

Txxxxx: represents the Sentinel-2 100x100 km Tile Number field.

yyyymmdd (year, month, day): the first one is the date of the first image used for BA detection; the second one is the date of the last image used for generating the reference fire perimeters.

The metadata of the reference files is written as an XML document following the international CSDGM and ISO 19115 standards. The metadata contains fields to cover all necessary information to be provided to external users: author names of the reference data file, affiliations/institutions, date of creation, the input data sources (names of satellite image files) and the reference of the website of the Fire_cci project.