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	20/12/2019
Version	1.1
Author	Daniela Stroppiana, Mirco Boschetti
Document Ref.	Fire_cci_ D1.3_PVP_v1.1
Document type	Public

To be cited as: Stroppiana D., Boschetti M. (2019) ESA CCI ECV Fire Disturbance: D1.3 Product Validation Plan, version 1.1. Available at: https://www.esa-fire-cci.org/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	Stephen Plummer (ESA)	stephen.plummer@esa.int	electronic copy
Project Team	Emilio Chuvieco (UAH) M. Lucrecia Pettinari (UAH) Joshua Lizundia (UAH) Gonzalo Otón (UAH) Mihai Tanase (UAH) Miguel Ángel Belenguer (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 miguel.belenguer@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 to be 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 Mirco Boschetti	26/06/2019
Reviewed	UAH – Project Manager	Lucrecia Pettinari	01/07/2019
Authorized	UAH - Science Leader	Emilio Chuvieco	01/07/2019
Accepted	ESA - Technical Officer	Stephen Plummer	20/12/2019

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

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

Table of Contents

List of Tables

List of Figures

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 that will be implemented in the Fire_cci builds on standard methods published in the literature (Padilla et al. 2017) and developed during previous Fire_cci phases (Fire_cci background). To this aim, this document reviews validation methods and protocols from Fire_cci background and the literature from the EO (Earth Observation) research community (Section 3). Sections 4 and 5 describe the protocol that will be implemented for validation and the activities to be carried out, respectively.

2 Introduction and objectives

The *Product Validation Plan* (PVP) describes the approaches and methods that will be used to assess the quality of burned area (BA) products obtained by applying the Fire_cci algorithms (Table 1).

The products object of validation are global and regional (Africa) covering one year (2019) or multi-annual (2017-2019) (Table 1). Furthermore, BA algorithms for Sentinel-1 and Sentinel-2 data will be developed and tested over three study areas in Africa. Validation is required after delivery of the products to assess their accuracy.

Table 1. BA products foreseen in the Fire_cci project, time period covered (in brackets), source data (Mission&sensor) and the years' object of validation.

Products	Sensor	Area	2017	2018	2019
FireCCI51 (2017-2019)	Terra-MODIS	**Global**	**X**	**X**	**X**
FireCCILT20 (1982-2019)	NOAA AVHRR-LTDR	**Global**	**X**	**X**	**X**
FireCCIS311 (2019-2020)	S-3 OLCI+SLSTR	**Global**			**X**
FireCCI60 (1982-2019)	Merged AVHRR-MODIS-OLCI-SLSTR BA	**Global**	**X**	**X**	**X**
FireCCISFD20 (2019)	S2 MSI	**Sub-Saharan Africa**			**X**
FireCCIS1S2AF10 (2019)	S-1 SAR and S-2 MSI	**Selected sites in Africa**			**X**

In the case of the FireCCI51, FireCCILT20 and FireCCI60 products, which cover a longer time period, validation will be focused on years 2017 to 2019. For the other products validation will cover 2019.

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 May 2019). 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 images of medium spatial resolution of around 10-30 m. At present, the major data source used for this purpose is Landsat mission with imagery acquired by TM/ETM+ and OLI sensors (Padilla et al., 2015; Vanderhoof et al., 2017; Roteta et al., 2019). In the next future Sentinel-2 data will complement Landsat data for validation of regional/global BA products.

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.

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.

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 (Figure 1). Temporal stability of the accuracy was also computed for multi-annual products.

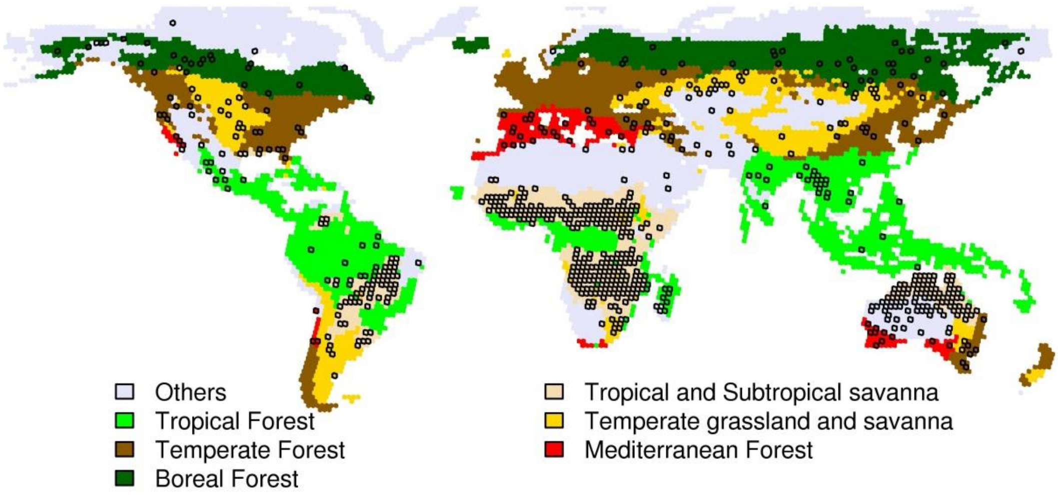

Figure 1: Thiessen scene areas (TSAs) with at least one unit selected in the sample and biome stratification based on a reclassification of the 14 Olson biomes (Olson et al. 2001).

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.

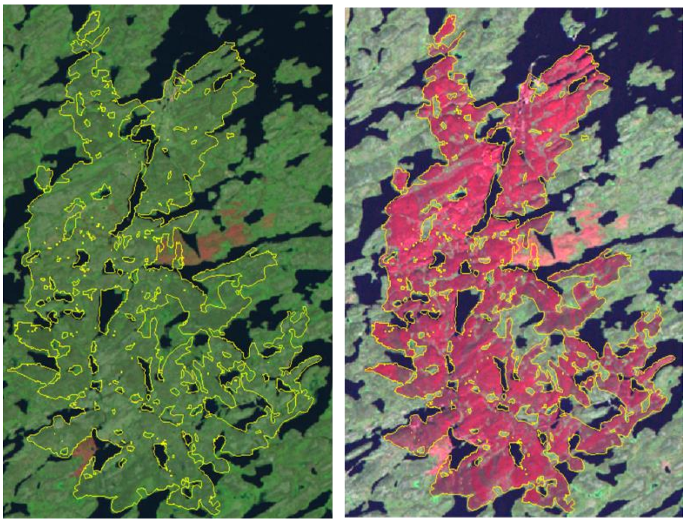

Figure 2: Example of reference data derived from classification and visual refinement of Landsat image pair over Canada. Burned area is highlighted by the yellow polygon.

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 S-2 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 S-2 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).

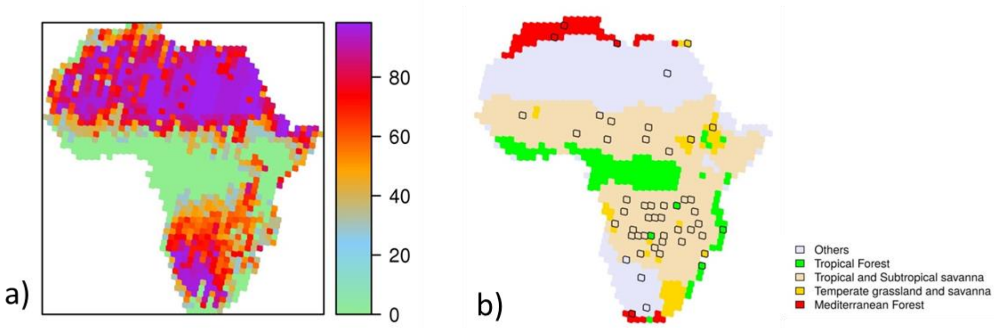

Figure 3: a) Spatial distribution of reference data availability for long sampling units shown as percentage of time; b) Thiessen scene areas (TSAs) with at least one unit selected in the sample and biome stratification based on a reclassification of the 14 Olson biomes (Olson et al. 2001).

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).

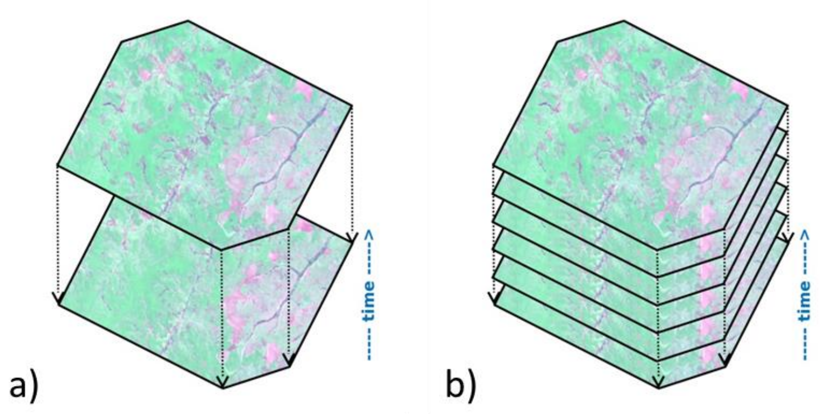

Figure 4: Illustration of short (a) and long (b) sampling units for a Thiessen scene area (TSA) on a three-dimensional space. Each sampling unit is delimited spatially by a TSA (two-dimensions) and temporally (the third dimension) by the time between two or more consecutive Landsat images. Images are 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 will be carried out for the three major categories of products that will be delivered as listed in Table 1: global BA products (FireCCI51, FireCCILT20, FireCCIS311, and FireCCI60), regional for Africa (FireCCISFD20) and BA products for three test sites in Africa derived from the combined use of S-1 and S-2. 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 5, along with information on major biomes of Africa, mean daily cloud cover and annual burned area as estimated by FireCCI51 2016 product.

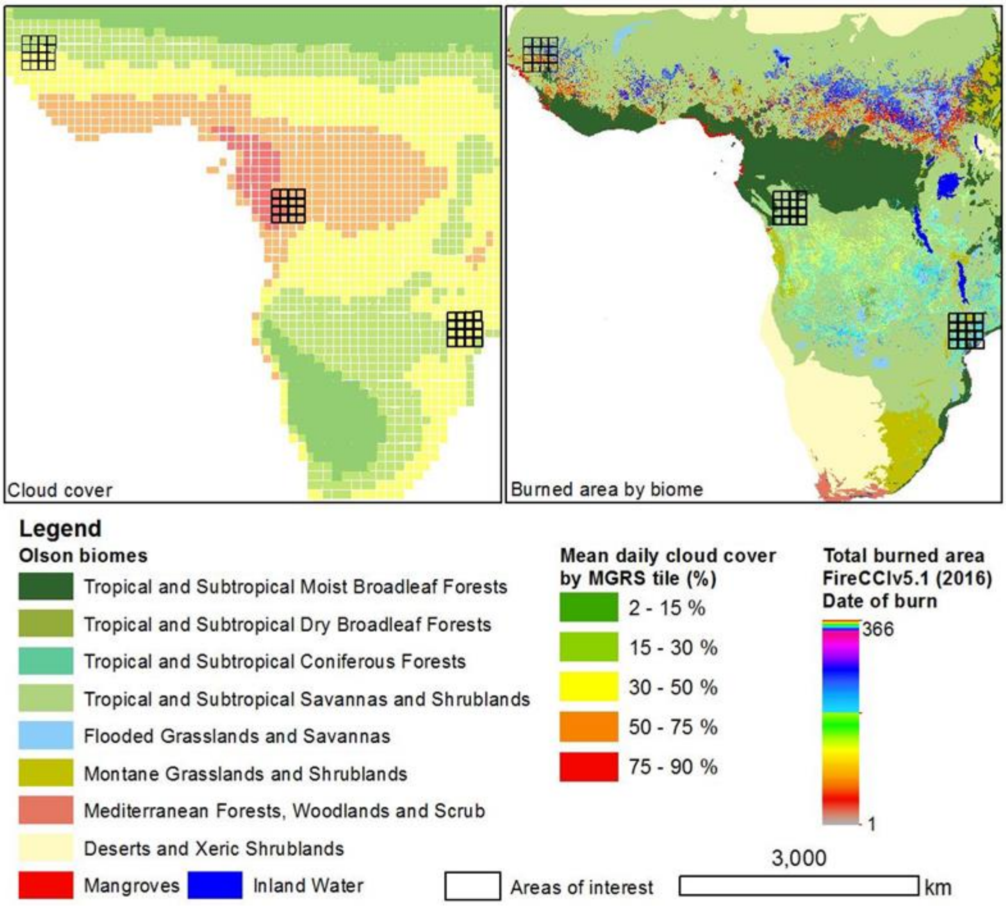

Figure 5: Location of the proposed areas of interest for SAR-O algorithm deployment. The daily mean cloud cover at tile level (Military Grid reference System - MGRS), the total burned area for year 2016 (pixel level, FireCCI51 product) and the main Olson biomes (Olson et al., 2001) are also displayed.

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.

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.

Moreover, fire perimeters are required to be as accurate as possible to be assumed as ground truth; this objective is generally achieved by visual refinement of fire perimeters derived with (semi-)automatic supervised classification algorithms.

Finally, independence is a critical characteristic of any validation assessment, since it assures that unbiased accuracies are obtained among products. Independence implies that validation datasets are not used during the design of BA algorithms, either for calibration or “tuning” processes. It is acceptable that EO 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.

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

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

4.2 Criteria to select validation sites and sampling scheme

Sampling is defined as the temporal and spatial selection of validation sites/units for the estimation of accuracy metrics. Sampling is critical in any validation, to make the most of the resources dedicated to generate reference data; moreover, a sampling scheme should be based on random and representative selection of units to assure unbiased collection of reference data. In Fire_cci, the spatial dimension of sampling units will be defined by contiguous non-overlapping areas/frames while the temporal dimension will be defined by the dates of (consecutive) pair(s) of medium resolution remotely sensed images.

Since sampling sites should be selected as to properly represent the variety of conditions that affect the accuracy of BA cartography, both in time and space, a stratified random sampling scheme will be adopted. In particular, following Padilla et al. (2017), stratification will be based on i) major Olson biomes (Olson et al. 2001) and ii) considering areas with high and low fire activity, as derived within each biome from operational products available at the global scale; FireCCI51 (Chuvieco et al., 2018) BA product will be the primary choice and MODIS-MCD64A1 (Giglio et al. 2009; Giglio et al., 2018) which will be used on as back-up solution. The specificity of sampling design for global and regional products is detailed in the following sections.

4.2.1 Global sampling design

For the global BA products, the spatial dimension of sampling units will be based on Thiessen Scene Areas (TSAs) constructed over the Landsat WRS-2 by Cohen et al. (2010) and Kennedy et al. (2010) and exploited in previous Fire_cci phases (Padilla et al. 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 2017-2019 (Chuvieco et al. 2018).

For each sampled TSA, a time series of consecutive pairs of Landsat images will be selected to create a "long sampling units". Each pair of Landsat images within the long unit constitutes a "short unit" (i.e. two images: pre- and post-fire images). By using "long sampling 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 represent an improvement over previous Fire cci phases where the use of "short units" for generating reference fire perimeters had an impact on the accuracy of time reporting. In case of persistent cloud coverage over a significant portion of the TSA, distance between consecutive clear sky Landsat images could be increased (Padilla et al., 2014a). When temporal distance between cloud-free images is long enough as to be uncertain on whether the area has been affected or not by fire (particularly in Tropical regions, with fast recovery rates), the area will be considered as non-observed.

For each sampled TSA an area of approximately 100 km by 100 km will be selected for extracting reference fire perimeters; with respect to previous Fire cci, the area has been increased to characterize a larger portion of the TSA and to allow more accurate analysis of fire size and patches.

4.2.2 Regional sampling scheme: Africa continent

For the regional Africa BA products, a specific sampling scheme will be designed based on Sentinel-2 imagery. Contiguous non-overlapping areas will be defined, based on S-2 orbit/tile characteristics, as previously done for Landsat frames (Padilla et al. 2014a; 2015). Following Roteta et al. (2019), a total of **50 TSAs** will be sampled for the year to be validated ("golden year").

For each sampled TSA, a S-2 frame will be univocally assigned and multiple consecutive pairs of S-2 images will be used ("long sampling unit") by adopting the same criteria as for previous Fire_cci phases (Padilla et al., 2014c; Padilla et al., 2018). As for global BA, the use of "long sampling unit" reduces the impact of errors in fire time reporting.

For each sampled TSA, an area of approximately 100 km by 100 km will be selected for defining reference fire perimeters; with respect to previous Fire cci, the area has been increased to characterize a larger portion of the sampled TSA and to allow more accurate analysis of fire size and patches.

4.2.3 Integration study sites sampling

Over the three study sites for integration of S-1 and S-2 (Figure 5), no stratified sampling design is applicable; hence, within each test area, validation sites will be selected to cover different fire and land cover conditions. Over the selected validation sites, the availability of suitable HR/VHR EO data will be inspected for spot/local evaluation of the Fire_cci BA products. To this aim, archives of data, such as PLANET and SPOT, available free of charge for the scientific community, will be searched to collect consecutive clear sky images to be used for generating HR/VHR fire reference perimeters. HR/VHR images collected should ideally cover 2/3-month periods with consecutive pairs. The location of the validation areas within each study site will depend on data availability and no sampling design will be implemented (stage 1 validation).

4.3 Protocol for creating and documenting BA reference files

Reference fire perimeters will be extracted by with semi-automatic mapping of burned areas: a machine learning algorithm (i.e. Random Forest classifier) (Breiman, 2001) followed by a systematic quality check performed by visual inspection and refinement. Training of a semi-automatic supervised classification algorithm and visual refinement of fire perimeters will be performed by independent interpreters to assure the least bias and high accuracy. Visual refinement will be carried out in a GIS environment.

Blind tests over a sub-sample of the selected TSAs will be carried out to investigate robustness of generated reference fire perimeters.

BA reference files will be delivered as ESRI© shapefiles and will 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 due to the presence of clouds and cloud shadows, sensors issues (i.e. SLC-off problems of ETM+), and other factors (e.g. very thick smoke plumes). Not valid pixels in the reference image datasets could be derived from mask and quality flags provided with the source data, (semi-)automatic classification as well as from photo-interpretation during the visual refinement.

Fire perimeters below a minimum size/group of pixels will be discarded to reduce the impact of noise, different resolutions of input data and geometric and positional errors. This also applies to unburned patches inside burned areas and to non-observed areas by the reference EO images.

Data structure and naming convention for reference files are described in details in Annex 3.

In particular, for both global and regional BA products, reference fire perimeters for each sampling unit will be generated from a time series of consecutive images (thus building up to a “long sampling unit”) of Landsat and S-2 data, respectively. Source EO data will be sub-sampled using a 100 km by 100 km window located in the centre of the unit; this choice represent an improvements over previous Fire cci that was decided for analysing fire size and patches.

For the three integration sites in Africa, reference fire perimeters will be generated from time multiple consecutive S-2 images and from pairs of HR/VHR over spot areas, if data is available. For all source images reference fire perimeters will be extracted following the same classification approach.

A summary of the main criteria for validation Fire_cci BA products is provided in Table 2.

Table 2. Summary of the validation protocol (n_y is the number of sampled units per year).

BA products				Validation			
Name	**Sensor**	**Scale**	**Years**	**Stratification**	**Spatial units**	**n_y**	**EO data**
FIRECCI51	Terra-MODIS	Global	2017-2019	Olson + BA	TSA	100	Landsat
FIRECCI60	MODIS + S-3 + AVHRR	Global	2017-2019	Olson + BA	TSA	100	Landsat
FIRECCILT20	NOAA-AVHRR-LTDR	Global	2017-2019	Olson + BA	TSA	100	Landsat
FIRECCIS311	S-3 OLCI+SLSTR	Global	2019	Olson + BA	TSA	100	Landsat
FIRECCISFD20	S-2 MSI	Africa	2019	Olson + BA	TSA	50	S-2
FIRECCIS1S2AF10	S-1 SAR + S-2 MSI	Selected sites	2019	*N/A*	*N/A*	*N/A*	HR/VHR

5 Description of validation activities

Validation activities are the processing steps necessary for the implementation of the protocol outlined above. For pixel-based Fire cci BA products (namely FIRECCI51, FireCCI60, FIRECCIS311, and FIRECCISFD20 (see Table 2)), major activities will be:

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

For what concerns the FIRECCILT20 product derived from the AVHRR LTRD, which is provided aggregated at 0.25 deg resolution, regressive analysis will be carried out to compare with reference fire perimeters.

5.1 Sampling scheme implementation and reference EO data

Reference fire perimeters for validation of **global Fire_cci BA products** will be derived from Landsat data. The sampling scheme detailed in 4.2.1 will be applied to select, for each year, a total of 100 TSAs for building the "long sampling units" of reference data. Each sampling unit will be 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 will be considered to extend the time lag or to shorten the length of the "long sampling units".

For the **Africa regional BA product** (FireCCISFD20), fire reference perimeters will be created for 50 TSAs by using Sentinel-2 data and the "long" sampling units approach. Although BA fire perimeters will be derived from the same dataset used for deriving Fire_cci BA product, the validation protocol will assure the extraction of independent reference perimeters (different algorithms) with the highest accuracy possible (operator supervision), hence ensuring to be as closest as possible to the ground truth.

For **BA products obtained over the three study sites in Africa** and focusing on testing the **integration of Sentinel-1 and Sentinel-2** data, no stratified sampling will be implemented. Since sites were selected only for algorithm testing purposes, validation will be carried out at stage 1. As for validation of global and regional products, reference EO dataset will be composed of consecutive image pairs for pre- and post-fire dates. An effort to use independent data will be conducted. The possibility of accessing archives of HR/VHR data, made available free of charge to the scientific community, will be investigated.

In all cases, Level 2 EO image products (i.e. Bottom of Atmosphere-BOA reflectance) will be preferred.

5.2 Processing steps for creating reference BA data

A supervised semi-automatic classification algorithm will be applied for all the above listed product categories to consecutive pair(s) of reference images to identify areas burned between pre- and post-fire dates. Training data will be extracted for image pair(s) by photointerpretation of RGB false colour composites.

The semi-automatic classification will be applied at the pixel level (20-30 m spatial resolution, depending on the source EO data). Output raster will be converted to ESRI shapefile © and cleaning of small polygons will be carried out to reduce the impact of isolated pixels, geo-location errors and different spatial resolutions.

ESRI shapefile © of reference fire perimeters will be uploaded in a GIS environment to fulfil the task of final visual refinement by photointerpretation and comparison with false colour composites of image EO data bands. Shapefile of consecutive image pairs will be summed up over the time period covered by the "long sampling unit" and it will be input for cross-tabulation with Fire_cci BA product.

ESRI shapefile © of reference fire perimeters will be accompanied by metadata including information about geographic projection, year of detection, information on the original input images (first and last dates/images of the sampling unit), name of interpreter, institution, etc. according Fire cci standards.

5.3 Processing steps for Fire_cci BA product extraction

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

- Projection conversion: Fire_cci BA products will be re-projected to UTM, WGS84, with the UTM zone coincident with the one covered by the major part of the reference L8/S-2 scene assigned to each TSA;
- Clipping to the specific sub-sampled window size (100 km by 100 km) centred over the sampling unit(s);
- Temporal sampling of Fire_cci BA between pre- and post-fire dates of the reference data (first and last date of the "long sampling unit";

5.4 Cross-tabulation and accuracy metrics

Reference and Fire_cci BA products will be overlaid and intersected to extract the confusion matrix; each cell of the confusion matrix e_{ij} (Table 3) expresses the proportion of area of agreement or disagreement for all pixels of the TSA between the BA product (map) and the reference classes (Padilla et al. 2014a; Padilla et al., 2017).

Confusion matrices will be summed up over all of the sampling units and, if applicable (i.e. multi-years BA products) for each year, to compute the accuracy metrics listed in Table 3.

5.5 Regressive analysis of gridded products

For what concerns the FIRECCILT20 product, which will be derived from AVHRR LTRD data, regressive analysis will be carried out to compare the BA product with reference fire perimeters. Reference fire perimeters will be aggregated over grid cells of 0.25 deg. The fraction of burned area in each cell will be compared to assess spatial agreement between the two datasets/products. Correlation metrics will be computed (R2, slope and intercept) and analysed.

5.6 Error analysis

Part of the validation activities will be 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 2017-2019 period will be presented and discussed. 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.

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.

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 Sensing of Environment, 217, 72-85.

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.

Morisette, J.T., Baret, F., & Liang, S. (2006). Special issue on global land product validation. IEEE Trans. Geosci. Remote Sens. 44, 1695–1697.

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.

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.

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.

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

ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BA	Burned Area
CBERS	China-Brazil Earth Resources Satellite
Ce	Commission error ration
CEOS	Committee on Earth Observation Satellites
CSDGM	Content Standard for Digital Geospatial Metadata
DC	Dice Coefficient
ECV	Essential Climate Variables
EO	Earth Observation
ESA	European Space Agency
ESRI	Environmental Systems Research Institute
ETM(+)	Enhanced Thematic Mapper (Plus)
FireCCI51	MODIS Fire_cci v5.1
FireCCISFD11	Small Fire Database Fire_cci v1.1
HR	High Resolution
ISO	International Organization for Standardization
L3JRC	Global Multi-year (2000-2007) Validated Burnt Area Product
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
NIR	Near InfraRed
Oe	Omission error ration
OLI	Operational Land Imager
PVR	Product Validation Report
relB	Relative bias
S-1	Sentinel-1
S-2	Sentinel-2
SLC	Scan Line Corrector
SFD	Small Fire Database
SPOT	Satellite Pour l'Observation de la Terre
SWIR	Short Wave InfraRed
TM	Thematic Mapper
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 3) which expresses the proportion of agreements or disagreements between product and reference classifications. The error matrix is derived for each sampling unit (TSA); matrix entries e_{ij} express the agreements (i=j) or disagreements (i≠j) for all pixels of the TSA between the BA product (map) class and the reference data summed up over the total number of sampling units (TSAs).

Table 3. 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 4.

Table 4. 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 will be delivered as ESRI shpeafiles © (.shp), along with the auxiliary files required (.dbf, .prj, shx, .sbn, .xml). The projection will be 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 6):

- 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 5.
- 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 5. Satellite-sensor codes naming convention

Satellite-sensor	Mission Code (MMM)	Reference system	
Landsat-8 OLI	LC8	Path (ppp)	Row (rrr)
Landsat-7 ETM+	LE7	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 6. 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 will contain 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.