ESA Climate Change Initiative – Fire_cci
O3.D2. Burned area database for candidate validation tiles.

Project Name	ECV Fire Disturbance: Fire_cci Phase 2
Contract Nº	4000115006/15/I-NB
Issue Date	30/01/2018
Version	1.1
Author	Mihai A. Tanase, Ángel Fernández Carillo
Document Ref.	Fire_cci_O3.D2_CAL-VAL_tiles_v1.1.docx
Document type	Internal

To be cited as: M.A. Tanase, A. Fernández Castillo (2018) ESA CCI ECV Fire Disturbance: O3.D2.Burned area database for candidate validation tiles, version 1.1. Available at: http://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) EHU – University of the Basque Country (Spain) UL – University of Leicester (United Kingdom) UCL – University College London (United Kingdom) ISA – School of Agriculture, University of Lisbon (Portugal)
System Engineering	BC – Brockmann Consult (Germany)
Climate Research Group	MPIC – Max Planck Institute for Chemistry (Germany) IRD - Research Institute for Development (France) LSCE - Climate and Environmental Sciences Laboratory (France) VUA - Vrije Universiteit Amsterdam (Netherlands)

Distribution

Affiliation	Name	Address	Copies
ESA	Stephen Plummer (ESA)	stephen.plummer@esa.int	electronic copy
Project Team	Emilio Chuvieco (UAH)	emilio.chuvieco@uah.es	electronic copy
	M. Lucrecia Pettinari (UAH)	mlucrecia.pettinari@uah.es	
	Joshua Lizundia (UAH)	joshua.lizundia@uah.es	
	Gonzalo Otón (UAH)	gonzalo.oton@uah.es	
	Mihai Tanase (UAH)	mihai.tanase@uah.es	
	Miguel Ángel Belenguer (UAH)	miguel.belenguer@uah.es	
	Aitor Bastarrika (EHU)	aitor.bastarrika@ehu.es	
	Ekhi Roteta (EHU)	ekhi.roteta@gmail.com	
	Kevin Tansey (UL)	kjt7@leicester.ac.uk	
	Marc Padilla Parellada (UL)	mp489@leicester.ac.uk	
	James Wheeler (UL)	jemw3@leicester.ac.uk	
	Philip Lewis (UCL)	ucfalew@ucl.ac.uk	
	José Gómez Dans (UCL)	j.gomez-dans@ucl.ac.uk	
	James Brennan (UCL)	james.brennan.11@ucl.ac.uk	
	Jose Miguel Pereira (ISA)	jmocpereira@gmail.com	
	Manuel Campagnolo (ISA)	mlc@isa.ulisboa.pt	
	Thomas Storm (BC)	thomas.storm@brockmann-consult.de	
	Johannes Kaiser (MPIC)	j.kaiser@mpic.de	
	Angelika Heil (MPIC)	a.heil@mpic.de	
	Florent Mouillot (IRD)	florent.mouillot@cefe.cnrs.fr	
	M. Vanesa Moreno (IRD)	mariavanesa.morenodominguez@cefe....	
	Philippe Ciais (LSCE)	philippe.ciais@lsce.ipsl.fr	
	Chao Yue (LSCE)	chaoyuejoy@gmail.com	
	Pierre Laurent (LSCE)	pierre.laurent@lsce.ipsl.fr	
	Guido van der Werf (VUA)	guido.vander.werf@vu.nl	
	Ioannis Bistinas (VUA)	i.bistinas@vu.nl	

Summary

This document describes the approaches used to generate the validation dataset used to estimate the quality of the Burned Area (BA) product estimate from Sentinel-1 synthetic aperture radar (SAR) data for the large demonstrator area (LDA) located in tropical South America. Amazon LDA is part of the Small Fire Database (SFD).

	Affiliation/Function	Name	Date
Prepared	UAH	Mihai A. Tanase	19/01/2018
Reviewed	UAH – Project Manager UAH - Science Leader	Lucrecia Pettinari UAH - Science Leader	30/01/2018
Authorized	UAH - Science Leader	Emilio Chuvieco	30/01/2018
Accepted	ESA - Technical Officer	Stephen Plummer	30/01/2018

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

Document Status Sheet

Issue	Date	Details
1.0	30/09/2017	First Issue of the document
1.1	30/01/2018	Addressing comments of CCI-FIRE-EOPS-MM-17-0092

Document Change Record

Issue	Date	Request	Location	Details
1.1	30/01/2018	UAH-ESA ESA	Section 1 Section 2.3 Section 3 Section 4 Section 4.3 Section 5	Small changes in the text. Clarified the independence between calibration and validation activities. New section included to replace the Section 4 of the previous version of the document. Section expanded, and sub-sections 4.1 and 4.2 added. Text updated (previously was section 3.1). New section added.

Table of Contents

1 Executive Summary 5
2 Introduction 6
2.1 Purpose of the document 6
2.2 Applicable Documents 6
2.3 Background 6
3 Sampling design..7
3.1 Sampling size 7
3.2 Stratification..7
3.3 Sample allocation...8
4 BA validation dataset..9
4.1 General considerations...10
4.2 Subsampling...10
4.3 Generation of reference fire perimeters 11
4.4 Data structure and naming convention 12
4.5 Metadata..13
5 Accuracy estimates..13
6 References 14
Annex 1: Acronyms and abbreviations 16

List of Tables

Table 1: Landsat TSAs selected for validation within Option 3 9
Table 2: Sampled error matrix on a sampling unit*. 13

List of Figures

Figure 1: Sampling units (upper left panel), eco-regions (upper right panel, Olson et al. 2001), stratification by biome (lower left panel), and stratification by biome and fire activity for the Amazon LDA. 8
Figure 2: Sampling units selected for validation (by biome and fire activity). 9
Figure 3: Cloud fraction over the Amazon LDA (grey rectangle) as seen by Terra/MODIS. The maps show how much of Earth`s surface was covered by clouds for each month in 2016 (the latest year of complete temporal coverage). The patterns for 2017 are assumed to be similar. 11

1 Executive Summary

Burned area (BA), as derived from satellites, is considered the primary variable that requires climate-standard continuity. It can be combined with information on burn efficiency and available fuel load to estimate emissions of trace gases and aerosols. Measurements of BA may be used as direct input (driver) to climate and carbon cycle models or, when long time series of data are available, to parameterize climate-driven models for BA. The aim of Option 3 is to provide Sentinel-1 burned area products over a large demonstrator area (LDA) located in tropical South America. As such, Option 3 extends the areas mapped within the “Small fire database” to tropical regions in South America.

Climate and carbon cycle models need information on the validity of the input BA products. Such information is described within Product Validation Reports. PVRs usually describe the i) approaches used to derive the reference dataset (i.e., the validation dataset), ii) methods used to assess the quality of BA products (e.g., omission errors, commission errors), and iii) results of the validation over specific BA products. **Within Option 3 two complementary documents are used describe the methods used to appraise BA product quality**:

- D2 Burned area database for the candidate validation tiles: focused on the methods used to generate the validation database (this document).
- D4 Product validation report (PVR): describing the results of the validation process.

Since reference fire perimeters are not available via the baseline project, Option 3 has generated its own validation dataset to appraise the accuracy of BA products coming from the Sentinel-1 algorithm in tropical South America (Amazon LDA). The validation dataset **follows the principles outlined in the main Fire_cci proposal** [RD-3] with changes accommodating the specific conditions encountered for Option 3: i) **reduced temporal frequency** of the generated BA product (one year, 2017[1]), ii) **reduced availability of cloud free optical imagery** (used to derive the reference dataset), and, iii) **reduced number of biomes** (i.e., tropical forest and tropical and subtropical savanna).

Burned areas classified based on Sentinel-1 imagery [RD-2] were validated using optical imagery as large-scale ground assessment of BA products is cost prohibitive. The reference dataset (i.e. validation database) was developed using medium resolution optical images, acquired close enough in time as to portray the same ground conditions as the input Sentinel-1 images from which the BA product is generated. The main optical sensors used to derive the reference BA are Landsat-7/8 as for the validation strategy followed in the baseline project. Sentinel-2 images were used where insufficient cloud-free image pairs were available. For a sample of 46 validation sites, a BA reference dataset was generated for year 2017 at selected time periods (i.e. cloud-cover permitted). The validation dataset was generated following the methods and procedures described in D4.1 Product Validation Report [RD-3] and Padilla et al.

[1] Due to significant Sentinel-1 data gaps in tropical South America during 2015-2016 the BA algorithm cannot be deployed as it needs at least one acquisition every month. Since Sentinel-1 data availability increased significantly starting with September 2016, it was decided, within the 7th Contract Progress Meeting of the Fire_cci project, to generate the BA product for the year 2017.

(2017). BA product validation over Amazon LDA was carried out within Option 3. The reference BA shall be compared with BA algorithm outputs (when available), with common temporal interval and spatial coverage. CEOS LPV (land product validation) protocols are used to generate the reference data and peer-reviewed methods were used to summarize and express the validation results (Padilla et al. 2017). The reference BA dataset is specifically designed for the SFD Amazon product to ensure temporal overlaps with Sentinel-1 derived BA estimates.

2 Introduction

2.1 Purpose of the document

The objective of this document is to describe the generation of the validation dataset used to assess the quality of the Sentinel-1 derived BA according to Option 3 product outputs. This document will be supplemented by D4. Product validation report (PVR). The document follows and draws from the D4.1 Product Validation Report provided within Fire_cci Phase 2 ([RD-3]).

2.2 Applicable Documents

[RD-1]	Option 3, Radar Burn Ratio for burnt area detection and mapping, Proposal prepared for ESA on September 12, 2016. Option to ECV Fire Disturbance Phase-2 project.
[RD-2]	Tanase M., Belenguer-Plomer M.A (2017) O3.D1 Algorithm Theoretical Basis Document (ATBD) – Small Fires Dataset (SFD) for the large demonstrator area (LDA) in South America, version 1.0.
[RD-3]	Padilla M., Wheeler J., & Tansey K. (2017). ESA CCI ECV Fire Disturbance: D4.1.1. Product Validation Report, version 1.2. Available at: http://www.esa-fire-cci.org/documents

2.3 Background

The ESA CCI initiative stresses the importance of providing a higher scientific visibility to data acquired by ESA sensors, especially in the context of the IPCC reports. This implies producing consistent time series of accurate Essential Climate Variables (ECV) products, which can be used by the climate, atmospheric and ecosystem scientists. The quality of such products needs to be characterized quantitatively, by independent means, to facilitate critical information and product reliability to the end users. Within the Fire_cci Phase 2, the BA products are validated using reference data collected by means of probabilistic sampling carried out both in space and time. In addition, the sampling allocation follows stratification criteria to properly allocate samples to each stratum and optimize resources dedicated to reference data generation (Padilla et al. 2017). Such complex validation designs recognize the shortcomings of earlier methods based on using a relatively reduced number of locations to validate global BA products (Chuvieco et al. 2008; Roy and Boschetti 2009; Roy et al. 2008; Tansey et al. 2008).

As part of an effort to promote the acceptance of the CCI products by external communities, an independent validation analysis, including the assessment of temporal trends of accuracy was adopted in the Phase 2 of the Fire_cci project. Validation samples' 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, neither for calibration nor for “tuning” processes. Independence is further ensured by separation of tasks (algorithm development and product validation) between different team members. Since the BA product over Amazon LDA covers one-year (2017) validation of temporal trends is not needed, thus simplifying the adopted approach. Option 3 follows a similar stratification (biome level) and sample allocation concept as the baseline project.

Accuracy is characterized through cross-tabulation of reference and classified datasets, by accounting for the spatio-temporal coincidences and disagreements. This approach is widely used in BA mapping projects (Boschetti et al. 2004; Boschetti et al. 2016; Chuvieco et al. 2008; Giglio et al. 2009; Padilla et al. 2017; Padilla et al. 2014a; Padilla et al. 2014b; Padilla et al. 2015; Roy and Boschetti 2009). The reference maps are derived from medium spatial resolution images using standard methods as described by the Committee on Earth Observation Satellites’ Land Product Validation Subgroup (CEOS-LPV) (Boschetti et al. 2016; Padilla et al. 2017).

3 Sampling design

As per the baseline project validation strategy, a stratified random sampling design was used to provide reference BA estimates and infer the accuracy of the BA product produced within Option 3. The sampling units were the Thiessen Scene Areas (TSAs) as defined based on the Landsat World Reference System II (WRS-II). TSAs contain the area closer to the centre of the corresponding Landsat frame and do not overlap (Figure 1). TSAs provide partitioning for selecting non-overlapping sampling units that allows for convenient computing of unbiased estimators (Padilla et al. 2014b).

3.1 Sampling size

The total number of samples were computed using Eq. 1 (Olofsson et al. 2014) under the following assumptions: i) proportion of disturbed area 20% (i.e., TSA with high fire activity), ii) expected user accuracy for burned areas, 60% (Padilla et al. 2014a; Padilla et al. 2014b), iii) expected user accuracy for not affected areas, 90%, and iv) expected standard error for the overall accuracy, 5%.

$$n = \frac{(\sum W_i S_i)^2}{[S(\hat{O})]^2 + (1/N)\sum W_i S_i^2} \approx \left(\frac{\sum W_i S_i}{S(\hat{O})}\right)^2 \qquad (1)$$

where: S(O) - expected standard error for the overall accuracy; W_i – class proportion, S_i - the standard deviation of stratum i, $Si = \sqrt{Ui(1-Ui)}$ with U_i being the expected user accuracy, N – sampling population. When N is large, the second term in the denominator can be ignored.

3.2 Stratification

The stratification of sampling units was designed to ensure sufficient sampling over major Olson ecoregions (Olson et al. 2001) with a focus on regions with high fire activity. The stratification consisted in assigning each sampling unit to the appropriate terrestrial ecoregion among those found within the Amazon LDA: i) Tropical and Subtropical Moist Broadleaf Forests and, ii) Tropical and Subtropical Grasslands, Savannas and Shrublands (Figure 1).

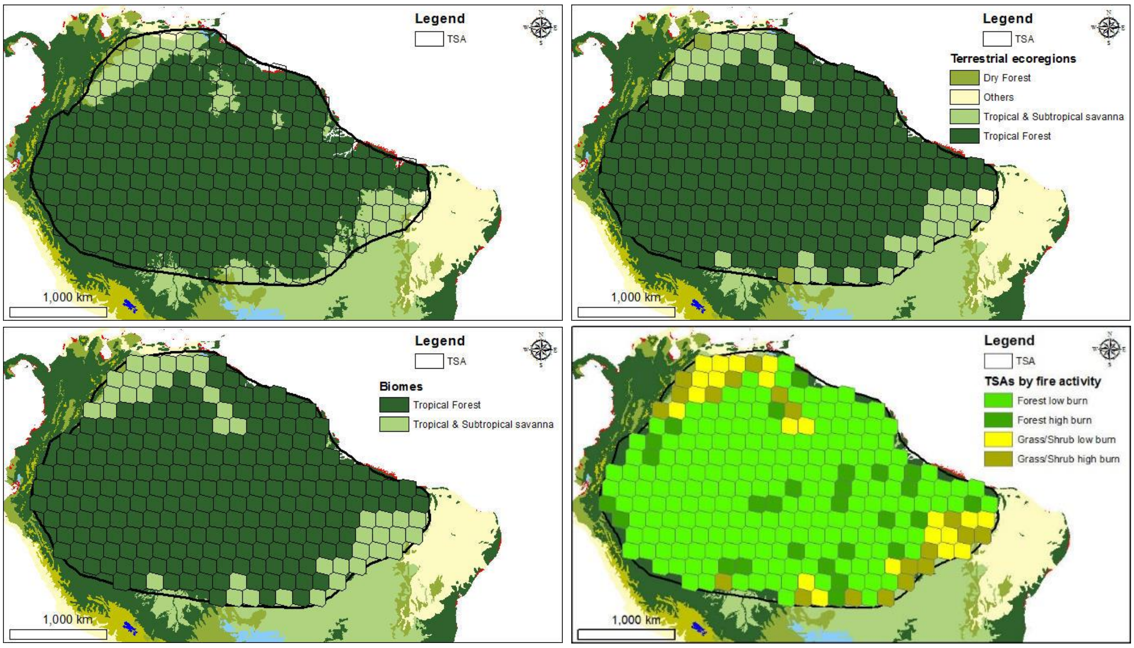

Figure 1: Sampling units (upper left panel), eco-regions (upper right panel, Olson et al. 2001), stratification by biome (lower left panel), and stratification by biome and fire activity for the Amazon LDA.

The sampling units (TSAs) were stratified by eco-region (Olson et al. 2001) and fire activity into four classes: Forest high burn (44 TSAs), Forest low burn (174 TSAs), Grasslands/Shrublands high burn (8 TSAs), and Grasslands/Shrublands low burn (32 TSAs). Notice that, the forest class includes both Tropical forest and Dry forest eco-regions as only two TSAs (out of 258 covering the Amazon LDA) were classified as dry forests. Similarly, the Grasslands/Shrublands class includes one TSA classified within the Desert and Xeric Shrublands eco-regions. The stratification by fire activity was based on the BA product MCD64 (Giglio et al. 2009) for the year 2016-2017. TSAs above 80th percentile were classified as high burn activity within each eco-region class.

3.3 Sample allocation

Samples were allocated by strata proportionally with the square root of the burned area in each class (Eq. 2), as suggested in Padilla et al. (2017). The 46 validation TSAs were split as follows: Forest high burn (14), Forest low burn (13), Grasslands/Shrublands high burn (6), and Grasslands/Shrublands low burn (13). The TSAs were drawn randomly from the entire population after eliminating the three TSAs used during the BA algorithm development. Their spatial distribution is showed in Figure 2 while their correspondence with the Landsat WRS-II reference system is shown in Table 1.

$$n_h \propto N_h \sqrt{\overline{BA}_h} \quad (2)$$

where: Nh – number of sampling units within class h, BA - mean burned area within class h.

Table 1: Landsat TSAs selected for validation within Option 3

Landsat WRS II reference system							
Forest high burn		Forest low burn		Grass/Shrub high burn		Grass/Shrub low burn	
006/058	226/61	001/067	228/63	004/056		001/068	220/065
007/059	227/066	007/060	231/62	221/064		002/054	222/067
220/063	227/068	008/064	232/55	222/066		003/055	226/069
223/064	227/069	009/063	232/63	223/067		006/055	229/069
224/061	228/067	224/62	233/57	223/068		006/056	232/057
224/066	230/069	228/61	233/63	224/067		006/057	233/056
225/064	233/067	228/62				219/065	

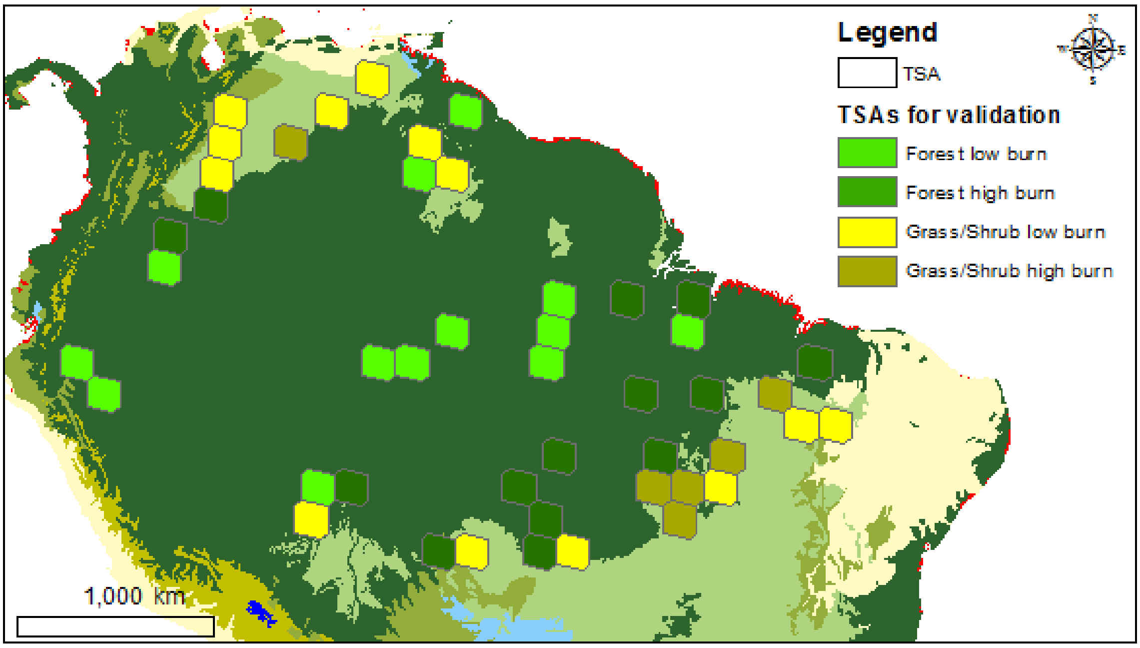

Figure 2: Sampling units selected for validation (by biome and fire activity).

4 BA validation dataset

In this document, burned area is defined as any vegetated area that has been completely or partially consumed by a fire, regardless of whether the fire was the result of human activities or natural phenomena, or whether the fire affected wildland areas or managed lands (agricultural, forests, or pastures).

This section describes the protocol used to generate reference information for BA validation as agreed by the partners involved in the Fire_cci consortium and as described in D4.1 Product Validation report of the Fire_cci main project [RD-3]. The methods, based on the CEOS-CalVal protocol for the validation of burned area products (Padilla et al. 2017), were adapted to suit Option 3 (temporal trend analysis was omitted as BA product is generated for only one year). Reference perimeters are generated from multi-temporal comparison of medium resolution satellite imagery, acquired before and after fire using a semi-automatic procedure. After the semi-automatic mapping, quality control is performed through visual inspection. All fire perimeters are reviewed and perimeters with errors are rectified. This procedure is iterated until no errors are identified.

4.1 General considerations

Following CEOS Validation protocol for BA products (Boschetti et al. 2016), reference data is generated from two consecutive images acquired over the same region. A sampling unit is delimited spatially by the WRS-2 / MGRS footprints and temporally by the acquisition dates of consecutive images (image pair). It is relevant to limit the time length between any consecutive observations to make sure the spectral signal of a fire occurred between acquisition times is still present in the latest image. Therefore, image pairs form a sampling unit whenever they were separated by 16 days or less (same as for the baseline project). Over areas with frequent cloud cover, the interval may be increased to 32 days to allow formation of additional sampling pairs as cloud cover is persistent over large parts of the Amazon LDA (Figure 3). The use of a 32 days interval of 32 days is still feasible as fast-growing vegetation (i.e., grasslands) need longer time to recuperate after fires. In addition, the Sentinel-1 data over the Amazon LDA is available at about one-month interval. Optical imagery with less than 40% of clouds (a slightly higher limit when compared to the 30% used for the main project) and the temporal requirements between image pairs specified above may limit the availability of useable images and thus the reference data generation. Although reference BA was generated for only one year the stratification by calendar year was considered as a convenient approach when planning data collection. It also makes easy to expand the temporal period of study by adding complete years if necessary. All sampling units were allocated to the same calendar year (2017). For consistency and simplicity, this assignation was based on the earliest acquisition date of the Landsat (or Sentinel-2) image pair. Over TSAs with few cloud-free Landsat image pairs, the use of Sentinel-2 datasets is envisaged for gap filling the temporal series of reference burn perimeters during the fire season.

4.2 Subsampling

The main advantage of subsampling is that it allows increasing the number of sampling units which helps to decrease the variance of accuracy estimates (Padilla et al. 2017). Each sampling unit selected was subsampled using a 30 km by 20 km window. The rectangular size makes possible the evaluation of single pixels (depending on differences in reflectance between a pixel and its neighbouring area) while the whole image is visualized (1:80000 scale) on a 27’’ screen. This reduces the necessity to navigate across the scene in the process of image exploration for the collection of training data and/or revision of image classification. The navigation across a scene is a time-consuming task with little or no contribution to reference data generation, thus it is to be avoided as much as possible. The subsampling is expected to produce a gain in the precision of BA estimates mainly due to the increase of *n* and within unit positive correlation (Padilla et al. 2017). The positive correlation implies that pixels within a unit provide similar information, and therefore a sample of them may provide a similar average when compared to the one obtained from all pixels on the unit.

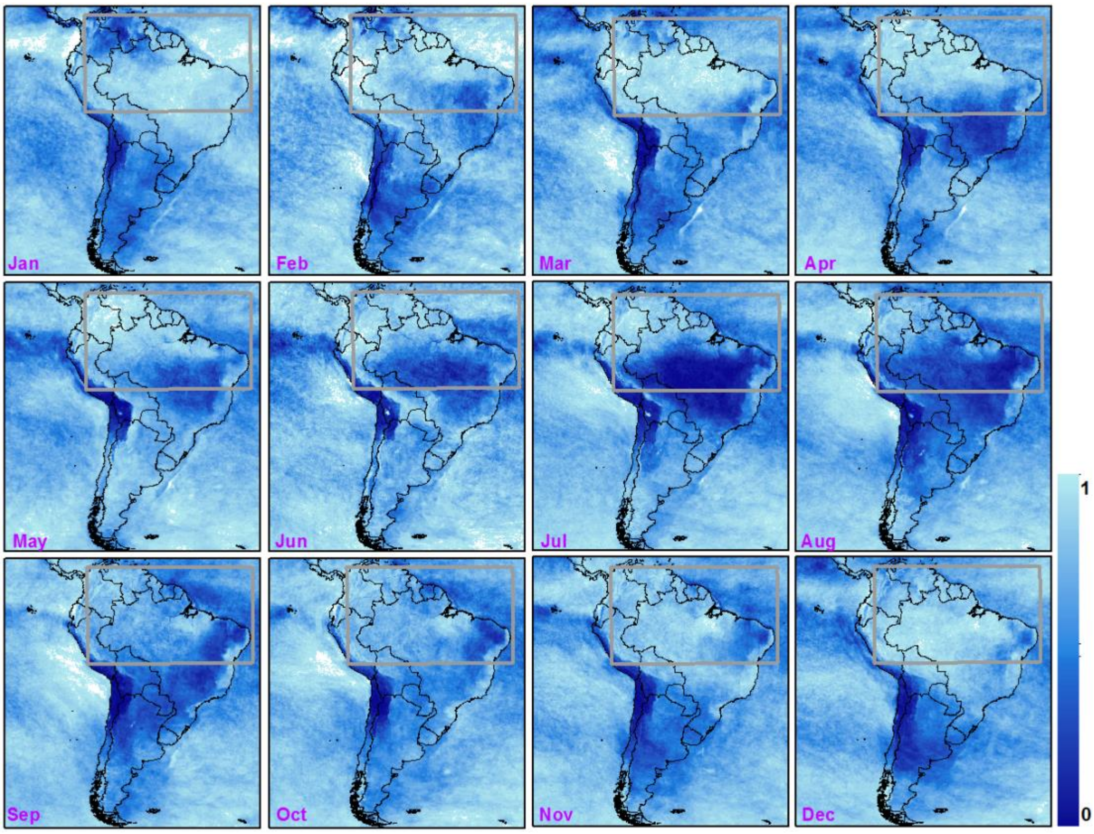

Figure 3: Cloud fraction over the Amazon LDA (grey rectangle) as seen by Terra/MODIS. The maps show how much of Earth`s surface was covered by clouds for each month in 2016 (the latest year of complete temporal coverage). The patterns for 2017 are assumed to be similar.

4.3 Generation of reference fire perimeters

The semi-automatic procedure to generate the reference dataset consists of two steps. On the first step, the image pair (pre- and post-) reflectance is reformatted for efficiency. The reformatting consists on co-registering the core region (30 km by 20 km) of the scene. The output is a raster file with six bands, with the SWIR, NIR and RED bands of the two images. This first step is automatic and can be parallelized. The second step is the semi-automatic BA classification. For the classification, the interpreter needs to digitize training polygons for burned and unburned areas, and optionally for clouds. The training data is used to fit a Random Forest Classifier (Breiman 2001; Pedregosa et al. 2011) on each sampling unit (pair of images), taking as input variables NBR, SWIR and NIR of the pre- and post- dates, and the multitemporal index dNBR. Those spectral regions and indices are specifically useful to discriminate burned areas (Giglio et al. 2009; Goodwin and Collet 2014). The classification procedure consists in repetitive iterations of visual inspection, delineation of training polygons and classification until no further errors can be perceived on the visual inspection. Optionally, the classification can be overwritten by polygons digitized manually. The output is an ESRI® shape file with the reference data and metadata as defined below. Within post-processing, the detected perimeters are filtered by area with polygons below 0.1 ha (pixel size polygons) being removed. Shape indices are then computed (Eq.3 and Eq.4) to reduce misclassification errors over cropping areas. A filter based on shape indices and size is applied to all polygons below 2 ha.

$$CR_i = \sqrt{\frac{A_i}{Ac_i}} \quad (3) \qquad RBF_i = 1 - \frac{A_i}{Bf_i} \quad (4)$$

where: CRi- compactness ration for polygon i, RBFi – related bounding figure, Ai- area, Acj – area for a circle with the same perimeter as the polygon, Bfi – convex hull for polygon i.

Parts of the scene that cannot be observed or interpreted, either due to clouds or sensor related difficulties (i.e. SLC-off problems of ETM+) in one of the two images pre- or post- are classified as no-data. As such, only areas with reliable information are included in the validation process.

4.4 Data structure and naming convention

Data structure and naming convention is the same as for the PVR of the Fire_CCI Phase 2 project. Each burned area reference file is an ArcGIS™ shape file (.shp), along with the auxiliary files required (*.dbf, *.prj, *.shx, *.sbn, *.xml). The projection is UTM, WGS84, with the UTM zone/row being the zone that is covered by the major part of the scene. The following attribute fields are included in the shape file:

- 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: satellite-code_Path_Row (e.g. LT5_201_032). Satellite codes as follows:
  - Landsat-7 ETM+ (LT5)
  - Landsat-8 OLI (LC8)
  - Sentinel-2 (S2)
- 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.

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

PRO_RD_YYYYMMDD_YYYYMMDD_PPPRRR

where:

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

RD = stands for Reference Data

yyyymmdd (year, month, date). The first one is the pre-fire date, which is the date of the first image used for BA detection; the second one is the post-fire date, which is the date of the last image used for generating the reference fire perimeters.

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.

4.5 Metadata

The metadata of the reference files is written as an XML document. The metadata contains the author of the reference data file, its institution, the date of creation, the input data sources (names of satellite image files) and the reference of the website of the Fire_cci project. An example of a metadata file can be found in D4.1 Product Validation report [RD-3].

5 Accuracy estimates

Accuracy estimates are based on the cross-tabulation approach (Latifovic and Olthof 2004). The result of the cross tabulation can be represented by the error matrix (Table 2) which express the amount of agreement and disagreement between product and reference classifications. A product pixel is coded as “burned” if it was detected as such between the dates defining the temporal dimension of the sampling unit, similarly as is done on the reference classification. All other sampled pixels are coded as “unburned” or “no-data”, the latter for unobserved pixels. The agreement and disagreement areas can be measured in each sampling unit by spatially comparing reference and product binary (burned or unburned) maps. This comparison is performed by overlaying the two vector polygons layers derived from the product and reference datasets. The product binary raster map is converted to polygons and then re-projected to the spatial reference system of the reference dataset.

Table 2: Sampled error matrix on a sampling unit*.

Product	Reference		Total
	Burned	Unburned	
Burned	e_{11}	e_{12}	e_{1+}
Unburned	e_{21}	e_{22}	e_{2+}
Total	e_{+1}	e_{+2}	

* eij express the agreements (diagonal cells) or disagreements (off diagonal cells) in terms of area (m^2) between the BA product (map) class and the reference class

Since the reference data set may include areas burned as several time periods (i.e. the reference data is generated from a consecutive series of image pairs) the product validation can be carried out at two scales: i) the entire interval with the binary maps defined by the first and last acquisition dates, and ii) by intervals, with the binary maps defined by the acquisition dates of image pair included in the data series. The difference in accuracy estimated from the two scales gives an indication of the effect of the size in time of the sampling unit onto the accuracy estimations.

Apart of the common ratios between combinations of error matrix cells (i.e. commission (Eq.5) and omission errors (Eq. 6)) the use of the Dice Coefficient (DC, Eq.7) may be useful when comparing product accuracies since it summarizes both commission and omission errors showing the accuracy of category 'burned' (Padilla et al. 2017; Padilla et al. 2015). Lastly, the error matrix can be used to compute the bias either as a total estimate (Eq. 8) or in relative terms to the reference BA (Eq. 9).

$$Ce = \frac{e_{12}}{e_{1+}} \text{ (5)}; \qquad Oe = \frac{e_{21}}{e_{+1}} \text{ (6)}; \qquad \text{DC} = \frac{2e_{11}}{2e2_{11} + e_{12} + e2_{21}} \text{ (7)}$$

$$bias = e_{12} - e_{21} \text{ (8)}; \qquad \text{relB} = \frac{e_{12} - e_{21}}{e_{+1}} \text{(9)}.$$

Since the Amazon LDA product covers a much smaller area (i.e. homogenous) when compared to global products, the size of the reference BA relative to the product coverage is relatively high, the validation temporal dimension spans only one year, and only two strata (biomes) are present, the global estimate of accuracy is computed without considering a stratified sampling design as for the global product.

6 References

Boschetti, L., Flasse, S.p.P., & Brivio, P.A. (2004). Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: The Pareto Boundary. *Remote Sensing of Environment*, *91*, 280-292

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., Del Valle, H., Anaya, J., Di Bella, C., Cruz, I., Manzo, L., López, G., Mari, N., González-Alonso, F., Morelli, F., Setzer, A., Csiszar, I., Kanpandegi, J.A., Bastarrika, A., & Libonati, R. (2008). Global Burned Land Estimation in Latin America using MODIS Composite Data. *Ecological Applications*, *18*, 64-79

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

Goodwin, N.R., & Collet, L.J. (2014). Development of an automated method for mapping fire history captured in Landsat TM and ETM+ time series across Queensland, Australia. *Remote Sensing of Environment*, *148*, 206-221

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

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., E.Woodcock, C., & Wulder, M.A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, *148*, 42-57

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. *Biogeosceinces*, *51*, 933-938

Padilla, M., Olofsson, P., Stephen V., S., Tansey, K., & Chuvieco, E. (2017). Stratification and sample allocation for reference burned area data. *Remote Sensing of Environment*, *203*, 240-255.

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., 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

Pedregosa, F.V.G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning*, *12*, 2825-2830

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

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

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*

Annex 1: Acronyms and abbreviations

ATBD	Algorithm Theoretical Basis Document
BA	Burned Area
CCI	Climate Change Initiative
CEOS	Committee on Earth Observation Satellites
dNBR	Difference Normalized Burn Ratio
ECV	Essential Climate Variables
ESA	European Space Agency
ETM+	Enhanced Thematic Mapper Plus sensor on Landsat satellites
IPCC	Intergovernmental Panel on Climate Change
LDA	Large Demonstrator Area
LPV	Land product validation
MGRS	Military Grid Reference System
MODIS	Moderate Resolution Imaging Spectroradiometer
NBR	Normalized Burned Index
NIR	Near Infrared
PVR	Product Validation Report
SAR	Synthetic Aperture Radar
SFD	Small Fire Database
SLC	Single Look Complex
SWIR	Short Wave Infrared
TSA	Thiessen Scene Area
UTM	Universal Transverse Mercator
WGS84	World Geodetic System 1984
WRS	Worldwide Reference System
XML	Extensible Markup Language