




ESA Climate Change Initiative – Fire_cci

D2.1.2 Algorithm Theoretical Basis Document (ATBD) – Small Fires Dataset (SFD)

Project Name	ECV Fire Disturbance: Fire_cci Phase 2
Contract Nº	4000115006/15/I-NB
Issue Date	01/10/2018
Version	1.0
Author	Aitor Bastarrika, Ekhi Roteta
Document Ref.	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0
Document type	Public

To be cited as: Bastarrika A., Roteta E. (2018) ESA CCI ECV Fire Disturbance: D2.1.2 Algorithm Theoretical Basis Document-SFD, version 1.0. Available at: <http://www.esa-fire-cci.org/documents>

	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
	Algorithm Theoretical Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2	Page	2		


Project Partners

Prime Contractor/ Scientific Lead & Project Management	UAH – University of Alcalá (Spain)
Earth Observation Team	UAH – University of Alcalá (Spain)
	EHU – University of the Basque Country (Spain)
	UL – University of Leicester (United Kingdom)
	UCL – University College London (United Kingdom)
System Engineering	ISA – School of Agriculture, University of Lisbon (Portugal)
	BC – Brockmann Consult GmbH (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 - Stichting VU-VUmc (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)	mlucaresia.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	
	Duarte Oom (ISA)	duarte.oom@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		
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		

	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
	Algorithm Theoretical Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2	Page	3		

Summary

This document describes the algorithm used for generating the small-fire dataset within the Fire_cci project, with data from the MSI sensor on board the Sentinel-2 satellite.

	Affiliation/Function	Name	Date
Prepared	EHU EHU	Aitor Bastarrika Ekhi Roteta	28/07/2018
Reviewed	UAH – Project Manager	Lucrecia Pettinari	01/10/2018
Authorized	UAH - Science Leader	Emilio Chuvieco	01/10/2018
Accepted	ESA - Technical Officer	Stephen Plummer	

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

Document Status Sheet

Issue	Date	Details
0.1	01/05/2017	Internal draft release of the document.
1.0	01/10/2018	First public release of the document

Document Change Record

Issue	Date	Request	Location	Details
1.0	01/10/2018	UAH	All document	Removed references to the S-1 algorithm, which are included in a separate document, and Landsat-8 tests.
		ESA	Section 3	Text expanded.
		ESA	Sections 4.1 and 4.2	Sections merged and text reduced.
		ESA, UL	Section 4.2	The previous section 4.6 was moved to section 4.2 and expanded.
		ESA	Section 4.3, 4.6.4, 4.6.8, 5.1	Small changes in the text.
		UL	Sections 4.5, 5.2	Sections updated.
		ESA	Section 4.6.3	Text expanded.
		ESA	Section 4.6.7	Clarification added on threshold selection.
		UL	Figure 14	Figure updated.
UAH, UL	Section 4.6.9	New section added.		
ESA-UL	Section 6	Text expanded, and merged with previous Section 7		




fire
cci

Fire_cci
Algorithm Theoretical Basis
Document – Small Fires Dataset: S2

Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
Issue	1.0	Date	01/10/2018
Page			4

Table of Contents

1	Executive Summary	6
2	Introduction.....	6
2.1	Background.....	6
2.2	Purpose of the document.....	7
2.3	Applicable Documents.....	7
3	The Small Fire Dataset	7
4	Data and Methods	8
4.1	Sentinel-2 products	8
4.2	Active fire products	8
4.3	Methodology for Site Selection with S-2	9
4.4	Spectral analysis of burned areas.....	12
4.5	Classification of Burned and Non-Burned Vegetation.....	15
4.5.1	Preparation of data.....	15
4.5.2	General overview of the BA algorithm	19
4.5.3	Not burnable mask.....	21
4.5.4	Initially Burned/Not Burned Area detection (IB/INB).....	22
4.5.5	Initially Burned (Not) Confirmed (IBC/IBNC).....	23
4.5.6	Burned seeds.....	23
4.5.7	Second Stage Probability of Burn (SEPB)	23
4.5.8	Probability of burn results	25
4.5.9	Resampling of the Probability of burn	27
5	Results of Test over the study areas	28
6	Limitations and Next Steps	30
7	References	30
	Annex: Acronyms and abbreviations	32


	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
	Algorithm Theoretical Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2		Page	5	

List of Tables

Table 1. Selected Sentinel-2 image pairs, the mapped burned area and the cloud cover	10
Table 2. List of spectral indices computed for the spectral analysis	12
Table 3. Burned versus Water/shadows/cloud separability for NBR2, MIRBI and NIR variables.....	14
Table 4. SCL Classification map created by Sen2cor algorithm.....	16
Table 5. fMask classification map.....	17
Table 6. Rules applied to obtain the not burnable mask.....	21
Table 7. Rules applied to obtain the Initially Burned Result	22
Table 8. Rules applied to obtain Burned Seeds	23
Table 9. Lookup Table for the rescaling process.....	28
Table 10. Quality assessment of the S2 algorithm in the 52 study areas.	28

List of Figures

Figure 1. Comparison of Landsat and Sentinel 2 spectral bands.....	8
Figure 3. For each S2 granule a) Mean cloud percentage b) Burned area sum in km ² derived from the Fire_cci BA product.....	9
Figure 4. Location of S2 study areas	10
Figure 5. Burned / Not Burned M separability for the best 10 bands	13
Figure 6. Boxplot for post and multitemporal difference of a) NBR2, b)MIRBI and NIR for Burned, Not Burned, Water, cloud shadow and cloud	14
Figure 7. Comparison of a false colour image (SWIR2-NIR-Red) and the Scene Classification of the same area, where the cloudy area on the right is not detected (it is labeled as Bare Soil).....	16
Figure 8. a) SWIR2-NIR-Red false color b) SCL obtained with previous sen2cor version (2.0.6) c) SCL obtained with the latest sen2cor version (2.2.1)	17
Figure 9. Comparison of SCL and Fmask; a), b) and c) represent an area of the tile 28PET, while d), e) and f) another of 36NVP. a) and d) are SWIR2-NIR-Red colour compositions, b) and e) the SCL for each area, and c) and f) the Fmask cloud masks.	18
Figure 10. Compositing of two images of the same date in the 33PTK study area of both false colour SWIR2-NIR-Red image and SCL (the right column is the compositing result).....	19
Figure 11. Aggregation process for the Sentinel 2 images.....	20
Figure 12. BA algorithm for each pair of Images considered	21
Figure 13. s-shape and z-shape sigmoid membership functions applied to MIRBI and NBR2 temporal differences, in tile 28PET between December 22 2015 and January 11 2016.	24
Figure 14. Extension of seeds' influence area in the SEPB for a sample zone of the 30PWQ study area, between 27 th December 2015 and 6 th January 2016; the legend is the same for all the maps.	26
Figure 15. Evolution of omission and commission errors depending on the threshold in two sample tiles (31PEN to the left and 34PET to the right).	27

	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
	Algorithm Theoretical Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2		Page	6	

1 Executive Summary

Fires emit greenhouse gases (GHGs) and aerosols, important climate forcing factors which need to be estimated and modelled to better understand climate and carbon cycling. Fires are also a major factor in land cover change, and hence affect fluxes of energy and water to the atmosphere. In this context, spatial and temporal monitoring of trace gas emissions from fires is of primary importance. These can be inferred using both land-surface and atmospheric measurements, preferably in combination. The Fire Disturbance Essential Climate Variable provides baseline products for the land-surface to allow this.

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 parameterise climate-driven models for BA (GCOS, 2016).

This document is the Algorithm Theoretical Basis Document (ATBD) corresponding to the generation of the small fires dataset of Phase 2 [AD-1]. It describes the algorithm method and approach that has led to the generation of the small fire dataset for the continent of Africa for the Phase 2 of the Fire_cci project, using optical data. Test sites have been selected that are representative of the vegetation that is found in Africa that burns on a regular basis. The theoretical basis described here identifies the data sets that were used to classify burned area and the methods used to derive the BA products.


2 Introduction

2.1 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 to produce consistent time series of accurate Essential Climate Variables (ECV) products, which can be used by the climate, atmospheric and ecosystem scientists for their modelling efforts. The importance of keeping long-term observations and the international links with other agencies currently generating ECV data is also stressed.

The fire disturbance ECV identifies BA (BA) as the primary fire variable. Accordingly, the Fire_cci project shall focus on developing and validating algorithms to meet GCOS ECV requirements for (consistent, stable, error-characterised) global satellite data products from multi-sensor data archives.

All global BA products are based on coarse resolution sensors (from 250 to 1000m). Therefore, the likelihood of detecting small burns (i.e. < 50ha) is very low, and therefore omission errors from these products may be quite high (Padilla et al., 2015; Giglio et al., 2009), particularly coming from small fires (Randerson et al., 2012). In order to improve the characterisation of small fires, one of the objectives of the project is to generate a small fires dataset based on medium resolution sensors (10 to 100m). Considering the huge processing effort to cover the whole planet, this SFD shall be focused on the African continent, which is the most burned worldwide (Chuvieco et al, 2016; Giglio et al., 2013). The dataset will add a great value to the global BA products to be generated within the Fire_cci project, as well as to the global characterization of fire activity.

	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
	Algorithm Theoretical Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2	Page	7		

2.2 Purpose of the document

This document describes the algorithm theoretical basis from which to map burned areas from moderate resolution optical data. It further provides initial results of a mapping exercise in specially selected test sites in Africa.

2.3 Applicable Documents

[AD-1]	ESA Climate Change Initiative (CCI) Phase 2 Statement of Work, prepared by ESA Climate Office, Reference CCI-PRGM-EOPS-SW-12-0012, Issue 1.3, date of issue 24 March 2015.
--------	--

3 The Small Fire Dataset

Burned Area (BA) is defined in this document as any vegetated area that has been completely or partially consumed by a fire, regardless of whether that fire was human or natural causes, or whether that fire affected wildland areas or human managed territories (agricultural or pastures).

Since several coarse-resolution burned area detection algorithms require that a substantial fraction of an individual pixel's area undergo burning for successful attribution (to avoid commission errors from other forms of land cover change), detection of small fires becomes difficult (Roy and Landmann 2005). At a global scale, it has been shown that accounting for small fires may increase burned area and global carbon emissions by approximately 35% (Randerson et al. 2012). The small fire dataset will be used to establish whether the estimate by Randerson et al. can be verified (using a continental data set as a first indicator) and address what the contribution of small fires might be.

It is known that the strategy of using active fire products may provide new information on small fire contribution to burned area and emissions because of the strong nonlinearity in radiative power as a function of fire temperature. Active fires have shown very low commission errors, but very diverse omission rates depending on the ecosystem and fire size (36–86%), the detection of those below 500 ha is quite high (omissions < 20%) (Hantson et al. 2013). Given this context, there is a real need to link coarse resolution products with data on small burned patches, in particular in locations where these are the dominant fire types (e.g. sub-Saharan Africa). This could be achieved with higher resolution remote sensing data such as Landsat-OLI (30 m) or Sentinel-2 optical data (10 m).

The capability of Landsat TM and ETM+ data to provide information on burned scars has been widely recognized in scientific literature due to its spatial coverage, medium spatial resolution, multispectral characteristics (covering the most important spectral areas for burned area mapping with one band in the near infrared and two in the shortwave infrared) and specially for the historical dataset, with data available since 1982 until now (the last satellite, Landsat-8, was launched in 2013). Although some local/regional burned area algorithms based on Landsat TM/ETM images have been developed (Bastarrika et al. 2014; Bastarrika et al. 2011b; Chuvieco et al. 2002), recent developments have shown the potential of these sensors even for global BA products. Obviously, global BA algorithms are much more challenging than local based, as the proposals by the USGS (http://remotesensing.usgs.gov/ecv/BA_overview.php, last

accessed July 2018) and Boschetti et al. (2015) have shown. For both, only some partial validation results have been published.

One of the most critical issues in Landsat is the temporal resolution (16 days), not enough for tropical ecosystems where the vegetation recovers very fast. This issue is improved with the Sentinel-2 mission, which allows a high revisit frequency (5 days at the equator when the Sentinel 2A and 2B satellites are working simultaneously). The Multi Spectral Instrument (MSI) that is carried by these satellites provides a unique combination of high spatial resolution (which varies from 10 m to 20 m), a wide field of view (290 km) and spectral coverage (13 spectral bands spanning from the visible and the near infrared to the short wave infrared), and represents a major step forward compared to current multi-spectral missions (Drusch et al. 2012).

4 Data and Methods

4.1 Sentinel-2 products

The data used for the SFD are Sentinel-2 MSI Level-1C product, which contains Top-of-atmosphere reflectances projected in UTM WGS84 system. Bottom-of-atmosphere reflectances and a Scene Classification (SCL) are generated by the Sentinel-2 ESA toolbox at the beginning of the BA classification. However, due to the limited availability of Sentinel-2 MSI imagery until December 2016, we first used Landsat-8 OLI images to generate the preliminary versions of the algorithm in selected study areas across Africa, as both sensors have similar spectral bands (Figure 1). However, to avoid confusion to the reader, we keep in this document only the algorithm that was developed for Sentinel-2, which was the basis of the final SFD delivered from the Fire_cci project.

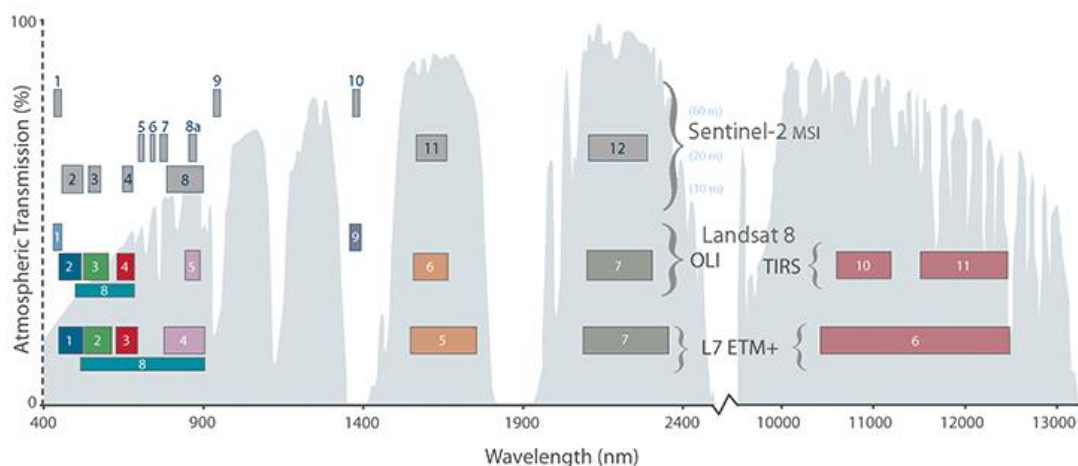



Figure 1. Comparison of Landsat and Sentinel 2 spectral bands

4.2 Active fire products

The proposed Sentinel-2 BA algorithm uses the active fire products to check that the spectral/temporal changes detected at the images are due to the fire. While the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor, aboard the Suomi-NPP satellite with its 375 meter product would be most adequate due to its better spatial resolution and a greater response over fires of relatively small areas, its active fire product was not

	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
	Algorithm Theoretical Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2	Page	9		

available at the beginning of this algorithm’s development and another hotspot product had to be used. Concretely, the vector file of Collection 6 MCD14DL hotspots downloadable by the Archive Download Tool (<https://firms.modaps.eosdis.nasa.gov/download/>, last accessed July 2018) has been used.

4.3 Methodology for Site Selection with S-2

Several factors were taken into account when selecting S2 study areas and the dates of the corresponding images:

- The cloud coverage (CC), generated from the MOD08 product during 3 years (2012-2014), applied to S2 tiles. Areas with not too many clouds are needed, so that optical S2 images can observe burned areas and can also be used to assessed how the algorithm works on cloudy areas (Figure 2a).
- The burned areas (BA), generated from the results of the Phase 1 of the Fire_cci project between 2006 and 2008 (Figure 2b).

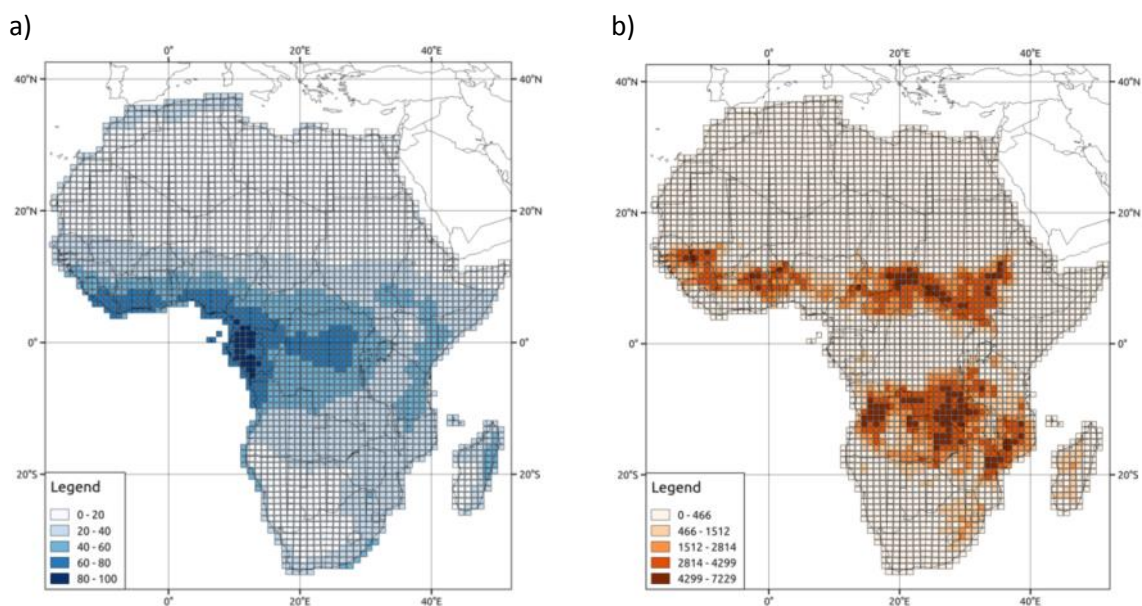


Figure 2. For each S2 granule a) Mean cloud percentage b) Burned area sum in km2 derived from the Fire_cci BA product.

In this case, the study areas for Sentinel-2 data were selected in a systematic way from areas where fires were expected, looking for cloud-free images at the same time. Due to difficulties to find pair of images with observable burned area between them, the temporal interval varies considerably; the difference between both dates is 10 or 20 days in more than half of the study areas, but it rises up to 4 months in one case. The location of all these areas, 52 totally, is represented in Figure 3.

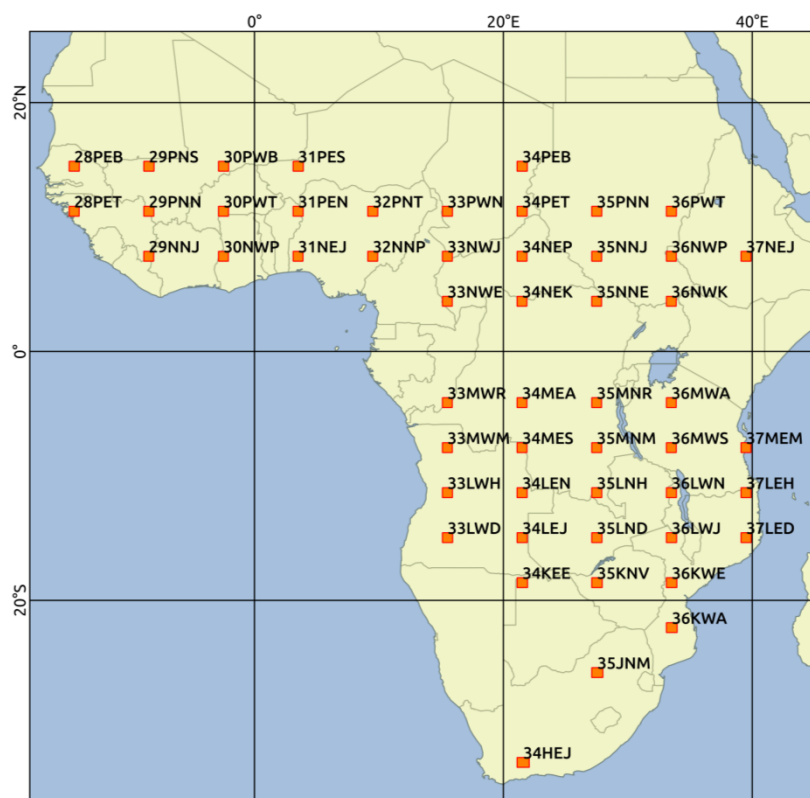


Figure 3. Location of S2 study areas

Burned area perimeters have been generated using the Burned Area Mapping Software (BAMS) (Bastarrika et al., 2014) in the same way as in Landsat-8 study areas (downloadable at <https://ehubox.ehu.eus/s/WoBHJaAdWzBYtSM>, last accessed July 2018). Some data about the scenes are described in the Table 1; surfaces covered by clouds have been discarded in order to calculate the percentage of burned areas.

Table 1. Selected Sentinel-2 image pairs, the mapped burned area and the cloud cover

Tile	Pre-fire date	Post-fire date	Burned area (km ²) and %	Cloud coverage %
28PEB	2015/12/22	2016/04/20	164 (1,4%)	0.0
28PET	2015/12/22	2016/04/20	1206 (11,6%)	14.0
29NNJ	2016/01/11	2016/03/11	1139 (10,3%)	7.9
29PNN	2016/01/02	2016/03/02	1012 (8,8%)	5.0
29PNS	2016/02/11	2016/04/21	23 (0,2%)	11.2
30NWP	2016/01/25	2016/03/15	543 (6,4%)	29.6
30PWB	2015/12/27	2016/01/16	22 (0,2%)	4.8
30PWT	2016/02/05	2016/03/26	1095 (9,1%)	0.4
31NEJ	2015/12/27	2016/01/16	909 (7,6%)	0.3
31PEN	2015/12/18	2016/01/07	819 (11,8%)	42.3



fire
cci

Fire_cci
Algorithm Theoretical Basis
Document – Small Fires Dataset: S2

Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
Issue	1.0	Date	01/10/2018
Page			11

Tile	Pre-fire date	Post-fire date	Burned area (km ²) and %	Cloud coverage %
31PES	2015/12/28	2016/01/17	1 (0,0%)	0.2
32NNP	2016/01/20	2016/02/19	3609 (30,3%)	1.2
32PNT	2015/12/22	2016/02/10	293 (2,4%)	0.6
33LWD	2016/01/01	2016/02/10	1666 (16,3%)	15.2
33LWH	2016/08/03	2016/08/23	1423 (12,5%)	5.9
33MWM	2016/05/18	2016/06/07	1958 (20,8%)	21.9
33MWR	2016/05/18	2016/06/07	449 (6,2%)	39.8
33NWE	2016/08/06	2016/09/05	860 (12,4%)	42.6
33NWJ	2016/01/02	2016/01/22	2044 (21,5%)	21.1
33PWN	2016/01/02	2016/01/22	849 (7,4%)	5.0
34HEJ	2015/12/26	2016/01/25	112 (0,9%)	1.1
34KEE	2016/01/04	2016/03/04	535 (4,5%)	0.5
34LEJ	2016/08/27	2016/09/06	526 (4,4%)	1.1
34LEN	2016/07/18	2016/08/07	2822 (26,9%)	13.0
34MEA	2016/05/29	2016/06/18	415 (14,2%)	75.8
34MES	2016/05/12	2016/06/11	3985 (40,0%)	17.4
34NEK	2016/05/12	2016/06/11	406 (10,9%)	69.1
34NEP	2015/12/27	2016/01/26	4919 (46,6%)	12.4
34PEB	2015/12/27	2016/02/05	0 (0,0%)	1.4
34PET	2015/12/30	2016/01/19	1150 (10,1%)	5.5
35JNM	2015/12/27	2016/02/05	67 (0,7%)	18.4
35KNV	2016/07/19	2016/08/08	135 (1,1%)	0.2
35LND	2016/09/07	2016/09/27	1274 (11,1%)	4.7
35LNH	2016/07/02	2016/07/22	1893 (16,4%)	4.4
35MNM	2016/07/02	2016/07/22	1597 (13,8%)	3.9
35MNR	2016/07/02	2016/07/12	116 (5,1%)	81.0
35NNE	2016/04/06	2016/05/16	2017 (30,7%)	45.5
35NNJ	2016/01/17	2016/04/06	997 (10,9%)	24.1
35PNN	2015/12/31	2016/01/30	76 (0,6%)	1.8
36KWA	2015/12/31	2016/01/10	21 (0,2%)	15.1
36KWE	2016/07/13	2016/07/23	1507 (14,0%)	10.9
36LWJ	2016/07/23	2016/08/22	1104 (9,3%)	1.8
36LWN	2016/08/12	2016/08/22	377 (4,2%)	25.3
36MWA	2016/07/16	2016/07/26	7 (0,1%)	0.6
36MWS	2016/06/26	2016/07/16	1638 (14,0%)	2.7
36NWK	2016/06/16	2016/07/16	4101 (42,2%)	19.4
36NWP	2015/12/22	2016/03/11	2026 (38,9%)	56.8

Tile	Pre-fire date	Post-fire date	Burned area (km ²) and %	Cloud coverage %
36PWT	2015/12/22	2016/01/11	136 (1,4%)	17.4
37LED	2016/01/04	2016/01/24	570 (5,4%)	11.7
37LEH	2016/09/05	2016/09/15	876 (10,2%)	28.5
37MEM	2016/07/27	2016/08/16	87 (2,7%)	73.0
37NEJ	2016/07/10	2016/07/20	3 (0,1%)	59.2

4.4 Spectral analysis of burned areas

Before developing the burned area algorithm, a spectral analysis of burned areas was carried out, with the intention of establishing which the best spectral bands or indices are in order to distinguish burned areas from the rest of land covers. As Sentinel-2 imagery was not available yet when this analysis was made, Landsat-8 imagery was used; that should not pose any problem due to the similarities between L8 and S2 spectral bands.

In addition to the original Landsat 8 bands (Blue, Green, Red, NIR, SWIR1 and SWIR2) the spectral indices in Table 2 were assessed.

Table 2. List of spectral indices computed for the spectral analysis

Acronym	Description	Spectral space
EVI	Enhanced vegetation index	Visible / NIR
NDVI	Normalized Difference Vegetation Index	Visible / NIR
SAVI	Soil Adjusted Vegetation Index	Visible / NIR
GEMI	Global Environmental Monitoring index	Visible / NIR
BAI	Burned Area Index	Visible / NIR
GEMI7S1	Global Environmental Monitoring index	NIR / Short SWIR
EVI7S1	Enhanced vegetation index	NIR / Short SWIR
NBRS1	Normalized Burned Ratio	NIR / Short SWIR
BAIMS1	Modified Burned Area Index	NIR / Short SWIR
GEMI7S2	Global Environmental Monitoring index	NIR / Long SWIR
EVI7S2	Enhanced vegetation index	NIR / Long SWIR
NBRS2	Normalized Burned Ratio	NIR / Long SWIR
BAIMS2	Modified Burned Area Index	NIR / Long SWIR
MIRBI	Mid- Infrared Burned Index	Short SWIR / Long SWIR
NBR2	Normalized Burned Ratio	Short SWIR / Long SWIR

Using validation perimeters obtained for 29 Landsat-8 scenes, 100.000 burned / not burned / cloud / shadow samples were extracted for each scene, in order to compute the



fire
cci

separability index between the burned and the other three categories, following the next equation

$$M = \frac{abs(\mu_A - \mu_B)}{\sigma_A + \sigma_B}$$

Where:

μ_A = Mean of the variable for category A

σ_A = Standard deviation for for category A

μ_B = Mean of the variable for the category B

σ_B = Standard deviation for category B

For the seven spectral bands (Blue, Green, Red, NIR, SWIR1 and SWIR2) and the 15 spectral indices, both the Post Variable (the value on the post-fire date for the corresponding band / spectral index) and the Multitemporal Difference Variable (the difference of values between post-fire and pre-fire dates on the validation period) were assessed. Figure 4 shows the 10 variables with the highest separability index. Note that although MIRBI and NBR2 are computed from the same two SWIR bands and show the highest separability among burned and unburned category, their behaviour in comparison with other covers like clouds and shadows is different (Table 3, Figure 5). The MIRBI multitemporal difference shows quite good separability with clouds (0.96) but poorer with cloud shadows (0.65), while the performance for NBR2 shows higher separability for cloud shadows (0.77) and lower for clouds (0.66). It is remarkable the good separability of the NIR post band with the water (1.55), due to the difference between the low reflectance of burned areas but nearly null value of water in this band.

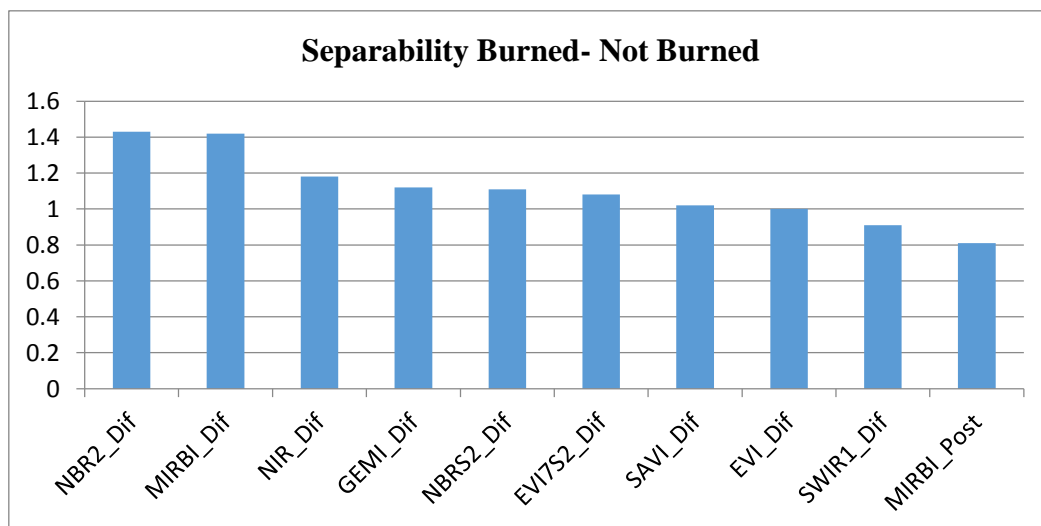
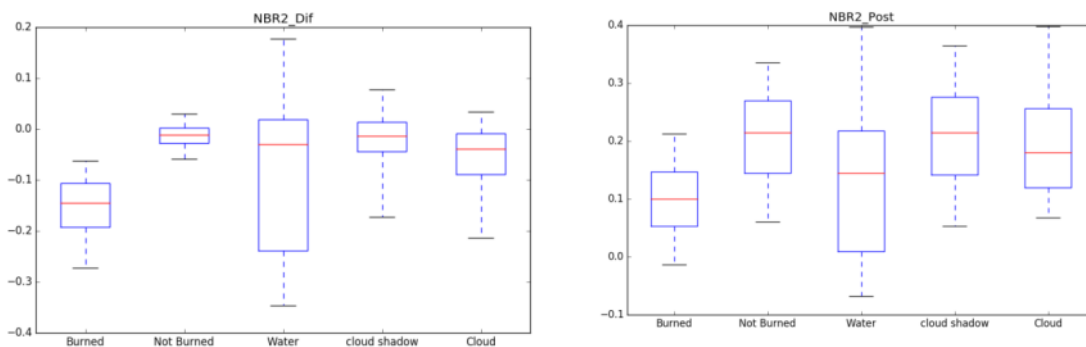


Figure 4. Burned / Not Burned M separability for the best 10 bands

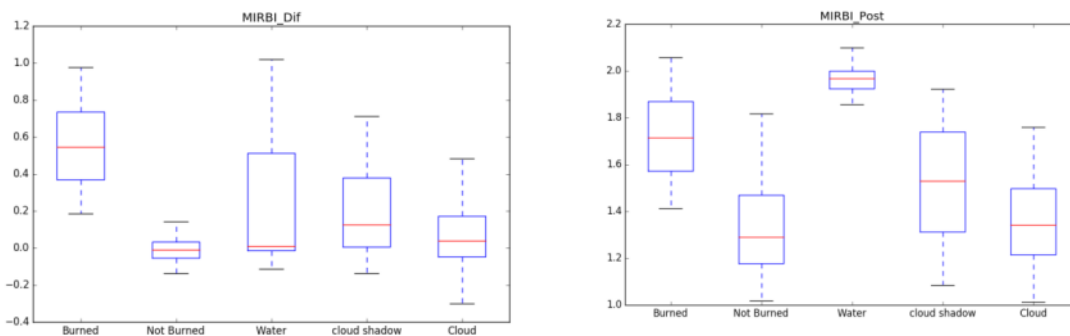
Table 3. Burned versus Water/shadows/cloud separability for NBR2, MIRBI and NIR variables

Variable	M Separability index			
	Burned vs Not Burned	Burned vs Water	Burned vs Cloud shadows	Burned vs Cloud
NBR2_Dif	1,43	0,16	0,77	0,66
MIRBI_Dif	1,42	0,47	0,65	0,96
NIR_Dif	1,18	0,33	0,29	0,75
MIRBI_Post	0,81	0,69	0,4	0,79
NIR_Post	0,79	1,55	0,25	0,78
NBR2_Post	0,69	0,11	0,58	0,56

a)



b)



c)

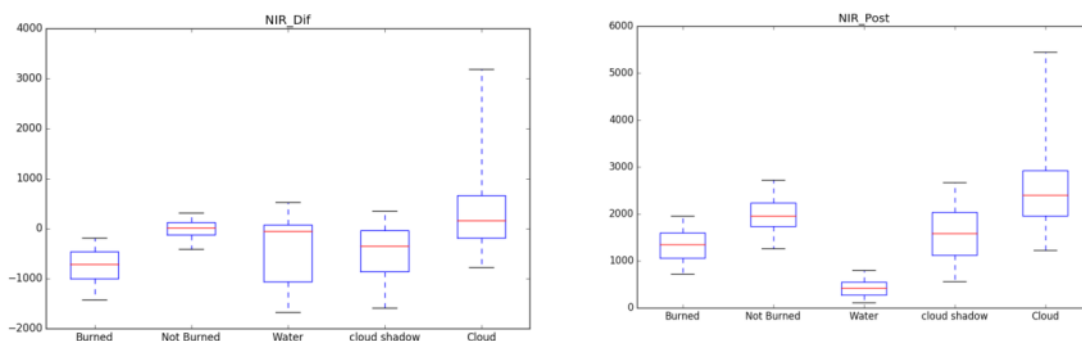



Figure 5. Boxplot for post and multitemporal difference of a) NBR2, b)MIRBI and NIR for Burned, Not Burned, Water, cloud shadow and cloud

	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
	Algorithm Theoretical Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2		Page	15	

4.5 Classification of Burned and Non-Burned Vegetation

4.5.1 Preparation of data

The pre-processing includes all the necessary steps to prepare the input data (Level-1C products downloaded from the Web) so that the algorithm can be applied directly on its output. This includes several phases:

- Generation of the Level-2A product from Level-1C using the sen2cor tool (Sentinel-2 Toolbox).
- Spectral index computation.
- Masks (SCL and fmask version 2015).
- Compositing of images for the same date.

4.5.1.1 Generation of the Level-2A product

The Sentinel-2 images downloaded from the Sentinels Scientific Data Hub (<https://scihub.copernicus.eu>, last accessed July 2018) are a Level-1C product, with regular 100x100 km tiles, geometrically corrected ortho-images (using a Digital Elevation Model) in UTM coordinates and WGS84 reference system. The pixel values are Top of Atmosphere (TOA) reflectances, which means that no atmospheric correction have been applied.

Sen2cor is a command-line-working tool of the Sentinel-2 Toolbox (downloadable at <http://step.esa.int/main/third-party-plugins-2/sen2cor/>, last accessed July 2018) which processes L1C products and generates a L2A product with Bottom of Atmosphere (BOA) reflectance images, as well as a Scene Classification (SCL). This new product has the same format as the L1C product: same tiling system, JPEG2000 format and three different resolutions (10, 20 and 60m).

According to the previous spectral analysis of burned areas, the NIR band and NBR2 and MIRBI spectral indices showed the highest separability, so only the bands B8A (NIR), B11 (SWIR1) and B12 (SWIR12) were required for the algorithm –though the band B04 (red) was also used to create colour composites for visual analysis–, whose original spatial resolution is 20m (10m for B04). For that reason, a L2A product with a 20 m spatial resolution was generated. For the NIR, the band B8 might be more appropriate than the B8A, but it is not generated within the L2A product (B01, B09 and B10 bands are not created either on this product) so B8A was used instead.

4.5.1.2 Comparison of SCL and fMask

The SCL classifies the entire image into 12 categories (Table 4), which consist of 4 different classes for clouds (including cirrus), together with six different classifications for shadows, cloud shadows, vegetation, soils / deserts, water and snow (<http://step.esa.int/thirdparties/sen2cor/2.2.1/S2PAD-VEGA-SUM-0001-2.2.pdf>, last access July 2018).

Table 4. SCL Classification map created by Sen2cor algorithm

Label	Classification
0	No Data
1	Saturated or defected
2	Dark area pixels
3	Cloud shadows
4	Vegetation
5	Bare soils
6	Water
7	Cloud low probability
8	Cloud medium probability
9	Cloud high probability
10	Thin cirrus
11	Snow

Analysing the classification result in the selected Sentinel 2 study areas, the following issues have been detected:

- Many thin clouds (and their consequent shadows) are not detected and are classified as vegetation or bare soils, classifying as clouds only the thickest ones; at the same time, some pixels of these clouds are classified as water (Figure 6).
- Very burned areas classified as dark area pixels or as cloud shadows, especially in the savannah (Figure 7). This error has considerably increased with the new version 2.2.1 of the sen2cor tool, released on May 4, 2016 (Figure 7c), while it was slighter in the older version 2.0.6 released on December 2015 (Figure 7b).
- Bordering the dark area pixels, several burned areas are labelled by mistake as Low probability clouds (Figure 7c).

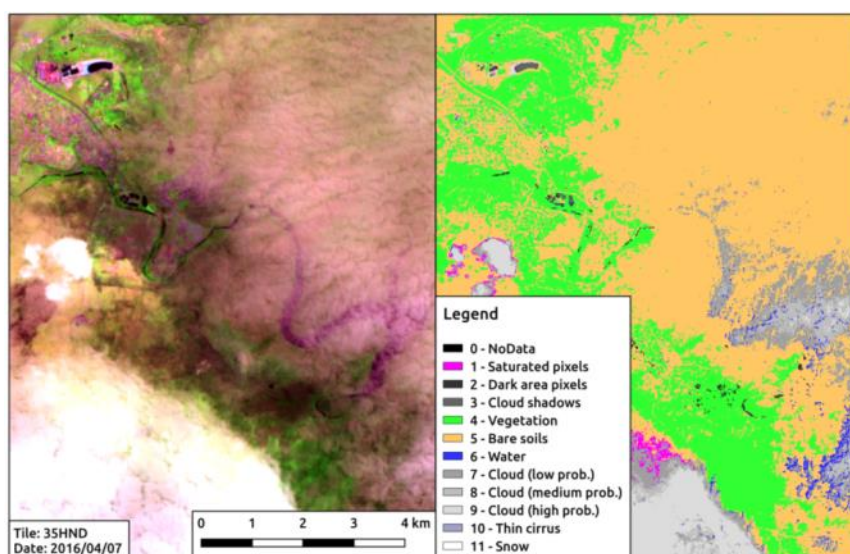


Figure 6. Comparison of a false colour image (SWIR2-NIR-Red) and the Scene Classification of the same area, where the cloudy area on the right is not detected (it is labeled as Bare Soil).

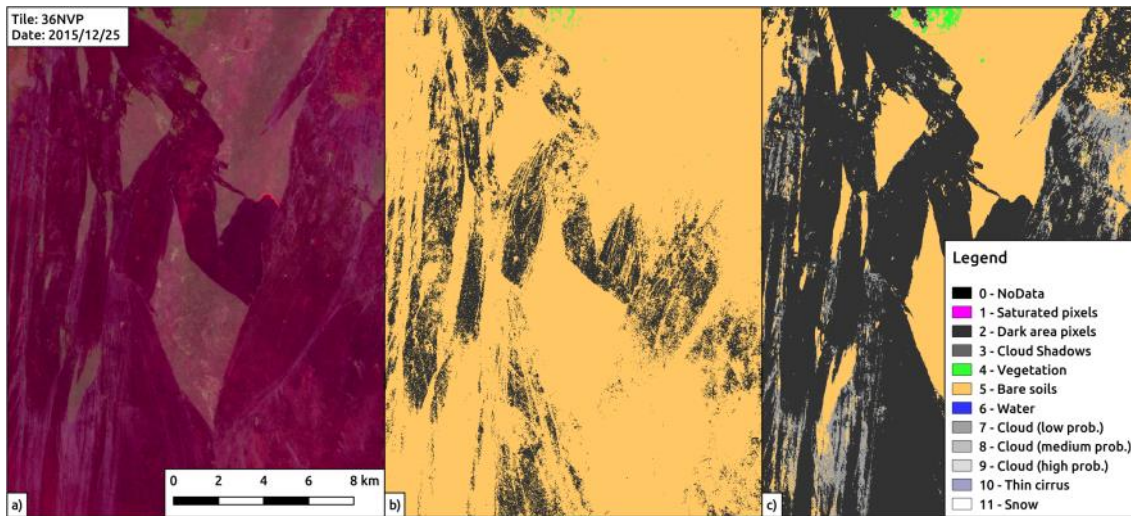


Figure 7. a) SWIR2-NIR-Red false color b) SCL obtained with previous sen2cor version (2.0.6) c) SCL obtained with the latest sen2cor version (2.2.1)

Trying to solve the detected problems with the SCL, the Fmask algorithm developed by Zhu & Woodcock (2014) to generate a cloud mask from both Landsat 4-8 and Sentinel-2 imagery was tested, whose Python module can be downloaded from <http://pythonfmask.org> (last accessed July 2018). This mask is simpler than the SCL, with the 6 categories in Table 5 considered.

The visual comparison of fmask classification results does not seem to improve the SCL obtained by sen2cor (Figure 8). Besides generalizing and rounding the mask, it does not resolve the problems that SCL showed:

- Clouds are well detected, it detects them even when they are not visible enough for a human eye to see in the false colour image, but most cloud shadows are not masked.
- Pixels with strong burned reflectance values not classified as ground, but as water.

In addition, the python implementation of the algorithm sometimes fails due to an internal error and does not create any cloud mask, henceforth only SCL will be used.

Table 5. fMask classification map

Label	Classification
0	No data
1	Ground
2	Cloud
3	Cloud shadows
4	Snow
5	Water

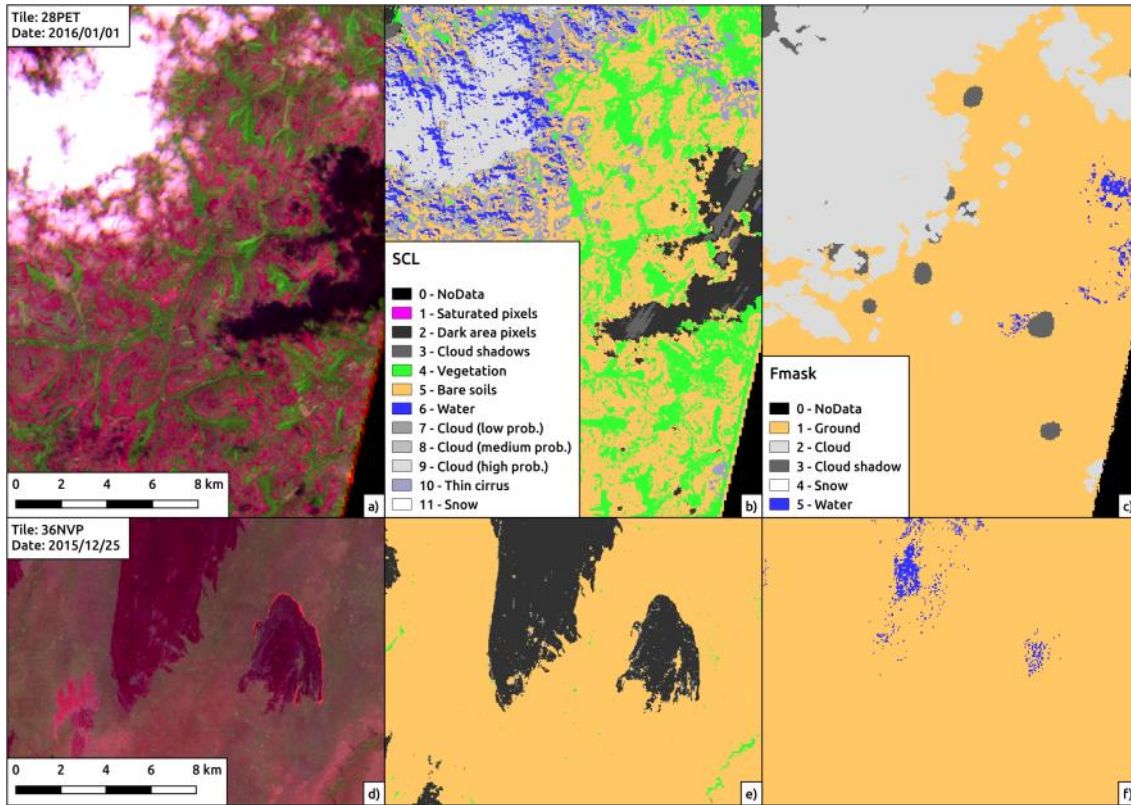


Figure 8. Comparison of SCL and Fmask; a), b) and c) represent an area of the tile 28PET, while d), e) and f) another of 36NVP. a) and d) are SWIR2-NIR-Red colour compositions, b) and e) the SCL for each area, and c) and f) the Fmask cloud masks.

4.5.1.3 Spectral indices

Two spectral indices were calculated: the Mid-Infrared Burned Index (MIRBI) and the Normalized Burned Ratio 2 (NBR2), both representing the relation between two very important spectral spaces for the burned signal detection, Long SWIR (B12) and Short SWIR (B11). These two spectral indices were selected because they got the highest separabilities in the spectral analysis of the BA reference perimeters between burned and not burned areas using Landsat data, along with the NIR band.

The NBR2 and MIRBI spectral index equation is as follows:


$$NBR_2 = \frac{\rho_{SWIRS} - \rho_{SWIRL}}{\rho_{SWIRS} + \rho_{SWIRL}}$$

$$MIRBI = (10 \cdot \rho_{SWIRL} - 9.8 \cdot \rho_{SWIRS} + 2)$$

Where:

- ρ_{SWIRS} = Short Short Wave Infrared Short reflectance [unitless] (B11 band divided by 10000 in the case of Sentinel 2)
- ρ_{SWIRL} = Short Wave Infrared Long reflectance [unitless] (B12 band divided by 10000 in the case of Sentinel 2)

For the NIR, the B8A band was used (divided by 10000 in the case of Sentinel 2).

	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
	Algorithm Theoretical Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2		Page	19	

4.5.1.4 Compositing of scenes of the same date

Some scenes, even if they were acquired in the same orbit at the same time, are split in two different files. This issue is quite easy to detect, as the *Sensing Time Start* and the *Sensing Time Stop* in the products filename is the same. In this case, both images must be joined using the corresponding SCLs as masks, filling the gap of an image (where SCL has the value 0 – NoData) with data from the other image (Figure 9).

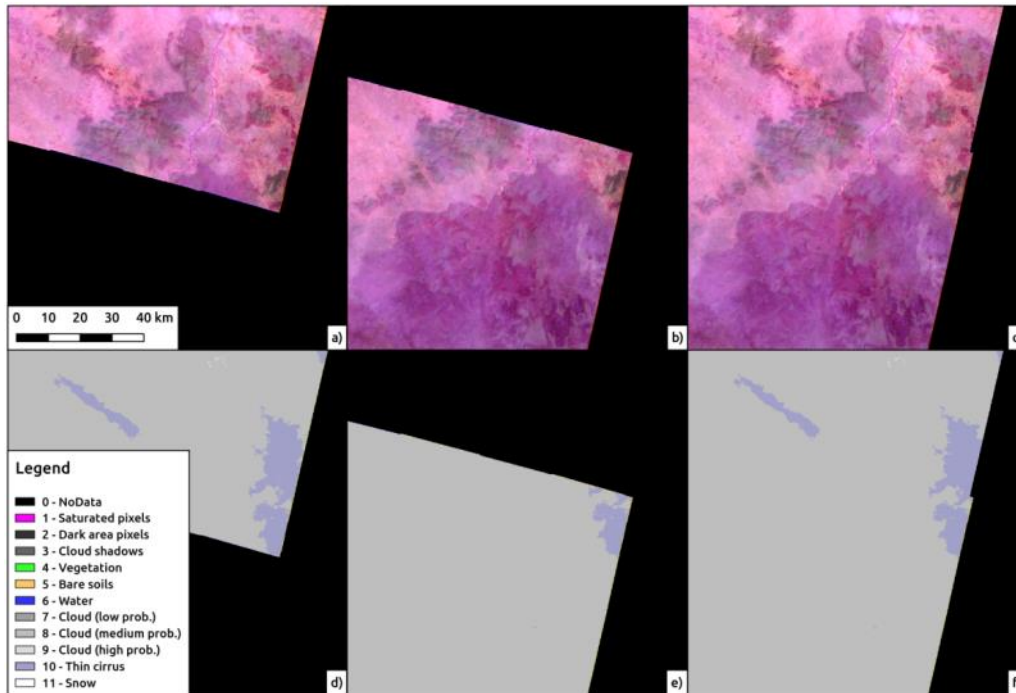


Figure 9. Compositing of two images of the same date in the 33PTK study area of both false colour SWIR2-NIR-Red image and SCL (the right column is the compositing result)

4.5.2 General overview of the BA algorithm

The proposed burned area algorithm compares two Sentinel 2 tiles each time (this algorithm it is also valid for Landsat data), using 6 variables, the multitemporal difference $[t] - [t-i]$ and the post values $[t]$ of the MIRBI, NBR2 spectral indices and the NIR, the variables that have shown a higher separability between the burned and not burned category on the spectral analysis using Landsat data (Table 3).

Each scene $[t]$ time is compared to the previous four scenes $[t-1, t-2, t-3, t-4]$ in order to complete the areas masked as not burnable in previous images (Figure 10). This criteria means that if $[t-1]$ image is cloud/shadows free, the previous scenes $[t-2, t-3, t-4]$ will not be used. In the current context where only S2A is working with a temporal resolution of 10 days (the S2B launched in March 2017 will ensure a temporal resolution of 5 days), we assume at most going back 40 days as the algorithm defined works better while the compared images are more similar. Our findings have shown high commission errors when comparing scenes older than 40 days in some study areas.

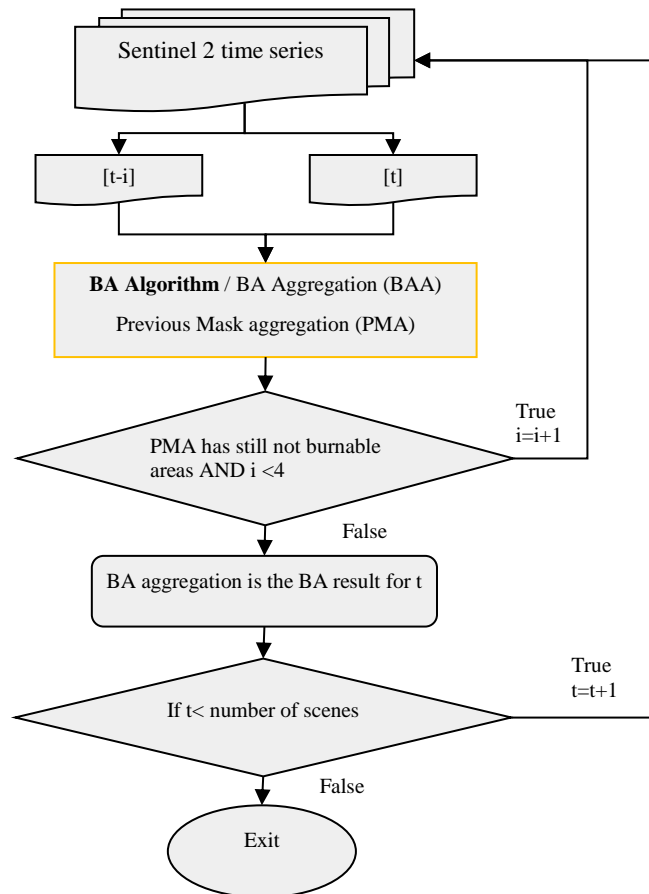


Figure 10. Aggregation process for the Sentinel 2 images

When comparing two scenes the algorithm starts computing the not burnable mask using, both [t] and [t-i] scene classification map (SLC). If the burnable area is below 5 km² or no MODIS Active Fire Hotspots have been detected in the scene extent during the [t-i, t] interval, the process ends (Figure 11).

If the flow continues, the algorithm starts applying fixed thresholds (see Section 0) for the MIRBI NBR2 and NIR multitemporal difference, and image based statistics for the MIRBI, NBR2 and NIR [t] images, obtaining the “Initial Burned/Not Burned” – IB and INB- cartography. The initially burned regions, once labelled using a connectivity of 2 (8 neighbour pixels), are crossed with the MODIS Active Fire hotspots points among the Pre and Post scene dates in order to have a confirmation that those regions are really burned, obtaining the “Initial Burned/Not Burned Confirmed”- IBC and INBC. These hotspots, initially read as vector points, are rasterized to the Sentinel spatial resolution (20m) and dilated to have a 1 km of diameter, the spatial resolution of the original MODIS active fire product. The BA process ends if no initially burned regions have been confirmed (IBC).

At this point of the algorithm, a two phase strategy is applied to balance omission and commission errors (Bastarrika, et al., 2011a). Initially Burned Confirmed (BC) based statistics (5th or 95th percentile depending on if the burned signal is higher or lower than the not burned one) are used to set the criteria to establish the seeds, based on the aforementioned 6 variables. The second stage criteria is based exclusively on two variables, the MIRBI and NBR2 multitemporal difference, where a sigmoid membership function is computed based on IB, INB, IBC, INBC statistics depending on

the case (two different options are considered depending on the separability between the confirmed and not confirmed burned areas). The process ends maintaining only the second stage result regions which have been marked by the seeds.

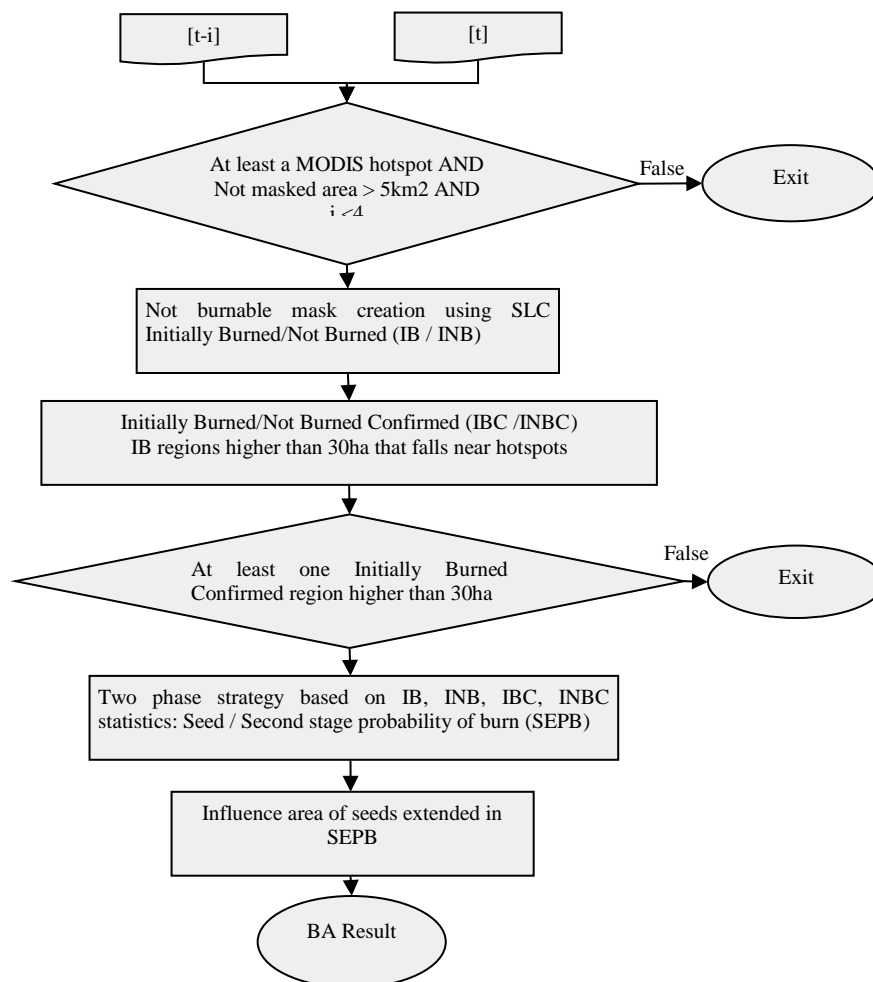


Figure 11. BA algorithm for each pair of Images considered

4.5.3 Not burnable mask

A not burnable mask was defined for each pair of images compared [t-i, t]. A pixel that satisfies any of the following rules was masked (Table 6):

Table 6. Rules applied to obtain the not burnable mask

[t] (Post image)	[t-i] (Previous image)
NO_DATA(0)	NO_DATA(0)
SATURATED_OR_DEFECTIVE (1)	SATURATED_OR_DEFECTIVE (1)
WATER (6)	WATER (6)
CLOUD_MEDIUM_PROBABILITY (8) with a dilation of 5 pixels	CLOUD_MEDIUM_PROBABILITY (8) with a dilation of 5 pixels
CLOUD_HIGH_PROBABILITY (9) with a dilation of 5 pixels	CLOUD_HIGH_PROBABILITY (9) with a dilation of 5 pixels
THIN_CIRRUS (10) with a dilation of 5 pixels	with a dilation of 5 pixels

[t] (Post image)	[t-i] (Previous image)
SNOW (11)	SNOW (11)
Long SWIR (B12) < 0.07	

Summarizing, in addition to not valid data (No data, saturated-defective), Medium-High probability clouds and cirrus with a dilation of 5 pixels (100 metres) are used in order to have a more reliable cloud mask, applied to the image pair considered. Low probability clouds are not used because as previously said, it would be masked a very significant quantity of burned areas. The “dark areas” are not masked for the same reason, and the Long SWIR band (B12) criteria is used, only in the Post image, where in general burned areas and shadows show a good separability (lower values for cloud shadows and higher for burned ones).

When only a reduced surface is observed between a pair of images, a considerable proportion of initially burned pixels (from the next step) can actually be cloud shadows, and the final burned areas will be estimated from wrong statistics extracted from those initially burned pixels. To avoid this, a pair of images will be processed only if there are at least 5 km² (12500 pixels) not burnable; this is an attempt to avoid wrong burned statistics. All the statistics computed in the next sections do not use the not burnable masked pixel values.

4.5.4 Initially Burned/Not Burned Area detection (IB/INB)


The initialization of the algorithm is crucial and it is based on the spectral analysis made with Landsat 8 data (Table 3), where NBR2, MIRBI and NIR multitemporal difference variables showed higher spectral separability between burned and unburned categories. Values dividing these two categories best are used as fixed thresholds. A pixel will be labelled as initially burned (IB/INB) if it satisfies the next 7 rules in Table 7.

The aim behind this first initialization is double. On one hand, to have burned candidates that will be crossed with the active fires, when they will become Initially Burned Confirmed (IBC). On the other hand, to obtain a good identification of the not burned background that will be used as comparison surface at the refinement process.

Table 7. Rules applied to obtain the Initially Burned Result

Rule number	Criteria
1	$MIRBI_t > \text{mean}(MIRBI_t)$
2	$(MIRBI_t - MIRBI_{t-i}) > 0.25$
3	$NBR2_t < \text{mean}(NBR2_t)$
4	$(NBR2_t - NBR2_{t-i}) < -0.05$
5	$NIR_t < \text{mean}(NIR_t)$
6	$NIR_t - NIR_{t-i} < -0.01$
7	Not masked

NOTE: To compute the mean of the post fire (t) data, only data that are not masked are used

	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
	Algorithm Theoretical Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2		Page	23	

4.5.5 Initially Burned (Not) Confirmed (IBC/IBNC)

The IB pixels were crossed with the MODIS derived hotspots, in order to obtain a confirmation that the spectral change detected on the previous step is due to the fire. For doing that, the IB pixels were labelled in regions (using a connectivity of 2, considering 8 neighbour pixels), and only the regions higher than 30 ha (750 pixels in Sentinel 2) were used to be checked by the active fires. Note that the active fires detections are derived from a resolution of 1 km², which is equivalent to 2500 Sentinel 2 pixels. To take into account the spatial disagreement among the active fire and Sentinel 2 detection, a dilation of 50 pixels (1 km of diameter) was applied to the active fire points rasterized in the Sentinel 2 resolution (20m) from the shapefile. Only the hotspots detected among the Previous [t-i] and Post [t] and located in the tile extent were considered.

4.5.6 Burned seeds

To compute the burned seeds, basic statistics derived from the Initially Burned Confirmed regions are used, specifically the 5th percentile for the variables which burned values are higher than the not burned ones, and the 95th percentile for the opposite case. A pixel will be labelled as seed if it satisfies all 7 rules in Table 8.

Table 8. Rules applied to obtain Burned Seeds

Rule number	Criteria
1	$MIRBI_t > 5^{th_Percentile_IBC}(MIRBI_t)$
2	$(MIRBI_t - MIRBI_{t-1}) > 5^{th_Percentile_IBC}(MIRBI_t - MIRBI_{t-1})$
3	$NBR2_t < 95^{th_Percentile_IBC}(NBR2_t)$
4	$(NBR2_t - NBR2_{t-1}) < 95^{th_Percentile_IBC}(NBR2_t - NBR2_{t-1})$
5	$NIR2_t < 95^{th_Percentile_IBC}(NIR_t)$
6	$NIR_t - NIR_{t-1} < 95^{th_Percentile_IBC}(NIR_t - NIR_{t-1})$
7	Not masked

4.5.7 Second Stage Probability of Burn (SEPB)

For the refinement process two behaviours were considered. To choose one over the other option, the separability between the IBC regions and the IBNC classes were computed for the three multitemporal variables ($MIRBI_t - MIRBI_{t-i}$, $NBR2_t - NBR2_{t-i}$, $NIR_t - NIR_{t-i}$) using the M separability index:

$$M = \frac{abs(\mu_{IBC} - \mu_{IBNC})}{\sigma_{IBC} + \sigma_{IBNC}}$$

Where

μ_{IBC} = Mean of the variable for the Initially Burned Confirmed category

σ_{IBC} = Standard deviation for the Initially Burned Confirmed category

μ_{IBNC} = Mean of the variable for the Initially Burned Not Confirmed category

σ_{IBNC} = Standard deviation for the Initially Burned Not Confirmed category

The aim behind this separability choice was to assess if the regions confirmed as burned by hotspots differ spectrally from those that have not been confirmed. A threshold of 1.0 is usually used to decide whether two categories are separable, but here a 0.75 value was chosen to assure that IBC and IBNC pixels are spectrally similar. Two cases are considered:

- If one of the following conditions is satisfied: $M_{MIRBI_t - MIRBI_{t-i}} > 0.75$ OR $M_{NBR2_t - NBR2_{t-i}} > 0.75$ OR $M_{NIR_t - NIR_{t-i}} > 0.75$
- If the three conditions are satisfied (Not a) case): $M_{MIRBI_t - MIRBI_{t-1}} < 0.75$ AND $M_{NBR2_t - NBR2_{t-1}} < 0.75$ AND $M_{NIR_t - NIR_{t-1}} < 0.75$

The refinement process for both cases is based on MIRBI multitemporal difference and NBR2 multitemporal difference variables. The main idea is to establish a threshold value dependent on the ‘not burned’ background and already detected burned areas, using the 90th percentile of not burned background pixels as the minimum value and the 50th percentile of burned pixels as the maximum value of a membership function.

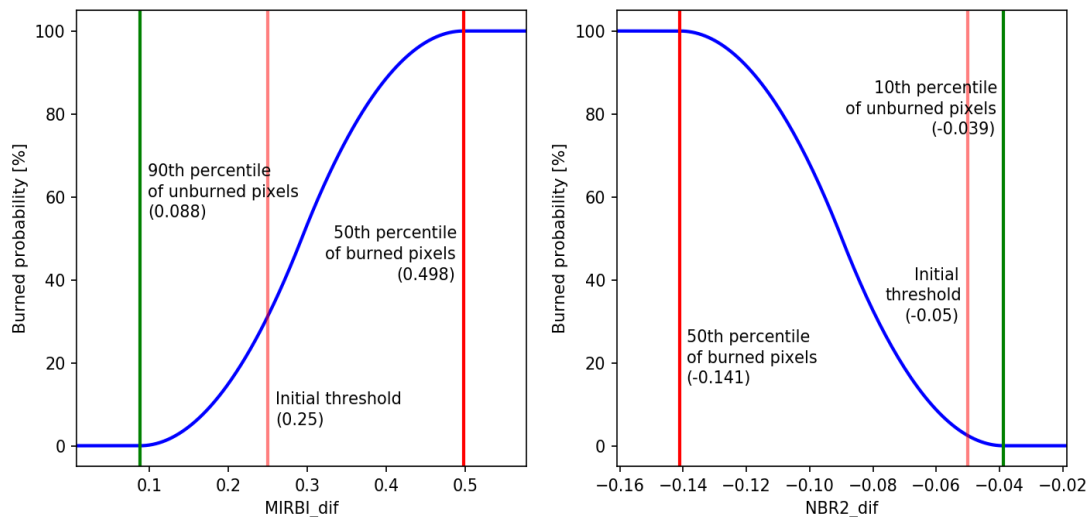


Figure 12. s-shape and z-shape sigmoid membership functions applied to MIRBI and NBR2 temporal differences, in tile 28PET between December 22 2015 and January 11 2016.


A sigmoid curve (s-shaped for MIRBI and z-shaped for NBR2) (Figure 12) is used to define the burned membership function so that the transition between both categories is more continuous. The minimum and maximum values for this function are extracted from not burned background and burned areas, depending on the case and following the next equations:

- IBC and IBNC pixels differ significantly, so IBNC is used as not burned background for the statistics (along with INB), and only IBC as burned.

$$MIRBI_{burned_membership} = scurve(90th_Percentile_IBNC_INB(MIRBI_t - MIRBI_{t-i}), 50th_Percentile_IBC(MIRBI_t - MIRBI_{t-i}))$$

$$NBR2_{burned_membership} = zcurve(10th_Percentile_IBNC_INB(NBR2_t - NBR2_{t-i}), 50th_Percentile_IBC(NBR2_t - NBR2_{t-i}))$$

where:

	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0				
	Algorithm	Theoretical	Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2			Page	25		

- *scurve*: is the sigmoid s-shape membership function
- *zcurve*: is the sigmoid z-shape membership function
- *90th_Percentile_IBNC_INB*: 90th percentile of those areas labelled as Initially Not Burned or Initially Burned Not Confirmed
- *10th_Percentile_IBNC_INB*: 10th percentile of those areas labelled as Initially Not Burned or Initially Burned Not Confirmed
- *50th_Percentile_IBC*: 50th percentile of those areas labelled as Burned on the Initially Burned Confirmed stage

b) IBC and IBNC pixels have similar values, so only INB is used as not burned background, and both IBC and IBNC as burned.

$$MIRBI_{burned_membership} = scurve(90th_Percentile_INB(MIRBI_t - MIRBI_{t-i}), 50th_Percentile_IBC_IBNC(MIRBI_t - MIRBI_{t-i}))$$

$$NBR2_{burned_membership} = zcurve(10th_Percentile_INB(NBR2_t - NBR2_{t-i}), 50th_Percentile_IBC_IBNC(NBR2_t - NBR2_{t-i}))$$

where:

- *scurve*: is the sigmoid s-shape membership function
- *zcurve*: is the sigmoid z-shape membership function
- *90th_Percentile_INB*: 90th percentile of those areas labelled as not burned on the initialization stage
- *10th_Percentile_INB*: 10th percentile of those areas labelled as not burned on the initialization stage
- *50th_Percentile_IBC_IBNC*: 50th percentile of those areas labelled as burned on the initialization stage

These membership functions are used as burned probability functions based on MIRBI and NBR2 temporal differences. The multiplication of both probability functions, called Second Stage Probability of Burn (SEPB), is used as the probability of burn for the final result. This lowers the probability values, but it is necessary because both spectral indices are independent variables. This is also why the recommended threshold of the final product between unburned and burned classes is so low.

$$SEPB = MIRBI_{burned_membership} \cdot NBR2_{burned_membership}$$

4.5.8 Probability of burn results

The burned seeds were applied in the Second Stage Probability of Burn image for the final result. Seeds were typically used in images with discrete values (rather binary values), as only patches containing a seed were selected removing areas with no seed. However, as the SEPB image is a continuous variable, a different approach is used in this case.

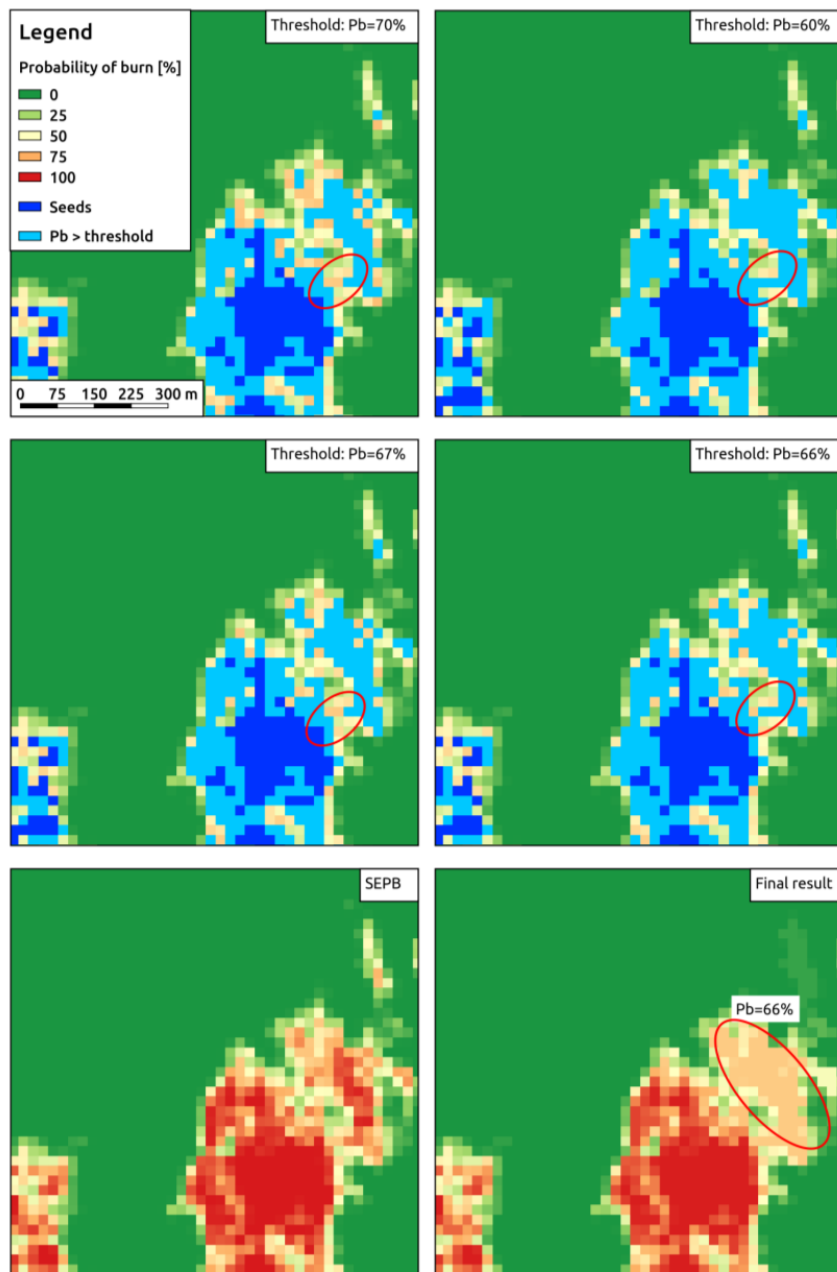


Figure 13. Extension of seeds’ influence area in the SEP algorithm for a sample zone of the 30PWQ study area, between 27th December 2015 and 6th January 2016; the legend is the same for all the maps.

The probability of a pixel being burned (P_b) of the final result for each pixel is the highest threshold that can be used in the SEP algorithm to classify burned and unburned areas, so that the pixel in question is part of a burned area containing a seed. That is to say, if a higher threshold than the P_b of a pixel is applied, the pixel will be part of a patch (using the connectivity of 2, considering 8 neighbour pixels) that does not contain any seed and classified as not burned; otherwise it will be connected to a burned area containing a seed, and it will be classified as a burned pixel. In Figure 13, results of using several thresholds are shown. The P_b value that connects the seed to the pixel group in the northeast of the red ellipse is found by decreasing the threshold levels; this threshold value is then applied to that group of pixels ($P_b=66\%$).

4.5.9 Resampling of the Probability of burn

The resulting layer of the algorithm has continuous values from 0 to 100 indicating the probability of burn of the pixel, but the user of the product may prefer a binary layer with two categories for unburned and burned areas. A threshold of 50% is typically used in these cases, but a lower value should be used in this product due to the following reasons:

- When generating sigmoid curves in the refinement process (section 4.5.7) the percentiles for burned and unburned categories are not symmetric, being the sigmoid curve closer to the burned class than to the unburned one, and thus causing the boundary between both classes to be in a lower probability value.
- The probability value of the SEPB is obtained by multiplying MIRBI and NBR2 probability functions, lowering even more the threshold value.

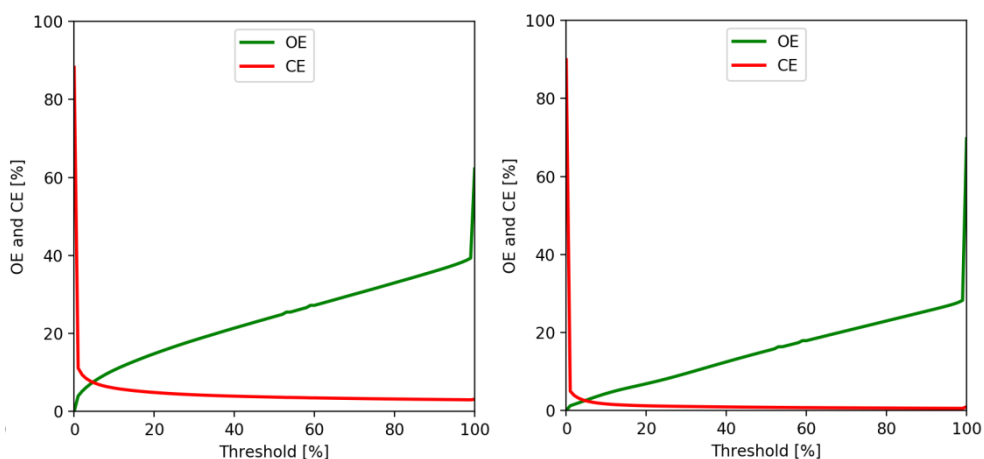


Figure 14. Evolution of omission and commission errors depending on the threshold in two sample tiles (31PEN to the left and 34PET to the right).

In order to decide which the best threshold was, the evolution of both omissions and commissions was analysed depending on the applied threshold. As it is shown in Figure 14, this threshold is around the 5% value. Since this value could be confusing for the product's users, the probability values were rescaled so that this original 5% value would become the common 50% threshold, following Table 9. Due to the scale of the values being rescaled, the new probability values have discrete values in intervals of 10%. As pixels with a probability of burn below 50% are considered as not burned, only pixels above this threshold are represented in the final SFD product.

Table 9. Lookup Table for the rescaling process.

Original probability	Rescaled probability (discrete values)
% 0	% 0
% 1	% 10
% 2	% 20
% 3	% 30
% 4	% 40
% 5-14	% 50
% 14-23	% 60
% 23-32	% 70
% 32-41	% 80
% 41-50	% 90
> =%50	% 100

5 Results of Test over the study areas

Quality assessment of the S2 algorithm was based on the sample scenes described in section 4.3. Table 10 shows the results of the algorithm in the 52 study areas and the aggregated results for the whole dataset. Omissions are much higher than commissions (24.5% and 8.1% respectively, for the whole dataset). Commissions are mainly related to cloud borders (clouds not classified by sen2cor and so not masked) and to a lesser detection of burned areas, especially around active fires where BA are overestimated; highest CE correspond to study areas where there were only small burned areas but many agricultural fields with similar spectral signal to burned areas. On the other hand, omissions are caused by either too strict thresholds or by the absence of hotspots (no BA is classified when no active fire can be found in the tile between pre- and post-fire dates).

Table 10. Quality assessment of the S2 algorithm in the 52 study areas.

Tile	Masked area [km ²]	Reference validation data				Algorithm results				OE [%]	CE [%]	Kappa
		BA [km ²]	Number fires < 25ha	Number fires 25 – 100ha	Number fires > 100ha	BA [km ²]	Number fires < 25ha	Number fires 25 – 100ha	Number fires > 100ha			
28PEB	6.0	164.0	75	6	16	247.1	15208	11	11	9.6	40.0	0.717
28PET	1715.6	1205.2	2292	322	223	787.5	27333	324	150	36.9	3.5	0.739
29PNS	1403.6	23.2	41	11	6	44.4	2156	12	9	6.1	50.9	0.643
29PNN	643.1	1011.1	11943	536	170	867.7	35640	438	127	22.5	9.7	0.819
29NNJ	1041.9	1138.7	9916	219	105	1571.2	28792	276	107	6.4	32.1	0.758
30PWB	876.6	22.2	25	1	2	20.1	128	0	2	12.7	3.7	0.916
30PWT	87.7	1093.6	11753	541	179	993.2	20918	513	143	11.2	2.2	0.925
30NWP	3689.3	536.5	2569	116	76	479.3	8083	124	73	14.8	4.7	0.893
31PES	3455.9	1.1	12	1	0	0.0	0	0	0	100.0	0.0	0.000
31PEN	5119.1	818.9	4860	333	105	795.8	11679	321	100	8.1	5.5	0.923
31NEJ	448.4	908.0	25027	478	60	525.7	38753	230	10	50.0	13.7	0.610
32PNT	206.7	292.7	8018	74	46	228.7	8055	62	44	23.3	1.9	0.858




fire
cci

Fire_cci
Algorithm Theoretical Basis
Document – Small Fires Dataset: S2

Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
Issue	1.0	Date	01/10/2018
Page	29		

Tile	Masked area [km ²]	Reference validation data				Algorithm results				OE [%]	CE [%]	Kappa
		BA [km ²]	Number fires < 25ha	Number fires 25 – 100ha	Number fires > 100ha	BA [km ²]	Number fires < 25ha	Number fires 25 – 100ha	Number fires > 100ha			
32NNP	275.3	3598.0	40904	937	281	3428.4	102607	1083	252	13.2	8.9	0.841
33PWN	660.3	847.7	1947	116	71	753.7	6003	127	85	11.9	0.9	0.928
33NWJ	2558.0	2044.3	12641	868	348	1585.5	43273	823	240	24.2	2.3	0.820
33NWE	5157.7	858.1	7205	540	143	811.0	16302	500	134	8.3	3.0	0.935
34PEB	1865.4	0.4	7	0	0	0.0	0	0	0	100.0	0.0	0.000
34PET	668.8	1149.8	209	67	104	1113.7	4859	79	108	6.4	3.3	0.946
34NEP	1502.2	4918.9	5029	318	158	4907.3	28907	324	178	6.6	6.4	0.878
34NEK	8346.9	405.8	6105	178	69	359.5	12451	154	60	15.1	4.2	0.889
35PNN	253.7	76.1	48	4	1	78.4	166	4	2	2.0	5.0	0.965
35NNJ	2923.3	996.1	2994	262	144	968.3	9336	255	133	4.4	1.7	0.966
35NNE	5481.5	2016.2	15175	694	356	1774.7	31961	605	292	15.9	4.4	0.852
36PWT	2119.7	135.9	336	28	17	130.9	1244	26	17	5.3	1.6	0.965
36NWP	6858.3	2020.8	1099	91	73	1575.1	23461	114	64	23.3	1.6	0.791
36NWK	2349.3	4098.8	11827	402	183	2747.0	85647	518	224	38.1	7.6	0.609
37NEJ	7153.4	3.2	100	3	0	40.8	5493	19	2	1.7	92.3	0.141
33MWR	5056.6	445.0	11323	260	35	552.2	49058	258	39	25.5	40.0	0.639
33MWM	2678.9	1957.7	19686	1248	325	1397.2	98428	823	141	34.7	8.6	0.711
33LWH	730.7	1421.7	10140	504	147	1088.1	52775	401	121	31.6	10.7	0.747
33LWD	1833.7	1665.9	2499	294	199	1734.4	13938	319	204	5.4	9.1	0.912
34MEA	9135.2	415.4	4692	262	70	334.1	7088	214	52	20.3	0.9	0.867
34MES	2123.7	3977.0	9518	979	478	2208.6	122971	925	283	48.6	7.4	0.526
34LEN	1656.3	2784.4	2406	230	224	1401.8	41075	324	197	49.9	0.5	0.593
34LEJ	140.1	526.0	1587	208	115	514.7	9462	207	108	10.6	8.6	0.899
34KEE	78.9	535.0	24	6	13	502.9	300	0	8	6.4	0.4	0.964
34HEJ	3830.4	99.3	472	3	4	74.2	12276	15	6	57.3	42.8	0.484
35MNR	9767.4	116.0	1389	104	18	73.7	4678	50	3	44.4	12.5	0.667
35MNM	496.7	1597.1	6485	569	264	1169.5	67354	441	180	33.1	8.6	0.743
35LNH	661.1	1887.6	4918	410	305	1718.1	76761	455	285	21.2	13.4	0.793
35LND	654.2	1261.2	6418	610	265	1200.9	50370	561	214	16.3	12.1	0.840
35KNV	32.7	134.9	406	14	17	128.3	1330	15	15	7.7	2.9	0.946
35JNM	2738.4	65.8	1535	32	7	82.2	5653	46	6	17.7	34.2	0.729
36MWA	112.1	6.5	409	3	0	2.0	228	1	0	74.0	14.5	0.399
36MWS	327.2	1638.2	5095	384	209	1388.1	30472	399	193	18.5	3.8	0.865
36LWN	3087.2	375.8	6656	210	55	383.3	23225	193	46	13.9	15.6	0.846
36LWJ	235.6	1103.5	6500	480	202	1106.9	16415	492	197	4.0	4.3	0.954
36KWE	2777.9	1220.0	9841	314	150	52.1	4211	28	8	96.1	9.7	0.064
36KWA	1906.8	20.6	388	15	2	16.9	810	9	2	25.1	9.0	0.821
37MEM	8815.7	87.4	1529	70	5	77.5	7180	48	3	28.4	19.2	0.753
37LEH	3525.8	875.0	4023	329	146	757.8	50918	290	113	22.4	10.4	0.814
37LED	1465.8	569.6	5083	205	102	509.5	16220	200	85	14.5	4.4	0.897
Aggregated results	130737.0	55171.8	305179	14889	6293	45310.3	1331651	13656	5076	24.5	8.1	0.809

	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
	Algorithm Theoretical Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2		Page	30	

6 Limitations and Next Steps

The limitations found in the S2 processing and algorithm development are:


1. Dependence on MODIS hotspots: The algorithm presented in this document depends totally on hotspots from the MCD14DL product, generated by detection of thermal anomalies in MODIS imagery. The lack of hotspots in a tile between pre- and post-fire dates or no hotspot being closer than 500m to IB (Initially Burned) areas means that no burned area is detected in the outcome of the algorithm. This leads to a 100% omission error even if there were some small areas burned in this period. Since the main reason not to use VIIRS hotspots was that they were not available at the time of the development of this algorithm, one possible solution is to replace MODIS hotspots with VIIRS, as they are more accurate. If even then there were no hotspot (or very few of them) in the S2 tile, statistics from adjacent tiles could be used; only the adjacent 8 tiles would be used, as going farther could change the biome of these tiles and thus the spectral signal of burned areas.
2. Confidence information: The result of the outcome indicates the probability of burn for every pixel with a value between 0% and 100% (not burned and burned, respectively), being this probability function extracted from IBC (Initially Burned Confirmed, IB areas crossed with hotspots) areas. However, the initialization of the algorithm is based in thresholds derived from initial spectral analyses, so the probability values are not totally dependent on image characteristics.
3. Missing data: There was a large gap of S2 images when developing this algorithm, as only images acquired later than December 2015 were available even if the Sentinel-2A satellite was operating since August 2015. This made it difficult to find study areas of different biomes that contained BA in the available dates, so tiles from temperate grassland and savannah and from the Mediterranean forest contain very few and small burned areas.

The next steps to perform for a future version are:

1. Implement confidence and uncertainty information based on image's properties, and remove thresholds.
2. Decrease the dependence of the algorithm on MODIS hotspots, either using VIIRS information, extracting statistics from the temporal series or extending their influence area, or a combination of those.
3. Prolong the temporal series.

7 References

- Bastarrika, A., Alvarado, M., Artano, K., Martinez, M., Mesanza, A., Torre, L., Ramo, R., & Chuvieco, E. (2014). BAMS: A Tool for Supervised Burned Area Mapping Using Landsat Data. *Remote Sensing*, 6, 12360-12380.
- Bastarrika, A., Chuvieco, E., & Martin, M.P. (2011a). Automatic Burned Land Mapping From MODIS Time Series Images: Assessment in Mediterranean Ecosystems. *IEEE Transactions on Geoscience and Remote Sensing*, 49, 3401-3413.
- Bastarrika, A., Chuvieco, E., & Martín, M.P. (2011b). Mapping burned areas from Landsat TM/ETM+ data with a two-phase algorithm: balancing omission and commission errors. *Remote Sensing of Environment*, 115, 1003-1012.

	Fire_cci	Ref.:	Fire_cci_D2.1.2_ATBD_SFD_S2_v1.0		
	Algorithm Theoretical Basis	Issue	1.0	Date	01/10/2018
	Document – Small Fires Dataset: S2	Page	31		

Boschetti, L., Roy, D.P., Justice, C.O., & Humber, M.L. (2015). MODIS–Landsat fusion for large area 30m burned area mapping. *Remote Sensing of Environment*, 161, 27-42.

Chuvienco, E., Martín, M.P., & Palacios, A. (2002). Assessment of different spectral indices in the red-near-infrared spectral domain for burned land discrimination. *International Journal of Remote Sensing*, 23, 5103-5110.

Chuvienco, E., Yue, C., Heil, A., Mouillot, F., Alonso-Canas, I., Padilla, M., Pereira, J.M., Oom, D., & Tansey, K. (2016). A new global burned area product for climate assessment of fire impacts. *Global Ecology and Biogeography*, 25, 619-629.

Chuvienco, E., M.L. Pettinari, A. Heil and T. Storm (2017) ESA CCI ECV Fire Disturbance: Product Specification Document, version 6.3. Available at: <http://www.esa-fire-cci.org/documents>

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., & Martimort, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25-36.

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., Randerson, J.T., & Werf, G.R. (2013). Analysis of daily, monthly, and annual burned area using the fourth generation global fire emissions database (GFED4). *Journal of Geophysical Research: Biogeosciences*, 118, 317-328.

GCOS (2016). *The Global Observing System for Climate: Implementation Needs*. Geneva, Switzerland: GCOS-200. World Meteorological Organization.

Hantson, S., Padilla, M., Corti, D., & Chuvienco, E. (2013). Strengths and weaknesses of MODIS hotspots to characterize global fire occurrence. *Remote Sensing of Environment*, 131, 152-159.

Padilla, M., Stehman, S.V., Hantson, S., Oliva, P., Alonso-Canas, I., Bradley, A., Tansey, K., Mota, B., Pereira, J.M., & Chuvienco, 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.

Randerson, J., Chen, Y., Werf, G., Rogers, B., & Morton, D. (2012). Global burned area and biomass burning emissions from small fires. *Journal of Geophysical Research: Biogeosciences* (2005–2012), 117 - G04012, 1-23.

Roy, D., & Landmann, T. (2005). Characterizing the surface heterogeneity of fire effects using multi-temporal reflective wavelength data. *International Journal of Remote Sensing*, 26, 4197-4218.

Zhu, Z., & Woodcock, C.E. (2014). Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. *Remote Sensing of Environment*, 152, 217-234.

Annex: Acronyms and abbreviations

AD	Applicable Document
ATBD	Algorithm Theoretical Basis Document
BA	Burned Area
BAMS	Burned Area Mapping Software
BOA	Bottom of Atmosphere
CC	Cloud Cover
CCI	Climate Change Initiative
ESA	European Space Agency
ECV	Essential Climate Variables
GCOS	Global Climate Observing System
GHG	Green House Gases
IBC	Initially burned confirmed
IB	Initially burned
INB	Initially not burned
INBC	Initially not burned confirmed
IPCC	Intergovernmental Panel on Climate Change
LDCM	Landsat Data Continuity Mission (LDCM)
LSC	Land Scene Clasification
MERIS	MEDium Resolution Imaging Spectrometer
MIRBI	Mid-Infrared Burned Index
MODIS	Moderate Resolution Imaging Spectroradiometer
MSI	Multi Spectral Instrument
NASA	National Aeronautics and Space Administration
NBR-2	Normalized Burned Ratio 2
NIR	Near InfraRed
OLI	Operational Land Imager
S2	Sentinel-2
SCL	Scene CLasification
SEPB	Second Stage Probability of Burn
SFD	Small Fire Dataset
SWIR	ShortWave Infra Red
TIRS	Thermal InfraRed Sensor
TOA	Top of Atmosphere
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
VIIRS	Visible Infrared Imaging Radiometer Suite