




fire
cci

**ESA Climate Change Initiative – Fire_cci
O3.D1 Algorithm Theoretical Basis Document (ATBD) –
large demonstrator area (LDA) in South America**

Project Name	ECV Fire Disturbance: Fire_cci Phase 2
Contract Nº	4000115006/15/I-NB
Issue Date	31/08/2018
Version	2.0
Author	Mihai A. Tanase, Miguel Ángel Belenguer-Plomer
Document Ref.	Fire_cci_O3.D1_ATBD-S1-SA_2.0
Document type	Public

*To be cited as: M.A. Tanase, M.A. Belenguer-Plomer (2018) ESA CCI ECV Fire Disturbance: O3.D1 Algorithm Theoretical Basis Document – S1 South America, v2.0.
Available at: <https://www.esa-fire-cci.org/documents>*

	Fire_cci	Ref.:	Fire_cci_O3.D1_ATBD-S1-SA_2.0	
	Algorithm Theoretical Basis Document	Issue	2.0	Date 31/08/2019
	LDA South America	Page	2	


Project Partners

Prime Contractor/ Scientific Lead & Project Management	UAH – University of Alcalá (Spain)
Earth Observation Team	UAH – University of Alcalá (Spain)
System Engineering	Cubenube (Spain)



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	
	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		
Vanesa Moreno (IRD)	mariavanesa.morenodominguez@cefe....		
Philippe Ciaï (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 Algorithm Theoretical Basis Document LDA South America	Ref.:	Fire_cci_O3.D1_ATBD-S1-SA_2.0	
		Issue	2.0	Date 31/08/2019
		Page		

Summary

This document describes the algorithms used for generating the small-fire database for the large demonstrator area (LDA) in tropical South America within the Fire_cci project. It includes the description of the Sentinel-1 C-band data pre-processing, as well as the ancillary data used to derive burned area over the tropical Amazon.

	Affiliation/Function	Name	Date
Prepared	UAH	Mihai A. Tanase	30/08/2018
	UAH	Miguel Ángel Belenguer-Plomer	
Reviewed	UAH – Project Manager	M. Lucrecia Pettinari	30/08/2018
Authorized	UAH - Science Leader	Emilio Chuvieco	30/08/2018

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

Document Status Sheet

Issue	Date	Details
1.0	31/07/2017	First release of the document.
2.0	30/08/2018	Second release of the document, after a thorough update of the text according to updates in the algorithm.


Document Change Record

Issue	Date	Request	Location	Details
2.0	30/08/2018	ESA- UAH	All document	Update of the whole of the document.



Table of Contents

1	Executive Summary	6
2	Introduction.....	7
2.1	Purpose of the document.....	7
2.2	Reference Documents	7
2.3	Background	7
3	The Small Fire Database	7
4	Data	9
4.1	Sentinel-1 system and data products.....	9
4.2	Ancillary datasets.....	10
4.2.1	Landsat-8 data products	11
4.2.2	Active fire products	11
4.2.3	Other datasets	12
5	Methods.....	12
5.1	Site selection for algorithm development and preliminary validation.....	12
5.2	Generation of burned area perimeters from Landsat 8 datasets.....	13
5.3	Classification of Burned and Non-Burned Vegetation from Sentinel-1	14
5.3.1	SAR data pre-processing	14
5.3.2	Burned area classification.....	15
5.3.2.1	Stage 1: Anomaly change detection.....	16
5.3.2.2	Stage 2: Burned and unburned regions of interest.....	17
5.3.2.3	Stage 3: Adjustment for temporal decorrelation	19
5.3.2.4	Stage 4: Random forests burned/unburned classification.....	19
5.3.2.5	Stage 5: Post processing	20
5.3.2.6	Stage 6: Burned area detection without hotspots	20
6	Uncertainty estimation framework	21
7	References	22
	Annex 1: Acronyms and abbreviations	25

	Fire_cci Algorithm Theoretical Basis Document LDA South America	Ref.:	Fire_cci_O3.D1_ATBD-S1-SA_2.0	
		Issue	2.0	Date 31/08/2019
		Page		

List of Tables

Table 1. Main characteristics of the Sentinel-1 satellite sensor. 10

Table 2. Landsat 8 OLI and TIRS bands description 11

Table 3. Correspondence between Landsat 8 and Sentinel -1 data used for algorithm development and preliminary validation (see D3 and D5 for validation results)... 13

Table 4. List of Sentinel-1 SAR data being used in the classification 15

List of Figures

Figure 1: Comparison of Landsat and Sentinel 2 spectral bands (source: NASA) 11


Figure 2: Areas selected for algorithm development and preliminary validation. Terrestrial eco-regions are also shown. 13

Figure 3: Flowchart for SAR data processing with Orfeo Toolbox 15

Figure 4: Burned area detection algorithm flowchart..... 16

Figure 5: Graphical representation of concepts needed to extract bROIs, HS-hotspot, distance. 18

Figure 6: Example of burn probability for two different tiles 22

	Fire_cci Algorithm Theoretical Basis Document LDA South America	Ref.:	Fire_cci_O3.D1_ATBD-S1-SA_2.0	
		Issue	2.0	Date 31/08/2019
		Page		

1 Executive Summary

Forest fires are an important source of atmospheric aerosols and greenhouse gases, with a causal relationship between biomass burning and inter-annual variability of related emissions being observed (Simmonds et al. 2005). Worldwide, about 350 million hectares are affected by large fires (i.e., mapped from 500 m spatial resolution data) annually (Giglio et al. 2013), exerting a major influence on carbon release from terrestrial ecosystems (Andreae and Merlet 2001; Simmonds et al. 2005). Fires are also a major factor in land cover changes, and hence affect fluxes of energy and water to the atmosphere. In this context, spatial and temporal monitoring of burned areas can be inferred using remote sensing, a cost effective, objective, and time-saving method to monitor and quantify location, extent and intensity of fire events (Chuvienco 1999; Laneve et al. 2006; Stroppiana et al. 2003). The Fire Disturbance Essential Climate Variable provides baseline products for the land-surface to allow such monitoring activities at global scales.

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. However, global burned area may be underestimated by up to 35% since small fires are often missed by coarse resolution sensors (Kloster et al. 2012).

The aim of Option 3 of the Fire_cci project is to provide Sentinel-1 burned area products, at high spatial resolution, over a large demonstrator area (LDA) located in tropical South America. The Option complements the baseline project by developing a C-band backscatter change detection algorithm adapted to the fire regimes encountered in tropical South America. Option 3 also extends the areas mapped within the “Small fire database” (SDF) to tropical regions in South America. The algorithm complements those proposed within the baseline (i.e., pixel-based coherence) and CCN-1: Indonesia Fires and El Niño Special Case (object-based backscatter) by integrating pixel-based approaches with object recognition and contextual information. The algorithm detects anomalous backscatter changes within an iterative multi-temporal analysis and takes advantage of ancillary information on land cover and thermal anomalies (hotspots) to label the fire affected areas being autonomous and adapting to the local conditions.

This document is the Algorithm Theoretical Basis Document (ATBD) corresponding to the generation of the small fires database over one LDA located in the tropical South America. **The document describes the algorithm, methods, and approaches that lead to the generation of the small fire database for the Amazon LDA** within the Phase 2 of the Fire_cci project. In this document approaches based on synthetic aperture radar (SAR) data are described. Test sites for algorithm development and calibration have been selected based on representative vegetation types (i.e., grasslands, crops and forests) that burn on a regular basis. The theoretical basis described here identifies the data sets used to classify burned area and the methods used to derive the cartographic products. **Assessment of algorithm’s performance and comparisons with the remaining SFD algorithms are described in separate documents [RD-2] and [RD-4].** Burned area maps derived from Sentinel-1 imagery were created for the tropical Amazon basin for year 2017. The resulting maps were validated using BA perimeters derived from multispectral optical datasets (e.g., Sentinel-2, Landsat-8).

2 Introduction

2.1 Purpose of the document

This document supplements existing ATBDs developed for Africa and South-East Asia. The document follows and draws from the SFD ATBDs already provided within Fire_cci Phase 2. It describes the theoretical basis of the algorithm used to map burned areas from Sentinel-1 SAR data over the tropical South America.

2.2 Reference Documents

[RD-1]	O3.D2: Burned area database for the candidate validation tiles
[RD-2]	O3.D3: Intermediate validation results: SAR pre-processing and Burned Area detection.
[RD-3]	O3.D4: Product validation report
[RD-4]	O3.D5: Radar - Algorithm intercomparison document

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

Global BA products are based on coarse resolution sensors (from 300 to 1000m). Therefore, the likelihood of detecting small burns (i.e. < 50ha) is low, and the products are frequently affected by omission errors (Giglio et al. 2009; Padilla et al. 2015), particularly coming from small fires (Kloster et al. 2012). To improve the characterization of small fires, one of the objectives of the Fire_cci Phase 2 project is to generate a small fires database based on medium resolution sensors (10 to 100m). Considering the massive processing effort when generating products at global level, the SFD is focused on the African continent, the most burned worldwide (Chuvienco et al. 2016; Giglio et al. 2013), with additional areas being *a posteriori* selected over tropical South-east Asia and South America. The SFD shall add value to the global BA products generated within the Fire_cci and the global characterization of fire activity.

3 The Small Fire Database

Burned area 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 of human or natural origin, 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 estimated that accounting for small fires may increase burned area and global carbon emissions by approximately 35% (Kloster et al. 2012).

The strategy of using active fire products for BA detection and mapping may provide additional information on small fires contribution to total burned area because of the strong nonlinearity in radiative power as a function of fire temperature. Active fires products have shown very low commission errors, but very diverse omission rates (36-86%) depending on the ecosystem and fire size (Hantson et al. 2013). In this context, there is a need to link coarse resolution products with data on small burned patches, in locations where these are the dominant fire types. This could be achieved with higher resolution remote sensing data such as Landsat-OLI (30 m) or Sentinel-2 optical data (down to 10 m). However, in areas with frequent clouds such as the Amazon, the mapping capability of optical data is diminished.

Sentinel-1 SAR data may provide the means to overcome the drawbacks of the optical sensors in areas with persistent cloud cover. Over the past decade, radar sensors have become of interest for tracking fire disturbance due to their independence of cloud cover and solar illumination (Bourgeau-Chavez et al. 2002; Tanase et al. 2010). In addition, radar sensors are ideal for long-term monitoring, as the signal is directly influenced by vegetation structure and thus more sensitive to temporal changes when compared to optical data (Kasischke et al. 2011). The radar backscatter includes different scattering mechanisms (e.g. crown, trunk, and ground) depending on the land cover type. The relative importance of such scattering mechanisms depends on the structural properties of the vegetation, with crown components being the primary source of scattering at C-band ($\lambda \approx 5$ cm; $f \approx 5.4$ GHz), the wavelength of interest for Option 3.

The standard derivative from radar data is the backscatter coefficient. This provides an indication of the amount of returned energy that is scattered in a backward direction towards the sensor. The contribution of understory vegetation and ground surface is primarily conditioned by the upper canopy architecture (e.g., crown closure, shape, etc.). Over sparsely vegetated areas most of the scattering comes from the ground surface, with some attenuation (two-ways) from the vegetation layer. *A priori*, the removal of vegetative material reduces the number of scattering elements and thus the total backscatter, as elements of a size similar with the wavelength constitute important sources for microwave scattering (Rignot and Zyl 1993). However, variations in the post-fire response caused by topographic and environmental conditions are widely documented (Huang and Siegert 2006; Menges et al. 2004; Rignot et al. 1999; Tanase et al. 2010). Overall, C-band co-polarized (VV or HH polarizations) backscatter showed little sensitivity to fire impacts, particularly for slopes pointing away from the sensor, in some environments (e.g., Mediterranean). Nevertheless, using co-polarized waves, fire affected areas could be delineated in tropical environments (Huang and Siegert 2006) suggesting differentiated sensitivities across biomes. Cross-polarized backscatter (HV or VH polarizations) show lower values over burned areas at most SAR frequencies, and the sensitivity to burns was rather constant over the entire range of local incidence angles (Tanase et al. 2010).

Regional to global burned area algorithms using Sentinel-1 data have not been yet published, but considering the better spatial and temporal resolution, improved characteristics when compared to past C-band sensors (e.g., signal to noise ratio, shallower incidence angle over land, higher temporal frequency, multiple polarizations)

	Fire_cci Algorithm Theoretical Basis Document LDA South America	Ref.: Fire_cci_O3.D1_ATBD-S1-SA_2.0
		Issue 2.0 Date 31/08/2019
		Page 9

past results regarding C-band sensitivity to fire impacts remain valid and change-detection radar-based algorithms used for fire impact assessment (Fieber et al. 2015) could be repurposed for BA detection within a multi-temporal framework. However, backscatter changes may result from non-fire-related landscape changes (e.g., crop harvest, logging, pests), as well as environmental variables (e.g., changes in the background humidity or surface roughness). As with all time-series change detection methods, it is essential to link the change in backscatter coefficient to fire activity. The complex task of identifying only the fire affected areas may pose an undetermined problem when only using the two backscatter channels (polarizations) regularly acquired over land by the Sentinel-1 mission. However, contextual information data may help reduce uncertainties by taking advantage of information from the temporal (e.g., higher fire probabilities for certain months) or spatial (e.g., burned area size, shape and distance to neighbours) domains. Furthermore, high resolution active fire products (e.g., Visible Infrared Imaging Radiometer Suite -VIIRS sensor, aboard the Suomi-NPP satellite) may be linked to changed patches based on location and acquisition time and thus provide additional means when identifying areas affected by relatively small fires.

The results of the SFD will be integrated with results from the Fire_cci project using a consistent product specification, detailed in the project's Product Specification Document (PSD). The products generated through Option 3 shall follow the PSD for the small fire database. However, considering the higher resolution of Sentinel-1 data, the use of Ground Range Detected (GRD) data sets and the developed algorithm, an adaptation of the product specification was necessary:

- Target resolution: 1 hectare or better. BA smaller than this threshold may not be mapped. The minimum mapping unit area shall contain several Sentinel-1 adjacent pixels thus avoiding spurious backscatter changes. Each pixel consists of multi-looked full resolution pixels (5 in range and 4 in azimuth, at 40 m resolution) for speckle reduction. Additional speckle reduction is achieved by temporal filtering.
- Projection: Geographic projection (latitude-longitude).
- Temporal resolution: Day of detection in monthly (bi-monthly for less than nominal S-1 acquisition frequency) composites.

4 Data

4.1 Sentinel-1 system and data products

Sentinel-1 is a two-satellite constellation (A - since April 2014, B - since April 2016) with the prime objectives of Land and Ocean monitoring (Table 1). The goal of the mission is to provide C-Band SAR data continuity following the end of ERS-2 and Envisat missions. The satellites carry a C-band SAR sensor, which offers medium and high-resolution imaging in all weather conditions making it useful for land monitoring. The operational lifespan of the satellites is 7 years (with consumable for 12). However, ESA ensured the procurement of replacement satellites (Sentinel-1 C/D) to extend the operational monitoring component of Copernicus programme at least until the end of 2030. The radar operates in two main modes, with the Interferometric Wide (IW) swath (250 km width) being the default operation mode over land. The IW mode images three sub-swathes using the Terrain Observation with Progressive Scans SAR (TOPSAR) to provide high quality, homogeneous images. The advantages of Sentinel-1 sensor over other C-band SAR missions, besides the free data access policy, are three-fold, i) high temporal frequency (6 days exact repeat cycle with two satellites), ii) high spatial resolution (5 m in azimuth

and 20 m in range) and, iii) dual-polarization (VV and VH). One should notice that nominal temporal frequency is not yet achieved over areas outside Europe and North America and that areas with frequent seismic activity (e.g., the Andes) are imaged in single polarization mode (VV) for increased spatial resolution.

Sentinel-1 products are released in two Level 1 formats, Ground Range Detected (GRD) and Single Look Complex (SLC). GRD products are projected, intensity images, radiometrically and terrain corrected. SLC data are designed for interferometric applications, containing both phase and intensity information. The most commonly available SLC and GRD data are acquired in IW mode.

As for the SFD Indonesia, this Option uses the systematically distributed Level-1 Ground Range Detected data, focused data that has been detected, multi-looked and projected to ground range using an Earth ellipsoid model (typical product size is 1GB for dual-pol IW mode). Pixel values represent detected magnitude (i.e., no phase information). The ellipsoid projection of the GRD products is corrected using the terrain height (as specified in the product general annotation) which varies in azimuth but is constant in range. The products are delivered as GeoTiffs together with ancillary files containing information about orbit, noise and calibration. Notice that for SFD Africa, Single Look Complex (SLC) images (contain phase information) are used as the Sentinel-1 algorithm in Africa is based on temporal changes of the interferometric coherence. The complex SLC data processing, drastically increased data storage (eight times more when compared to GRD) and computing requirements, and C-band high temporal decorrelation over vegetated areas were considered when selecting the GRD datasets as the main product for the development of the BA algorithm in South America.

Table 1. Main characteristics of the Sentinel-1 satellite sensor.

Mission and instrument	Characteristics
Sentinel-1 (A/B)	<p>Orbit: near-polar, sun-synchronous, 180° orbit phasing between A and B</p> <p>Repeat-pass: 12 days per satellite. Combined A/B satellite passes: 6 days. Combined A/B satellites Ascending/Descending passes: 3 days.</p> <p>Instrument: C-band (5.3 GHz) synthetic aperture radar</p> <p>Acquisition modes: Strip map (SM) – on demand Extra Wide Swath (EW) – basic operation mode over sea/ocean Interferometric Wide Swath (IW) – basic operation mode over land</p> <p>Spatial resolution (IW): range: 5 m / azimuth: 20 m</p> <p>Swath width (IW): 250 km</p>

4.2 Ancillary datasets

The main ancillary datasets used for BA algorithm development and calibration are Landsat 8 imagery (used to derive reference burned perimeters), hotspots detected by the Visible Infrared Imaging Radiometer Suite (VIIRS) and Moderate Resolution Imaging Spectroradiometer (MODIS) sensors. In addition, ESA CCI Land cover dataset (300 m spatial resolution) is used to provide information on land cover types and non-burnable areas (e.g., water bodies).

4.2.1 Landsat-8 data products

Due to the limited availability of Sentinel-2 imagery over the period considered for algorithm development (2015-2016 fire seasons), Landsat-8 images were used to generate reference burned areas maps over the sites selected for the calibration and preliminary validation of the algorithm. Landsat-8 was launched on February 2013 as a combined effort of NASA and the United States Geological Survey (USGS). Landsat-8 ensures the continued acquisition and availability of Landsat data. It carries two observation sensors, the Operational Land Imager (OLI) and the Thermal InfraRed Sensor (TIRS) collecting data in nine shortwave bands and two longwave thermal bands (Table 2, Figure 1). Landsat-8 Pre-Collection¹ products were used when generating the reference burned area maps for algorithm development and calibration sites.

Table 2. Landsat 8 OLI and TIRS bands description

Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 - Coastal aerosol	0.43 - 0.45	30
Band 2 – Blue	0.45 - 0.51	30
Band 3 – Green	0.53 - 0.59	30
Band 4 – Red	0.64 - 0.67	30
Band 5 - Near Infrared (NIR)	0.85 - 0.88	30
Band 6 - SWIR 1	1.57 - 1.65	30
Band 7 - SWIR 2	2.11 - 2.29	30
Band 8 - Panchromatic	0.50 - 0.68	15
Band 9 – Cirrus	1.36 - 1.38	30
Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100 * (30)
Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100 * (30)

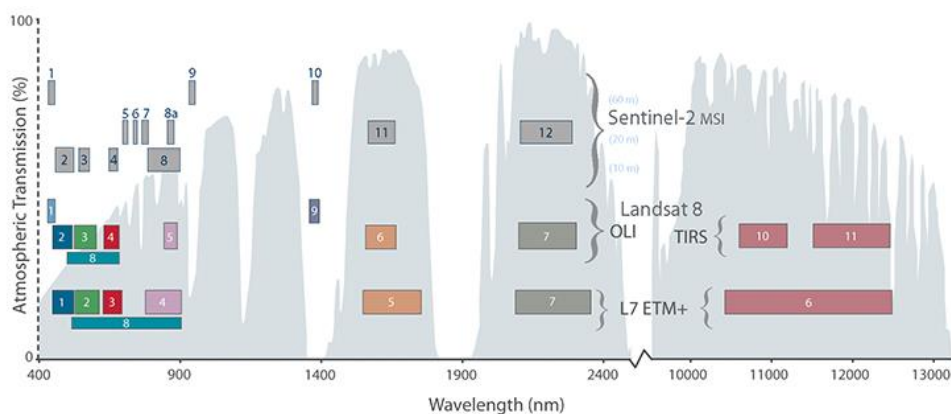



Figure 1: Comparison of Landsat and Sentinel 2 spectral bands (source: NASA)

4.2.2 Active fire products

Option 3 Sentinel-1 based BA detection algorithm uses active fire products to attribute changes in backscatter to fire affected areas. The hotspots detected by the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor, onboard the Suomi-NPP satellite (375 m product) seem the most adequate over the areas selected for algorithm development due

¹ In 2016, the USGS began making changes to manage the Landsat archive as a tiered Collection of Landsat data. Since October 2017 Landsat Pre-Collection data are no longer available.

	Fire_cci Algorithm Theoretical Basis Document LDA South America	Ref.:	Fire_cci_O3.D1_ATBD-S1-SA_2.0	
		Issue	2.0	Date 31/08/2019
		Page	12	

to its better spatial resolution and a greater response over fires of relatively small areas. Although the MODIS acquired thermal anomalies were of significant less quality over the selected areas they are used to complement the VIIRS ones. Therefore, vector files of VIIRS and MODIS hotspots downloadable from the Archive Download Tool (<https://firms.modaps.eosdis.nasa.gov/download/>) are used during various stages of the algorithm. One should notice that VIIRS hotspots are affected by errors which are currently being solved by the FIRMS team.

4.2.3 Other datasets

The ESA CCI Land cover dataset (300 m spatial resolution) is used by the algorithm to provide information on non-burnable areas (e.g. water bodies) and drive the BA detection and mapping algorithm as a function of land cover type. To account for the different spatial resolution, this layer is resampled to the spatial resolution used within Option 3 (i.e. 40m). The LC_cci land cover product is an annual series of global land cover maps at 300 m spatial resolution, covering the 1992-2015 period. The maps were produced using a multi-year and multi-sensor strategy to make use of all suitable data and maximize product consistency. The most recent ESA land cover map (i.e., 2015) is used within Option 3.

The Shuttle Radar Topography Mission DEM (SRTM DEM) provides the reference for SAR data geocoding. The NASA SRTM provides digital elevation information over 80% of the globe. The data is distributed by USGS and is available for download from the National Map Seamless Data Distribution System or the USGS ftp site. The SRTM DEM was derived from single pass SAR interferometric data acquired in 2000. For Option 3, the enhanced DEM is used. The enhanced DEM was released worldwide in 2016 at 1 arc-seconds resolution (30 m at equator) in 1-degree tiles.

5 Methods

5.1 Site selection for algorithm development and preliminary validation

Five areas (defined by Landsat path/row number) were selected for algorithm development and preliminary validation² considering: i) the fire activity (from active fire products, i.e., hotspots), ii) the main types of land cover in tropical Amazon (i.e., grasslands, cropping areas, and forests), and iii) the availability of concurrent cloud-free Landsat 8 and Sentinel-1 datasets. The selected Landsat path/rows were imaged by the Sentinel-1 sensors within the relative orbital paths 10, 54, 77, 149, and 170 (Table 3). As explained in the following sections, the selected Sentinel-1 pre-processing chain uses the 100 km Military Reference Grid System (MRGS) for data tiling. Therefore, only the 10 MGRS tiles corresponding to the selected Landsat path/rows were processed (Figure 2 and Table 3). Once the areas for algorithm development and preliminary validation were chosen, burned area perimeters were generated for the 2015-2016 fire seasons, Sentinel-1 data were processed, and the BA algorithm was developed, calibrated, and a preliminary validation³ was carried out.

² the validation of the BA product over the Amazon LDA was carried out independently as described in *O3.D2: Burned area database for the candidate validation tiles* and *O3.D4: Product validation report*.

³ the results of the preliminary validation are described in *O3.D3: Intermediate validation results: SAR pre-processing and Burned Area detection*. Notice that the preliminary validation was carried out within the MGRS tiles not used during the algorithm development and calibration).

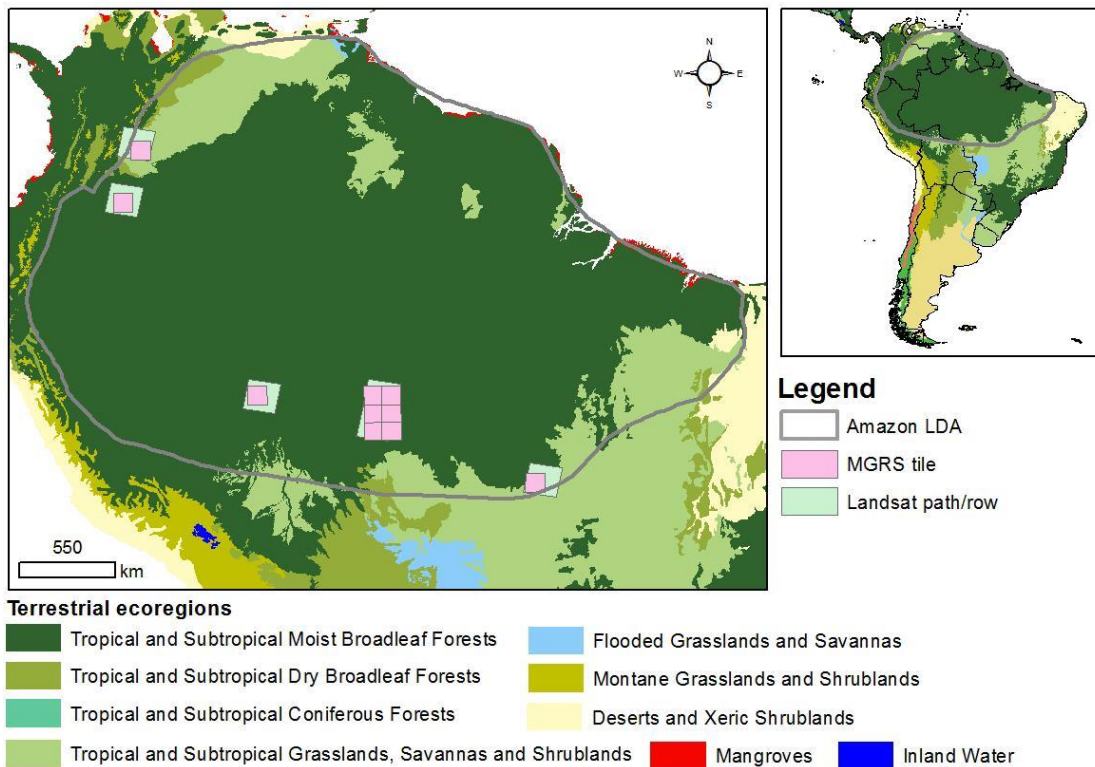


Figure 2: Areas selected for algorithm development and preliminary validation. Terrestrial ecoregions are also shown.


Table 3. Correspondence between Landsat 8 and Sentinel -1 data used for algorithm development and preliminary validation (see D3 and D5 for validation results)

Landsat-8				Sentinel-1		
Path/Row	Period	Image pairs*	Mean time-range (days)	Relative orbit	MGRS tile	Period
001/066	2016.06.30-2016.08.18	1	32	54	19LGL**	2016.05.01-2016.10.15
007/057	2015.09.26-2016.02.01	1	128	77	18NYK***	2014.10.22-2016.12.16
007/059	2016.10.30-2017.02.19	1	111	149	18NXG**	2016.09.10-2017.03.30
224/069	2016.05.23-2016.08.11	5	29	170	22LCL***	2016.04.01-2016.09.30
230/066	2016.05.01-2016.09.06	5	26	10	20LQR, 20LRR, 20LQQ, 20LRQ***	2016.04.01-2016.10.30
230/067	2016.07.20-2016.09.22	3	21	10	20LQQ, 20LRQ, 20LQP, 20LRP**	2016.04.01-2016.10.30

*cloud free image pairs used to derive multi-temporal reference burned areas; ** tile used for independent validation; *** tile used for algorithm calibration.

5.2 Generation of burned area perimeters from Landsat 8 datasets

Reference burned area perimeters were generated using the Burned Area Mapping Software, BAMS. BAMS uses several spectral indexes, commonly used in burned area detection, through a two-phase supervised strategy to map areas burned between two Landsat multitemporal images (Bastarrika et al., 2014). The algorithm needs user interaction in form of visual delimitation of burned areas, from which statistics are extracted. After the discrimination of burned patches the user can visually assess the

	Fire_cci Algorithm Theoretical Basis Document LDA South America	Ref.:	Fire_cci_O3.D1_ATBD-S1-SA_2.0	
		Issue	2.0	Date 31/08/2019
		Page		

results and iteratively select additional sampling areas to improve classification. The output is a vector layer with polygons being classified as:

- **Burned** (class 1): areas detected as burned
- **Non-observed** (class 2): areas commonly covered by clouds or by missing sensor data. If there are clouds in one of the two images, this area is masked. The mask is generated using the quality information provided with the Landsat data.
- **Unburned** (class 3): areas not detected as burned between the two dates; burned areas that were already burned in the earlier images are also classified in this category.

BAMS results were visually assessed, and missed burned area perimeters were added when changes of reflectance values matched VIIRS hotspots. In each MGRS tile, multi-temporal BA perimeters were created. Their number depended on the available Landsat images pairs with a cloud cover below 30% (Table 3). In total, 16 vector files, corresponding to 16 Landsat image pairs (pre-fire and post-fire images) were generated.

5.3 Classification of Burned and Non-Burned Vegetation from Sentinel-1

The algorithm is based on the use of temporal time-series of Sentinel-1 SAR data to identify changes in C-band backscatter and associate them with biomass burning events. The backscatter coefficient gives an indication of the amount of energy that is returned from the surface. The algorithm considers multi-temporal changes of incoherent SAR-based metrics (i.e. backscattering coefficient intensities), for each SAR channel (i.e., VV and VH) as well as changes of incoherent radar indices (RI) computed from the available radar channels. To derive the BA maps the algorithm is divided in three steps: SAR data pre-processing, detection of backscatter changes, and BA classification.

5.3.1 SAR data pre-processing

The outputs of two separate SAR chains were evaluated for Sentinel-1 data pre-processing over the LDA Amazon. The first chain is based on commercial software (i.e., Gamma Remote Sensing, GRS) while the second chain uses open-source libraries (Inglada and Christophe 2009) available through Orfeo ToolBox⁴, (OTB). The comparison was necessary as the Burned Area (BA) algorithm was developed using S-1 images pre-processed through the GRS chain, a chain available at the start of Option 3. The comparison allowed for validating BA algorithm's transferability from the development environment (GRS-based) to a cloud computing environment (OTB-based).

Following the evaluation process, OTB outputs were deemed compatible with GRS outputs (i.e., no modification was needed for the BA algorithm) and the OTB pre-processing chain was selected for implementation into a cloud computing environment as it provides for a platform-independent pre-processing solution. The agreement between the two pre-processing chains was evaluated regarding product geometric and radiometric accuracies. In addition, the BA algorithm was applied to Sentinel-1 data processed through both pre-processing chains and the results were compared⁵.

⁴ OTB is developed by the National Centre for Space Studies (CNES), France

⁵ A detailed description of the two SAR pre-processing chains (GRS and OTB) together with the results of the evaluation analysis are available in *O3.D3: Intermediate validation results: SAR pre-processing and Burned Area detection*.

The Sentinel-1 processing chain with OTB (S1-OTB) was developed by the Centre for the Study of the Biosphere from Space (CESBIO) as an operational tool for Sentinel-1 GRD data tiling and processing per the 100 km NATO MGRS (Military Grid Reference System) used by the Sentinel-2 processing system. The chain is highly scalable (multithreading/multiprocessor) and autonomous once few parameters are set. The chain also deals with data download from the PEPS (Plateforme d'Exploitation des Produits Sentinel) repository that mirrors ESA's SciHub. Alternatively, ESA's Sentinels hub may be used for data download. S1-OTB processing may be grouped in several steps, pre-processing, and geo-reference, and multi-temporal filtering (Figure 3).

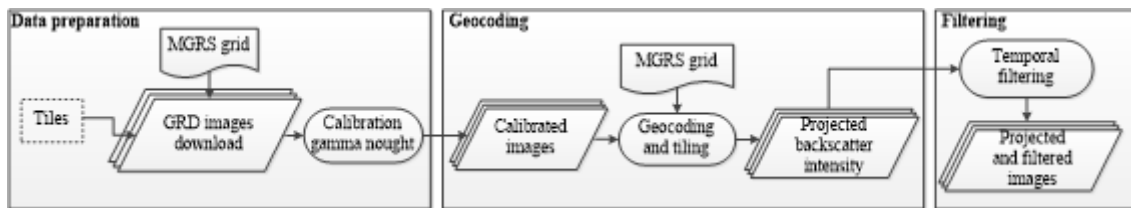


Figure 3: Flowchart for SAR data processing with Orfeo Toolbox

The pre-processing steps include data download (ascending and descending passes) for the specified MGRS tiles, calibration to gamma nought and multi-look to the desired spatial resolution. The geocoding steps include orthorectification to the desired spatial resolution, subset Sentinel-1 data to the current processing tile as well as slice assembly for data acquired from the same orbital path but provided within different slices. The last step of the OTB chain is multi-temporal filtering of the products according to satellite pass. Notice that, OTB pre-processing chain setup provides images at 20 m spatial resolution. To further reduce speckle and, more importantly, the BA algorithm processing time, the temporally filtered Sentinel-1 images are resampled (i.e. aggregated) to a coarser (40 m) spatial resolution.

5.3.2 Burned area classification

The algorithm considers multi-temporal changes of incoherent SAR-based metrics (i.e. backscattering coefficient intensities), for each SAR channel (i.e., VV and VH). The algorithm assumes time series of images that have been collected with a reasonably short time gap between them (Table 4. 4). As such, the algorithm is sensitive to the timing of images acquired in epoch -1 and +1 which need to be acquired within a reasonable amount of time as recovering vegetation may obscure the burn signal.

For the classification, consecutive acquisitions of the study area are needed to detect the burned area. The reason is that, in general, for undisturbed vegetated areas, Sentinel-1 backscatter coefficient tends to be relatively stable; for backscatter measured between scenes before and after a burn event, the backscatter coefficient drops significantly (up to 3 dB over forested areas) and is again stable between scenes following a burn event.

Table 4. List of Sentinel-1 SAR data being used in the classification

Classification Input Layers (from 4-scene time series)	Description (data means co- and cross-polarized backscatter)
Epoch -2 Backscatter Intensity (γ°)	Radiometric, Temporally Filtered and Terrain Corrected C-band backscatter – power – acquired two dates before burn.
Epoch -1 Backscatter Intensity (γ°)	Radiometric, Temporally Filtered and Terrain Corrected C-band backscatter – power – acquired before burn.

Epoch +1 Backscatter Intensity (γ°)	Radiometric, Temporally Filtered and Terrain Corrected C-band backscatter – power – acquired after burn.
Epoch +2 Backscatter Intensity (γ°)	Radiometric, Temporally Filtered and Terrain Corrected C-band backscatter – power –acquired two dates after burn.

We classify burned area between epochs -1 and +1. Epoch -2 is an earlier acquisition date; epoch +2 is a more recent acquisition date. Epoch -2 is needed as an early check on previous changes and to establish the status of the area before burn. Between epoch -1 and +1 is where we detect burned area. Epoch +2 is a check that burns have indeed taken place. Epochs -2 and +2 provide also a means to account for the observed temporal decorrelation between fire and backscatter coefficient decrease. Temporal indices are computed: i) between epochs -2 and -1, ii) between epochs -1 and +1, and iii) between epoch +1 and +2. The temporal indices may be computed based on the co-polarized channel (over areas where only VV polarization is available), and the cross-polarized channel for areas where both polarizations are acquired. The algorithm has six stages with its simplified structure being provided in Figure 4. The following paragraphs explain in detail each stage.

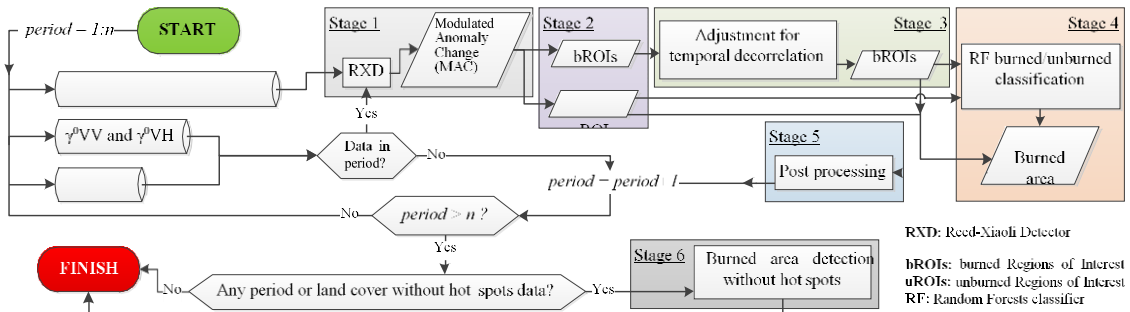


Figure 4: Burned area detection algorithm flowchart

5.3.2.1 Stage 1: Anomaly change detection

An anomalous change implies variations outside the typical behaviour expected for a given area and time. Burned areas are considered anomalies since fires are inconsistent spatial and temporal events. The Reed-Xiaoli detector (RXD), proposed by (Reed and Yu 1990), detects signatures that are distinct from the surroundings without the need for *a priori* information. Anomalies have two characteristics that make them outliers: (i) spectral signatures differ when compared to the surrounding pixels; and (ii) low occurrence probability (Kwon and Nasrabadi 2005; Stein et al. 2002). RXD uses the covariance matrix which calculates the Mahalanobis distance from a given pixel to the mean of the background pixels (Dabbiru et al. 2012). For a given pixel x , the score of the Anomalous Change (AC) for the current detection period (CDP) is given by Eq. 1:

$$AC(x) = (x - \mu)^T C^{-1} (x - \mu) \quad (1)$$

where μ is a vector composed by the mean value of the background pixels in each image band and C is the covariance matrix of the images bands of background pixels.

The background value may be computed as the global sample mean of a subset image. When *a priori* information is available the background value may be computed from areas where anomalies are not expected. For BA detection, *a priori* information was provided by MODIS and VIIRS active fire databases. MODIS and VIIRS hotspots corresponding



to the CDP were used to mask areas likely affected by fires while the remaining pixels were used to calculate the background values. The burned area masks were derived by taking a buffer of 0.75 km around each hotspot. The buffer was considered the influence area of each individual hotspot (IAhs) and it roughly corresponds to the pixel size for VIIRS and MODIS thermal channels while also considering location uncertainty.

The RXD was applied to a set of temporal ratios of backscatter coefficient in power units (Eq. 2 and 3). Such temporal indices were previously used for estimating the impact of different disturbance agents (e.g., fire, insect, wind) on vegetation (Tanase et al. 2018; Tanase et al. 2015). The selected temporal radar indices mainly use the VH backscatter which is more responsive to volumetric scattering from vegetation and less affected by changes in surface properties (e.g. soil moisture, surface roughness) when compared to the co-polarized (VV polarization) channel (Freeman and Durden 1998; van Zyl et al. 2011; Yamaguchi et al. 2005).

$$RI_1 = \gamma^0 VH_{t-1} / \gamma^0 VH_{t+1} \quad (2)$$

$$RI_2 = (\gamma^0 VH_{t-1} / \gamma^0 VV_{t-1}) / (\gamma^0 VH_{t+1} / \gamma^0 VV_{t+1}) \quad (3)$$

where γ^0 is the backscatter intensity (linear scale) of VV or VH polarizations, $t - 1$ and $t + 1$ are pre- and post-detection dates, being the CDP delimited by these.

To reduce errors related to signal variation due to fire unrelated sources (e.g. variation in soils surface moisture, vegetation regrowth) the AC values for the CDP, are modulated by the AC values recorded for the previous detection period (Eq. 4). Practically, AC scores for the most recently detected change are subtracted from the current AC scores. The result is a Modulated Anomalous Changes (MAC) score used in all subsequent detection stages.

$$MAC(x) = AC(x)_{[t-1..t+1]} - AC(x)_{[t-2..t-1]} \quad (4)$$

5.3.2.2 Stage 2: Burned and unburned regions of interest

In this stage, burned and unburned Regions of Interest (ROIs) were automatically extracted using the MAC scores and ancillary information from hotspots and land cover data. Burned ROIs (bROI) were extracted in two steps seeding and growing, an approach previously used for global burned area mapping algorithms (Alonso-Canas and Chuvieco 2015; Bastarrika et al. 2011). To obtain the seeds, spatially connected IAhs pixels were first grouped in uniquely identified objects ($q_1 : n$, where n is the number of the unique objects). A pixel x inside an object q , was considered burned seed (bSeed) if conditions in Eq. 5 were met.

$$x = \text{bSeed if} \quad (5)$$
$$\begin{aligned} &MAC(x) \geq \min(s, v) > 0 \text{ OR} \\ &MAC(x) \geq \max(s, v) > 0 \text{ AND } \min(s, v) < 0 \end{aligned}$$

where $s = \mu(MAC_{q'})$, with μ being the mean and q' a region around q bounded by $dist_q$ and $dist_q + \sqrt{dist_q}$, with $dist_q$ being the maximum span of object q . Thus, q' delineates likely unburned areas near q ; and $v = \mu(MAC_{N_G})$, with N_G being the neighbor pixels of G , where G is a pool of pixels inside q with MAC values below $\mu(MAC_q)$.

The bSeed pixels were extracted considering the major land cover type for each q object. Therefore, pixels in q' region were stratified by land cover type with only pixels of the same land cover type as q being used for computations. In addition, the selected q' pixels need to be outside the IAhs of any other hotspot. Figure 5 shows graphically the concepts of q , q' , and $dist_q$. Once bSeed pixels for q were extracted, an open morphological operator (3×3 window) was used to eliminate isolated pixels.

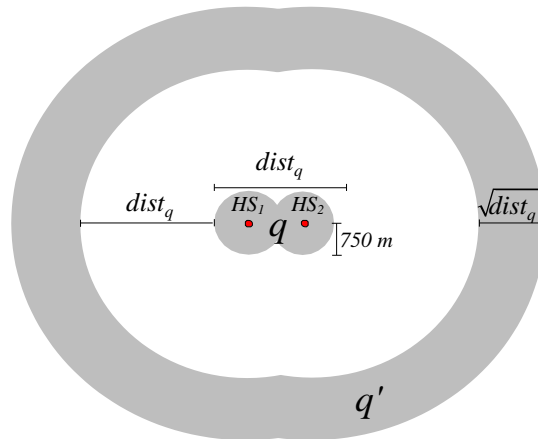



Figure 5: Graphical representation of concepts needed to extract bROIs, HS-hotspot, dist-distance.

The growing phase started by masking out image pixels when their MAC values were below the mean value for the predominant land cover class (k) of the object q . The remaining pixels were used to compute the MAC value used as the minimum threshold to label Likely Burned Pixels (LBP) (Eq. 6).

$$x = \text{LBP if} \quad \text{MAC}(x) > \mu(\text{MAC}_k) \quad (6)$$

Connected LBP pixels were grouped and subsequently overlapped with the extracted bSeed pixels for the q objects. LBP groups overlapping bSeed pixels were assigned to the bROIs and constitute the first component of the detected burned areas. The second component was detected using no parametric classification (i.e., random forests) as explained in Section 5.3.2.4.

The unburned ROIs (uROIs) were derived iteratively by land cover type. The histogram of bROIs pixels identified in the previous step was used to calculate the MAC values for the 25 and 75 percentiles (P_{25} and P_{75} , respectively). These values constituted thresholds used to classify the MAC image in burned (1) and unburned (0). Pixels with MAC values below P_{25} or above P_{75} were considered possible unburned seeds since: (i) MAC values below P_{25} indicate small changes, likely unrelated to fires (e.g. vegetation growth, changes in vegetation water content); and (ii) MAC values above P_{75} were usually associated with significant changes such as logging, crop harvesting, or floods. One should note that, high severity fires may also result in MAC values above P_{75} . However, such areas are regularly associated to hotspots and therefore were not labelled as uROIs. An open morphological operator (3×3 window) was applied to the classified binary image to remove noise. The effect of the open morphological operator is increasing the number of unburned (0) pixels. The uROIs, for the current land cover type, were obtained by: (i) filtering out pixel overlapping IAhs or bROIs; and (ii) adding all pixels from the not burnable LC map classes (i.e. bare soils, water, snow and ice, urban areas). Additionally, for the crop land cover class, groups of unburned pixels over 56 ha (0.75×0.75 km) not overlapping hotspots were included as uROIs to account for fire-unrelated changes such as crop harvesting or changes in surface properties (roughness) due to agricultural works (e.g., ploughing).

	Fire_cci Algorithm Theoretical Basis Document LDA South America	Ref.: Fire_cci_O3.D1_ATBD-S1-SA_2.0
		Issue 2.0 Date 31/08/2019
		Page 19

5.3.2.3 Stage 3: Adjustment for temporal decorrelation

During algorithm development, a temporal decorrelation between fire events (i.e., hotspots date) and backscatter coefficient change was observed. Such decorrelation events may be the result of delayed backscatter decrease after fire due to multiple factors including: (i) pre-fire conditions, e.g. drier than usual weather may result in low values for the pre-fire backscatter coefficient; (ii) post-fire weather, e.g. precipitations may temporally increase the backscatter coefficient; and (iii) vegetation-dependent backscatter response to fire events. For example, over forests, VH backscatter decrease may be delayed as there are still sufficient scattering elements (tree trunks and branches) present after fire. As time passes, trunks and branches dry up which results in decreased backscattering from vegetation.

To account for temporal decorrelation, the burned area was detected iteratively for each period. Delayed changes in backscatter were accounted for by using the bROIs detected in periods formed by the current pre-fire image ($t-1$) and images acquired during the following 90 days past the CDP (i.e., $t+2$, $t+3$). Such bROIs were labelled as burned in the CDP ($t-1$ to $t+1$) when overlapping hotspots from the CDP. Additionally, these bROIs must not overlap hotspot recorded past the CDP.

5.3.2.4 Stage 4: Random forests burned/unburned classification

Only a fraction of the anomalous pixels was labelled as burned based on information from hotspots due to the rather restrictive criteria used in Stage 2. Pixels not meeting the imposed criteria also needed labelling. To avoid subjectivity, such pixels were labelled using a land cover specific non-parametric classifier (i.e., random forests) trained with data extracted from bROIs and uROIs. Random forests (RF) is an ensemble classifier that consists in a group of decision trees $\{h(x, \Theta_z, z=1, \dots)\}$, where x is the input vector, and Θ_z are independently bootstrap sampled vectors, with replacement, in each decision tree (z). Each tree provides a unique class for x , being the class of x assigned as the most popular voted class by the trees group (Breiman 2001). In this study, *TreeBagger* from the MATLAB® software package was used to construct the RF classifiers.

RF classifiers are customizable through different parameters such as: (i) number of trees; (ii) number of training samples; (iii) proportion of training samples by class; and (iv) number of independent variables. The number of trees is a key adjustment in RF classification since for more trees the generalization error converges and models are not over-fit (Breiman 2001; Pal 2005; Rodriguez-Galiano et al. 2012). On the other hand, using more trees demands more computational resources. An empirical analysis (not shown) concluded that 250 trees provided the best trade-off between speed and accuracy for burned area classification in this study. Since the number of pixels in bROIs and uROIs is high, computational costs may be reduced by using just a fraction for training purposes. This fraction was determined, by land cover classes, as 1% of all bROIs and uROIs pixels divided by the number of trees (250).

Unbalanced training samples may result in infra-classification of the minority classes. According to (Chen et al. 2004), several approaches may be used to address such problems: (i) reducing the overall learning cost, with high costs being assigned to the misclassification of the minority classes (Pazzani et al. 1994); (ii) under-sampling the majority and over-sampling the minority classes; or (iii) a combination of both techniques (Chawla et al. 2002). The latter approach was used in this study. Depending on the misclassification cost, the *TreeBagger* function generates in-bag samples by oversampling



the burned class and under sampling the unburned class. The proportion of training data was empirically adjusted to 40% and 60% for burned and unburned classes respectively.

The number of variables considered for trees growing in each split was computed as the square root of the total number of variables (Gislason et al. 2006), as it reduces the correlation of trees and thus improves global accuracy (Gislason et al. 2006; Rodriguez-Galiano et al. 2012). In addition to the SAR based metrics used for RXD (Eq. 2 and 3), up to 30 SAR metrics were used for RF classification. These metrics were computed as in Eq. 7 to 12. The non-parametric classification was carried out considering the land cover type with specific models being built for each land cover class. The burned area detected by RF was added to the one detected in Stage 2 (bROIs) and formed the total BA for the CDP.

$$\mu(\gamma^0 XY_{[t\oplus, t-1]}) - \gamma^0 XY_{t+i} \quad (7)$$

$$\mu(\gamma^0 XY_{[t\oplus, t-1]}) / \gamma^0 XY_{t+i} \quad (8)$$

$$\gamma^0 XY_{t-1} - \gamma^0 XY_{t+i} \quad (9)$$

$$\gamma^0 XY_{t-1} / \gamma^0 XY_{t+i} \quad (10)$$

$$(\gamma^0 VH_{t-1} / \gamma^0 VV_{t-1}) / (\gamma^0 VH_{t+i} / \gamma^0 VV_{t+i}) \quad (11)$$

$$\mu(\gamma^0 VH_{[t\oplus, t-1]} / \gamma^0 VV_{[t\oplus, t-1]}) / (\gamma^0 VH_{t+i} / \gamma^0 VV_{t+i}) \quad (12)$$

where $\gamma^0 XY$ is the backscatter intensity (linear scale) of VV and VH polarizations, t' is $t-1$ minus the double of days distance between $t-1$ and $t+1$, and i is 1 or 2, being 30 the maximum number of indices computed.


5.3.2.5 Stage 5: Post processing

Post-processing was needed to account for temporal decorrelation and improve detection results over problematic land covers such as cropping areas. To adjust for temporal decorrelation, the BA detected by the non-parametric classifier for the CDP was compared to the IAhs of previous detection periods, up to 90 days before the pre-fire image ($t-1$). If burned areas detected in the current CDP (i.e., objects formed by contiguous pixels) overlap previous IAhs (objects) by more than 75% they were masked out and considered previous burns. Three additional post-processing steps were then carried out to further improve the results: (i) on cropping lands, groups of burned pixels (objects) with areas above 56 ha that did not overlap IAhs (i.e., no local hotspot) are removed. The rationale was that the lack of hotspots over a large changing cropping area is an indication of harvesting rather than fire; (ii) burned objects below one hectare were removed; and (iii) a modal filter with a convolution kernel of 3×3 pixels was applied to smooth the salt and pepper effects typical for SAR based classifications.

5.3.2.6 Stage 6: Burned area detection without hotspots

As clouds may prevent the propagation of radiation from active fires to the thermal sensors on board satellites, the algorithm was built with a backup mechanism to cope with the absence of hotspots for a specific land cover type and detection period. However, for the algorithm to work, hotspots need to be available for each land cover class at some point during the analysed fire season.

The algorithm first detected the burned area for all land cover types during detection periods for which hotspots were available. For detection periods with no hotspots, the data were temporally stored for later processing. During detection, the algorithm saved a database containing the P_{25} and P_{75} MAC values (Stage 2) and the trained RF models (Stage 4) for each land cover class. The database is hereinafter referred to as the Classifier Model


	Fire_cci Algorithm Theoretical Basis Document LDA South America	Ref.:	Fire_cci_O3.D1_ATBD-S1-SA_2.0	
		Issue	2.0	Date 31/08/2019
		Page	21	

and Criteria (CMC). Once detections for land cover classes and detection periods with hotspots ended, the CMC database was used to classify the temporally stored data (i.e., land cover types during detection periods without hotspots) if two conditions were met: (i) the current detection period was within the fire season. The fire season was computed using the hotspots daily frequency, for the processed MGRS tile, as the interval between the dates corresponding to the 5th and respectively the 95th percentiles; and (ii) the difference between the detection period and the date for the nearest CMC was less than a month, thus avoiding possible confusions due to changes in vegetation phenology. When CMC entries from different detection periods met the second condition, the one closest to the CDP was used. The MAC image for the CDP was segmented into burned and unburned based on the CMC P_{25} and P_{75} with the possible burned pixels being subsequently classified using the stored RF models by land cover class. When CMC entries are spaced equally in time when compared to the CDP (i.e., one entry is from a previous period and one from a posterior period), each entry was used separately and only the commonly detected burned area was kept. The post processing operations from stage 5 were carried out on the detected BA from this stage.

An additional operation was carried out to reduce possible commission errors during this stage. The operation was carried out over burned areas detected on different relative orbits. Note that detections always used time-series of images from the same relative orbit. If several relative orbits intersect a given tile, the algorithm worked through the data from each relative orbit separately. BA products composites were subsequently formed using detections from different relative orbits and the same detection period. For each detection period, burned area pixels detected in different relative orbits were grouped in objects. If all pixels of an object were classified as not burned in one orbit the object was removed from the detected burned area for the CDP. Since, dual pass (ascending and descending) acquisition were not available for all tiles and spatially overlapping relative orbits only partially cover any given tile, this additional operation reduces commission errors where burned area detections intersect.

6 Uncertainty estimation framework

Burn probability of each pixel was estimated through an analysis based on the statistical similarity of temporal radar indices (Eqs. 2 and 3) between each pixel with the mean of bROIs of the same land cover type (k). The RXD (Eq. 1) is employed again, but using as background the bROIs(k). Thus, the RXD(x) show the statistical distance between a given pixel x of land cover k and the bROIs(k). Notice that bROIs were selected based on significant changes of the SAR signal and the exitance of hotspots at the same location which indicates a very strong likelihood of burned areas. The conversion of RXD(x) values to burn probability was based on an empirical cumulative distribution function (eCDF) where a low RXD(x) value corresponds to low differences to bROIs statistics which in turn indicates a high burn probability. Pixels detected as burned that are located within a 0.75 km buffer centered at each hotspot receive a maximum burn probability (100%). Pixels detected as unburned received a burn probability of 0%. Therefore, uncertainty values are relevant for pixels labels as burned which is a caveat of the current uncertainty framework adopted.

	Fire_cci	Ref.: Fire_cci_O3.D1_ATBD-S1-SA_2.0
	Algorithm Theoretical Basis Document	Issue 2.0 Date 31/08/2019
	LDA South America	Page 22

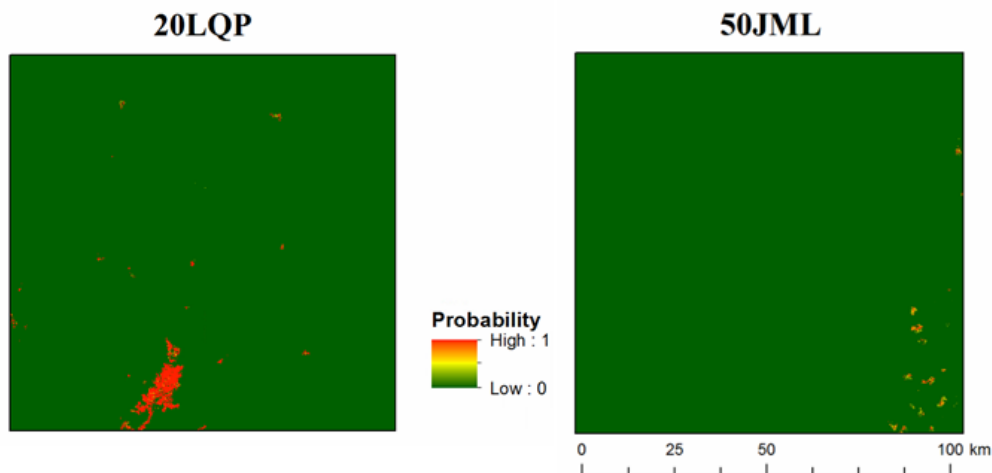




Figure 6: Example of burn probability for two different tiles

7 References

- Alonso-Canas, I., & Chuvieco, E. (2015). Global burned area mapping from ENVISAT-MERIS and MODIS active fire data. *Remote Sensing of Environment*, 163, 140-152
- Andreae, M.O., & Merlet, P. (2001). Emission of trace gases and aerosols from biomass burning. *Global Biogeochemical Cycles*, 15, 955-966
- Bastarrika, A., Chuvieco, E., & Martin, M.P. (2011). Mapping burned areas from Landsat TM/ETM+ data with a two-phase algorithm: Balancing omission and commission errors. *Remote Sensing of Environment*, 115
- Bourgeau-Chavez, L.L., Kasischke, E.S., Brunzell, S., & Mudd, J.P. (2002). Mapping fire scars in global boreal forests using imaging radar data. *International Journal of Remote Sensing*, 23, 4211-4234
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32
- Chawla, N.V., Bowyer, K.W., Hall, L.O., & Kegelmeyer, W.P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357
- Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, 110
- Chuvieco, E. (1999). Measuring changes in landscape pattern from satellite images: short-term effects of fire on spatial diversity. *International Journal of Remote Sensing*, 20, 2331-2346
- Chuvieco, 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
- Dabbiru, L., Aanstoos, J.V., Mahrooghy, M., Li, W., Shanker, A., & Younan, N.H. (2012). Levee anomaly detection using polarimetric synthetic aperture radar data. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International* (pp. 5113-5116)
- Fieber, K.D., Davenport, I.J., Tanase, M., Ferryman, J.M., Gurney, R.J., Becerra, V.M., Walker, J.P., & Hacker, J.M. (2015). Validation of Canopy Height Profile methodology for small-footprint full-waveform airborne LiDAR data in a discontinuous canopy environment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104, 144-157
- Freeman, A., & Durden, S.L. (1998). A Three-Component Scattering Model for Polarimetric SAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, 36, 963-973

	Fire_cci	Ref.: Fire_cci_O3.D1_ATBD-S1-SA_2.0
	Algorithm Theoretical Basis Document	Issue 2.0 Date 31/08/2019
	LDA South America	Page 23

- 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 (GFED). *Journal of Geophysical Research: Biogeosciences*, 118, 317-328
- Gislason, P.O., Benediktsson, J.A., & Sveinsson, J.R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27, 294-300
- Hantson, S., Padilla, M., Corti, D., & Chuvieco, E. (2013). Strengths and weaknesses of MODIS hotspots to characterize global fire occurrence. *Remote Sensing of Environment*, 131, 152-159
- Huang, S., & Siegert, F. (2006). Backscatter Change on Fire Scars in Siberian Boreal Forests in ENVISAT ASAR Wide-Swath Images. *IEEE Transactions on Geoscience and Remote Sensing Letters*, 3, 154-158
- Inglada, J., & Christophe, E. (2009). The Orfeo Toolbox remote sensing image processing software. In IEEE (Ed.), *Geoscience and Remote Sensing Symposium* (pp. 733-736). Cape Town, South Africa: IEEE
- Kasischke, E.S., Tanase, M.A., Bourgeau-Chavez, L.L., & Borr, M. (2011). Soil moisture limitations on monitoring boreal forest regrowth using spaceborne L-band SAR data. *Remote Sensing of Environment*, 115, 227-232
- Kloster, S., Mahowald, N.M., Randerson, J.T., & Lawrence, P.J. (2012). The impacts of climate, land use, and demography on fires during the 21st century simulated by CLM-CN. *Biogeosciences*, 9
- Kwon, H., & Nasrabadi, N.M. (2005). Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 43, 388-397
- Laneve, G., Castronuovo, M.M., & Cadau, E.G. (2006). Continuous Monitoring of Forest Fires in the Mediterranean Area Using MSG. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 2761-2768
- Menges, C.H., Bartolo, R.E., Bell, D., & Hill, G.J.E. (2004). The effect of savanna fires on SAR backscatter in northern Australia. *International Journal of Remote Sensing*, 25, 4857-4871
- Padilla, M., Stehman, S.V., Hantson, S., Oliva, P., Alonso-Canas, 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
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26, 217-222
- Pazzani, M., Merz, C., Murphy, P., Ali, K., Hume, T., & Brunk, C. (1994). Reducing misclassification costs. *Machine Learning Proceedings 1994* (pp. 217-225): Elsevier
- Reed, I.S., & Yu, X. (1990). Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 38, 1760-1770
- Rignot, E., Despain, D.G., & Holecz, F. (1999). The 1988 Yellowstone fires observed by imaging radars. In *Proceedings of the Joint Fire Sciences Conference and Workshop* (pp. 1-9). 15-17 June, Boise, Idaho: Univ. of Idaho and the Int. J. of Wildland Fire
- Rignot, E.J.M., & Zyl, J.J.v. (1993). Change Detection Techniques for ERS-1 SAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, 31, 896-906

	Fire_cci Algorithm Theoretical Basis Document LDA South America	Ref.:	Fire_cci_O3.D1_ATBD-S1-SA_2.0	
		Issue	2.0	Date 31/08/2019
		Page	24	

- Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J.P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104
- 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
- Simmonds, P.G., Manning, A.J., Derwent, R.G., Ciais, P., Ramonet, M., Kazan, V., & Ryall, D. (2005). A burning question. Can recent growth rate anomalies in the greenhouse gases be attributed to large-scale biomass burning events? *Atmospheric Environment*, 39, 2513-2517
- Stein, D.W.J., Beaven, S.G., Hoff, L.E., Winter, E.M., Schaum, A.P., & Stocker, A.D. (2002). Anomaly detection from hyperspectral imagery. *IEEE signal processing magazine*, 19, 58-69
- Stroppiana, D., Tansey, K., Grégoire, J.-M., & Pereira, J.M.C. (2003). An Algorithm for Mapping Burnt Areas in Australia Using SPOT-VEGETATION Data. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 907-909
- Tanase, M.A., Aponte, C., Mermoz, S., Bouvet, A., Le Toan, T., & Heurich, M. (2018). Detection of windthrows and insect outbreaks by L-band SAR: A case study in the Bavarian Forest National Park. *Remote Sensing of Environment*, xxx, xxx-xxx
- Tanase, M.A., de la Riva, J., Santoro, M., Le Toan, T., & Perez-Cabello, F. (2010). Sensitivity of X-, C- and L-band SAR backscatter to fire severity in mediterranean pine forests. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 3663-3675
- Tanase, M.A., Kennedy, R., & Aponte, C. (2015). Radar Burn Ratio for fire severity estimation at canopy level: An example for temperate forests. *Remote Sensing of Environment*, 170, 14-31
- van Zyl, J.J., Arii, M., & Kim, Y. (2011). Model-Based Decomposition of Polarimetric SAR Covariance Matrices Constrained for Nonnegative Eigenvalues. *IEEE Transactions on Geoscience and Remote Sensing*, 49, 3452-3459
- Yamaguchi, Y., Moriyama, T., Hiroyoshi, M.I., & Shinji, Y. (2005). Four-Component Scattering Model for Polarimetric SAR Image Decomposition. *IEEE Transactions on Geoscience and Remote Sensing*, 43, 1699-1706



Annex 1: Acronyms and abbreviations

ATBD	Algorithm Theoretical Basis Document
BA	Burned Area
BAMS	Burned Area Mapping Software
CBERS-4	China-Brazil Earth Resources Satellite 4
CCI	Climate Change Initiative
CCN	Contract Change Notice
CESBIO	Centre for the Study of the Biosphere from Space
DEM	Digital Elevation Model
ECV	Essential Climate Variables
ERS-2	European Remote Sensing 2 satellite
ESA	European Space Agency
FIRMS	Fire Information for Resource Management System
GCOS	Global Climate Observing System
GRD	Ground Range Detected
GRS	Gamma Remote Sensing
INPE	Instituto Nacional de Pesquisas Espaciais (Brazil)
IPCC	Intergovernmental Panel on Climate Change
IW	Interferometric Wide swath
LC	Land Cover
LC_cci	Land Cover CCI
LDA	Large demonstrator area
MGRS	Military Grid Reference System
MLI	Multi-Looked Image
MODIS	Moderate-Resolution Imaging Spectroradiometer
MSI	Multi Spectral Instrument
NASA	National Aeronautics and Space Agency
NTDI	Normalized Difference Temporal Index
OLI	Operational Land Imager
OTB	Orfeo ToolBox
PEPS	Plateforme d'Exploitation des Produits Sentinel
PSD	Product Specification Document
RI	Radar Indices
SAR	Synthetic Aperture Radar
SFD	Small Fire Database
SLC	Single Look Complex
SPOT	Satellite Pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
TIRS	Thermal InfraRed Sensor
TOA	Top of Atmosphere
TOPSAR	Terrain Observation with Progressive Scans SAR
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
VIIRS	Visible Infrared Imaging Radiometer Suite
WGS84	World Geodetic System 1984
WRS	Worldwide Reference System