ESA Climate Change Initiative – Fire_cci
D2.1.3 Algorithm Theoretical Basis Document (ATBD) – MODIS

Project Name	ECV Fire Disturbance: Fire_cci Phase 2
Contract №	4000115006/15/I-NB
Issue Date	11/06/2018
Version	1.1
Author	Joshua Lizundia-Loiola, M. Lucrecia Pettinari, Emilio Chuvieco, Thomas Storm, José Gómez-Dans
Document Ref.	Fire_cci_D2.1.3_ ATBD-MODIS_v1.1
Document type	Public

Project Partners

Prime Contractor/ Scientific Lead & Project Management	UAH – University of Alcala (Spain)
Earth Observation Team	UAH – University of Alcala (Spain) EHU – University of the Basque Country (Spain) UL – University of Leicester (United Kingdom) UCL – University College London (United Kingdom) ISA – School of Agriculture, University of Lisbon (Portugal)
System Engineering	BC – Brockmann Consult 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 - Vrije Universiteit Amsterdam (Netherlands)

Distribution

Affiliation	Name	Address	Copies
ESA	Stephen Plummer (ESA)	stephen.plummer@esa.int	electronic copy
Project Team	Emilio Chuvieco (UAH)	emilio.chuvieco@uah.es	electronic copy
	M. Lucrecia Pettinari (UAH)	mlucrecia.pettinari@uah.es	
	Joshua Lizundia Loiola (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	
	M. Vanesa Moreno (IRD)	mariavanesa.morenodominguez@cefe...	
	Philippe Ciais (LSCE)	philippe.ciais@lsce.ipsl.fr	
	Chao Yue (LSCE)	chaoyuejoy@gmail.com	
	Pierre Laurent (LSCE)	pierre.laurent@lsce.ipsl.fr	
	Guido van der Werf (VUA)	guido.vander.werf@vu.nl	
	Ioannis Bistinas (VUA)	i.bistinas@vu.nl	

Summary

This document presents the technical basis of the algorithms used to generate the Fire_cci Burned Area product v5.0, based on MODIS 250m data. The document analyses the input requirements and the process to create the product, including the stages of the processor to get the burned area data and the formatting of the data to obtain the released product.

	Affiliation/Function	Name	Date
Prepared	UAH BC UCL	Joshua Lizundia Loiola M. Lucrecia Pettinari Emilio Chuvieco Thomas Storm José Gómez-Dans	11/06/2018
Reviewed	UAH – Project Manager	Lucrecia Pettinari	11/06/2018
Authorized	UAH - Science Leader	Emilio Chuvieco	11/06/2018
Accepted	ESA - Technical Officer	Stephen Plummer	11/06/2018

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

Document Status Sheet

Issue	Date	Details
1.0	20/03/2018	First issue of the document.
1.1	11/06/2018	Addressing comments of ESA-CCI-FIRE-EOPS-MEM-18-0135

Document Change Record

Issue	Date	Request	Location	Details
1.1	11/06/2018	ESA	Sections 2.1, 2.2.2, 2.2.4, 2.3.2, 2.5, 2.5.1, 2.5.3, 2.6	Minor changes in the text to improve the explanations.
		ESA	Figures 6, 10, 14	Title expanded
		ESA	Figures 17, 22	Figure updated

<u>**Table of Contents**</u>

List of Tables

List of Figures

1. Introduction

The ESA Climate Change Initiative (CCI) 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 Variable (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.

Fire is one of the ECVs due to its importance in emissions calculation and their consequence at social, economic and environmental levels. The Fire disturbance ECV identifies burned area (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.

This document explains the theoretical basis of the global burned area (BA) algorithm MODIS Fire_cci v5.0, developed as part of the Fire_cci project. The predecessor of this algorithm used images from the MERIS sensor (Alonso-Canas and Chuvieco 2015). Because of the availability of images from this sensor, the time series of the MERIS Fire_cci v4.1 product only covers the years 2005-2011 (Chuvieco et al., 2016). This new algorithm based on the MODIS sensor extends that time series to the period 2001-2016. The algorithm was developed for the two highest-spatial resolution bands of the MODIS sensor (Red and Near Infrared, NIR) and follows a similar approach to the MERIS algorithm, as it combines information from hotspots and temporal reflectance changes to detect the burned pixels. Differences in the characteristics of both sensors and their derived products (spatial and temporal resolution, view angles, bands characteristics, etc.) and problems found with Fire_cci BA v4.1 product advised to extend previous developments and introduce some adaptations for this MODIS BA algorithm.

2. BA Algorithm description

2.1. General scheme

Following our previous experience (Alonso-Canas and Chuvieco 2015) and well-known BA products (Giglio et al. 2009), we considered the hybrid approach the most effective for global BA detection. Consequently, active fire detections (hotspots, HS) and post fire reflectance information were combined. Using both thermal and reflectance information, BA are detected more unambiguously, as the thermal characterisation allows detecting active fires while final burned patches have a longer impact on post-fire reflectances.

Following the scheme used for the MERIS BA algorithm (Alonso-Canas and Chuvieco 2015, Alonso-Canas et al. 2016), the general scheme of the MODIS BA algorithm follows a two-phase approach (seed and growing phases). The former aims to detect the most clearly burned pixels, while the latter tries to reduce omission errors using contextual analysis. The algorithm is also based on *temporal* composites, using the date of the HS and distance operators to create monthly syntheses of daily data. Finally, A fuel mask, based on the LC_cci product, was used to mask out non burnable areas and to reduce some potential confusion of BA with other land cover (LC) categories.

The general scheme of the algorithm is shown in Figure 1. MODIS NIR information is combined with HS and auxiliary data to obtain monthly composites. These composites are the basis to perform the seed identification phase. Once the seeds are retrieved, MODIS input data are used to build the monthly Global Environmental Monitoring Index (GEMI) image, based on the composite dates, and the maximum GEMI of the previous month. The difference between these two images is the second variable used by the algorithm. Finally the growing phase of the algorithm is performed on the composite and GEMI difference images to obtain the BA maps.
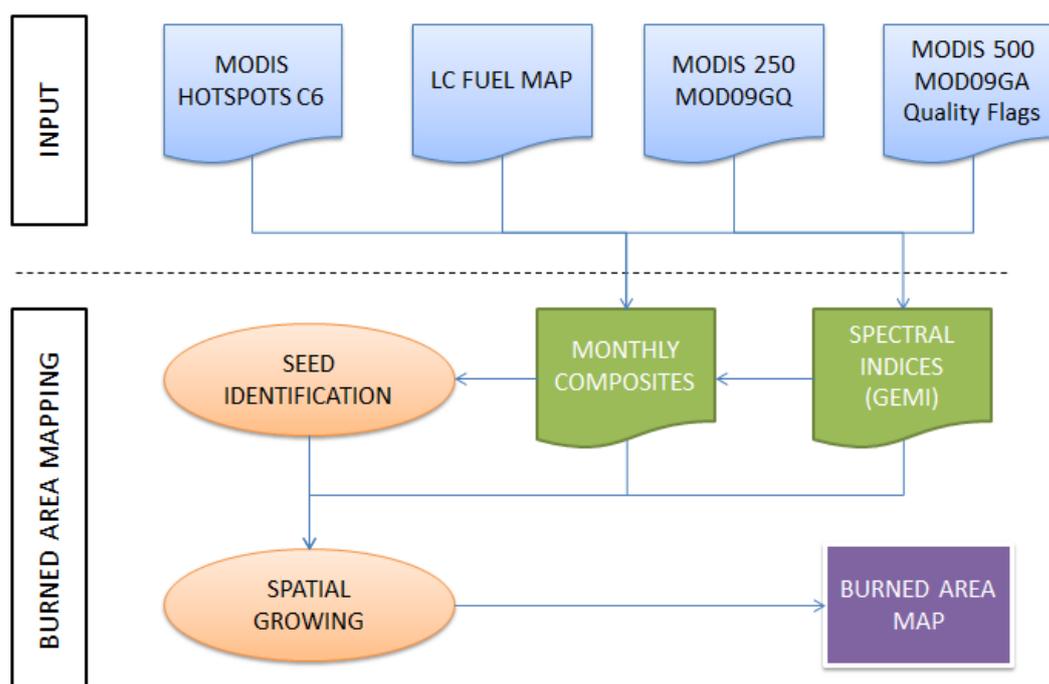
Figure 1: Algorithm main scheme definition

2.2. Algorithm inputs

2.2.1. Surface Reflectance

The objective of the algorithm is to create a BA product at a 250m spatial resolution, and for this reason the main source of data is the daily MOD09GQ Collection 6 product (https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod09gq_v006 last accessed March 2018), which offers surface reflectance information in the red (R) and Near Infrared (NIR) bands. A summary of the characteristics of the product is shown in Table 1.

Table 1: Science data set of the MOD09GQ product, extracted from Table 3 of Vermote et al. (2015).

Science Data Sets (HDF Layers)	Units	Data Type	Fill Value	Valid Range	Scale Factor
num_observations: number of observations within a pixel	none	8-bit signed integer	-1	0-127	NA
sur_refl_b01_1: 250m Surface Reflectance Band 1 (620-670 nm)	Reflectance	16-bit signed integer	-28672	-100 - 16000	0.0001

Science Data Sets (HDF Layers)	Units	Data Type	Fill Value	Valid Range	Scale Factor
sur_refl_b02_1: 250m Surface Reflectance Band 2 (841-876 nm)	Reflectance	16-bit signed integer	-28672	-100 - 16000	0.0001
QC_250m_1: 250m Reflectance Band Quality	Bit Field	16-bit unsigned integer	2995	0 - 4096	NA
obscov_1: Observation Coverage (percentage of the grid cell area covered by the observation)	Percent	8-bit signed integer	-1	0 - 100	0.01
iobs_res_1	none	8-bit unsigned integer	255	0 - 254	NA
orbit_pnt_1	none	8-bit signed integer	-1	0 - 15	NA
granule_pnt_1	none	8-bit unsigned integer	255	0 - 254	NA

The unit of analysis of the MODIS products is the standard tile (1200 x 1200 km) in sinusoidal coordinates in which they are delivered (Figure 2). All other input data has been reprojected to this coordinate system, if necessary, prior to processing.
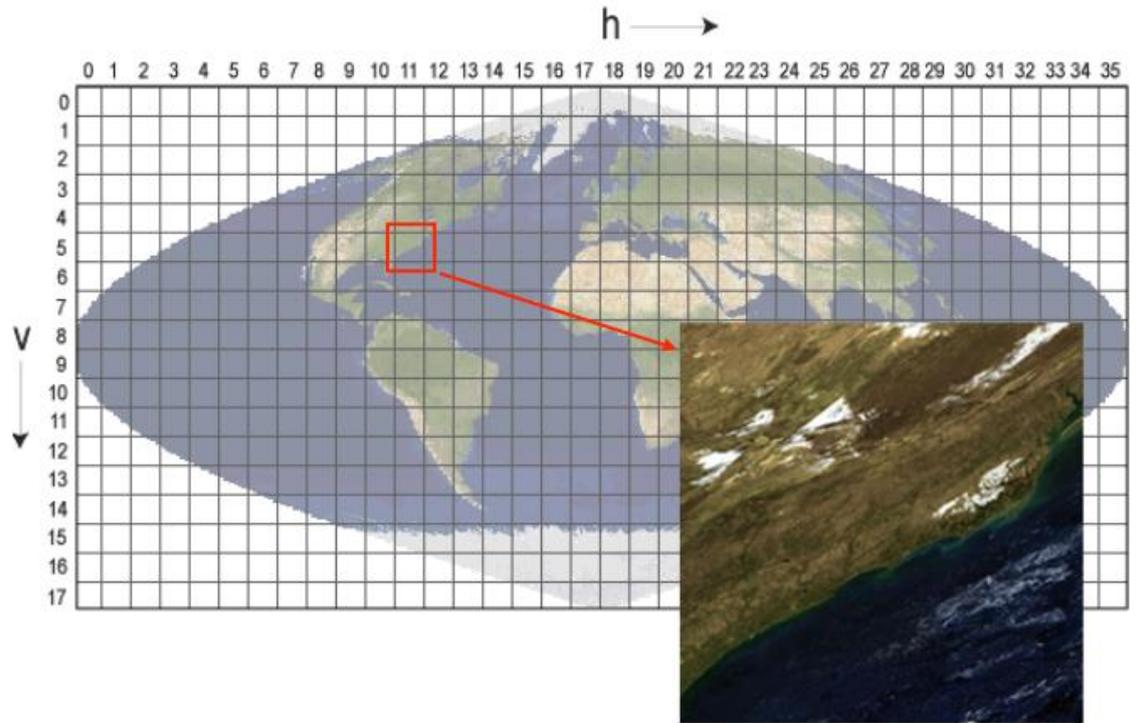
Figure 2: MODIS Sinusoidal Grid, as shown in Vermote et al. (2015).

2.2.2. Quality flags

Complementary to the surface reflectance product, the daily MOD09GA Collection 6 product is also used (https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod09ga_v006 accessed March 2018), which in Band 2 (state_1km_1: Reflectance Data State) contains the State Quality Assessment Description (QA) that offer information of the state of each pixel at a 1000m spatial resolution. Even though the use of this product at this coarser resolution might introduce some errors, the Band Quality Description of the GQ product (band 4, QC_250m_1: 250 m Reflectance Band Quality), which gives information related to issues like detectors quality, atmospheric correction quality and pixel saturation (Table 8 of Vermote et al. 2015), did not include sufficient information to perform an adequate masking of the daily images (see Figure 3), and hence it was decided to use the GA product's QA flags.

The QAs offer information stored in integer numbers ranging from 0 to 65535 (being this last number the NoData code), which translated to their binary expression offer in 16 bits the information on the quality of each pixel (Table 2).

Table 2: Bit structure of the quality flags of MOD09GA. For further information refer to Vermote et al. (2015)

Bit position	Description	Bit position	Description
0	Cloud state	8	Cirrus detected
1		9	
2	Cloud shadow	10	Internal cloud algorithm flag
3	Land/water flag	11	Internal fire algorithm flag
4		12	MOD35 snow/ice flag
5		13	Pixel is adjacent to cloud
6	Aerosol quantity	14	Salt Pan
7		15	Internal snow mask

| (a) | (b) | (c) |
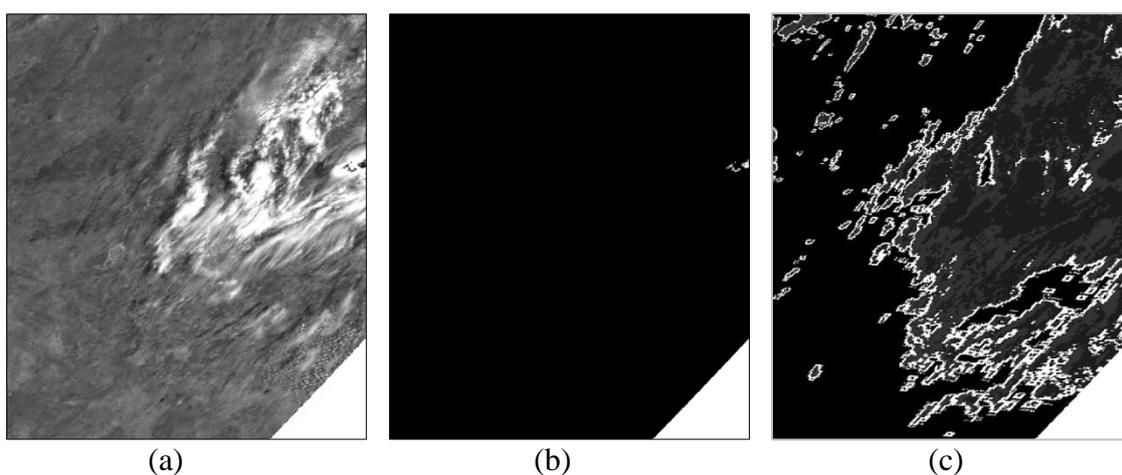
Figure 3: Images corresponding to a zone of tile h30v10 (Australia) of January 6, 2008. (a) NIR reflectance, where the clouds appear white. (b) QA information of the MOD09GQ product, indicating that all pixels in black have a good quality. (c) QA of MOD09GA, which is capable of detecting the clouds with a better precision.

The use of these QAs is necessary to avoid further problems in the BA classification, such as cloud or cloud shadows or poor observations. The impact of cloud shadows can

be observed in Figure 4. The quality flags were used to eliminate clouds and cloud shadows. Information of bits 0, 1, 2 and 10 of Table 2 was masked in the input images, keeping the rest of the values as valid.

(a)   (b)
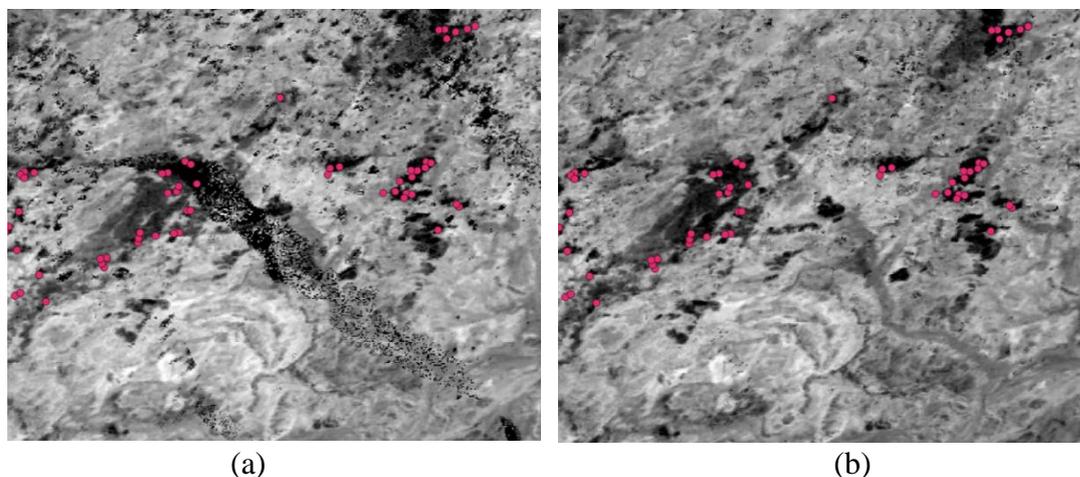
Figure 4: Example of June 2008 for tile h30v10. (a) Composite (see Section 2.3) without QA correction. (b) Composite with QA applied to the images that create it. The red dots are the hotspots for that period. The QA eliminates the dark pixels in the middle of the image that correspond to cloud shadows and not burned area.

2.2.3. Hotspots

The HS information was derived from the MCD14ML Collection 6 product. This new collection improves the detection of active fires from C5, which was used for the Fire_cci MERIS algorithm. This may increase the commission errors due to possible occurrence of hotspots caused by small fires (undetectable at our resolution) in isolated pixels whose statistics introduce noise in the analysis, but it also may decrease the omission ones (Giglio et al. 2016). This product provides global monthly information of global HS since 2001, covering the entire MODIS time series.

The MCD14ML product has a layer that specifies the presumed origin of the fire: 0 – presumed vegetation fire; 1 – active volcano; 2 – other static land source; 3 – offshore. Of these categories, only the HS classified as 0 have been used in the algorithm.

2.2.4. Land cover

The last product used is the Land Cover CCI product (LC_cci, Kirches et al. 2013), which provides information on global land cover. The collection used for the first global MODIS BA processing is v1.6.1, containing three epochs of land cover: 1998-2002 (called LC_cci 2000), 2003-2007 (LC_cci 2005) and 2008-2012 (LC_cci 2010). As the purpose of this layer is to provide information on the land cover prior to the fire, the LC_cci used is the one of the years prior to the BA analysis, i.e. LC_cci 2000 was used for the BA calculation of 2001-2007, the LC_cci 2005 for the 2008-2012 BA product, and the LC_cci 2010 for the 2013-2016 product. This version of the LC_cci product was used because when the algorithm was developed and the process started, the new version of the LC_cci v2.0.7 product (Santoro et al. 2017) was not yet available.

The LC_cci product has 37 classes (including no data), which were reclassified into 4 general classes to create a burnable mask and to differentiate between vegetation of different heights (see Table 3 and Figure 5):

- 0: non burnable
- 1: low vegetation
- 2: medium-height vegetation
- 3: high vegetation

The rationale for this reclassification is to group the land cover classes that could have a similar spectral behaviour during fire events.

Values 1, 2 and 3 determine the burnable mask, while 0 indicate the non-burnable pixels that are not taken into account when calculating statistics.

(a)   (b)
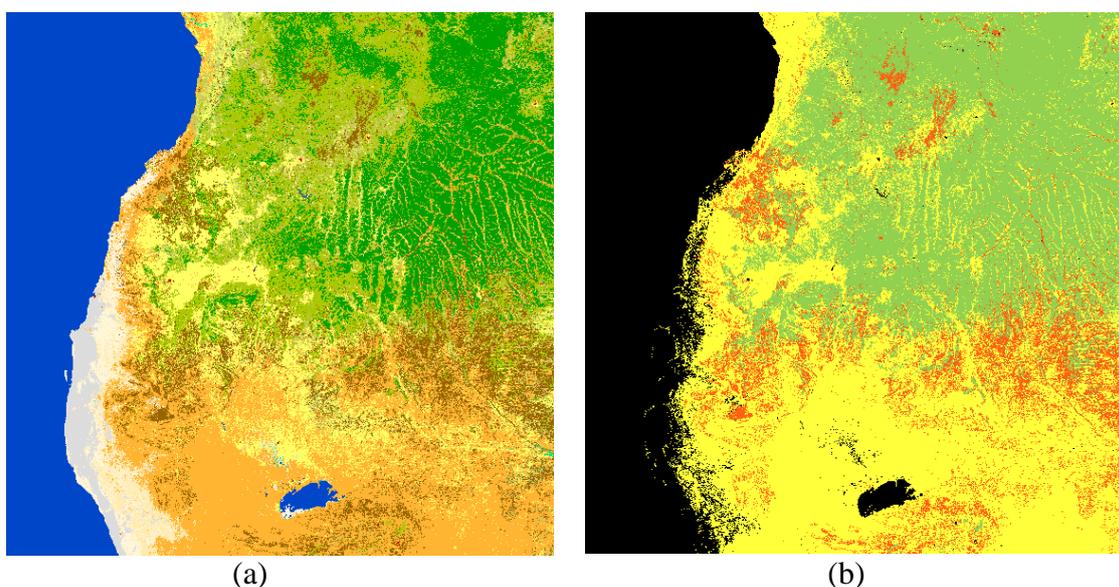
Figure 5: (a) land cover classes of the LC_cci. (b) Reclassified land cover, with the following colours: 0: black; 1: yellow; 2: orange; 3: green.

Table 3: Reclassification of the land cover extracted from LC_cci into 4 general classes

Label	LC_cci Value	Fire_cci class
No Data	0	0
Cropland, rainfed	10	1
Herbaceous cover	11	1
Tree or shrub cover	12	2
Cropland, irrigated or post-flooding	20	1
Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)	30	1
Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)	40	1
Tree cover, broadleaved, evergreen, closed to open (>15%)	50	3
Tree cover, broadleaved, deciduous, closed to open (>15%)	60	3
Tree cover, broadleaved, deciduous, closed (>40%)	61	3
Tree cover, broadleaved, deciduous, open (15-40%)	62	3
Tree cover, needleleaved, evergreen, closed to open (>15%)	70	3
Tree cover, needleleaved, evergreen, closed (>40%)	71	3

Label	LC_cci Value	Fire_cci class
Tree cover, needleleaved, evergreen, open (15-40%)	72	3
Tree cover, needleleaved, deciduous, closed to open (>15%)	80	3
Tree cover, needleleaved, deciduous, closed (>40%)	81	3
Tree cover, needleleaved, deciduous, open (15-40%)	82	3
Tree cover, mixed leaf type (broadleaved and needleleaved)	90	3
Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	100	3
Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	110	1
Shrubland	120	2
Evergreen shrubland	121	2
Deciduous shrubland	122	2
Grassland	130	1
Lichens and mosses	140	1
Sparse vegetation (tree, shrub, herbaceous cover) (<15%)	150	1
Sparse shrub (<15%)	152	2
Sparse herbaceous cover (<15%)	153	1
Tree cover, flooded, fresh or brackish water	160	3
Tree cover, flooded, saline water	170	3
Shrub or herbaceous cover, flooded, fresh/saline/brackish water	180	1
Urban areas	190	0
Bare areas	200	0
Consolidated bare areas	201	0
Unconsolidated bare areas	202	0
Water bodies	210	0
Permanent snow and ice	220	0

2.3.   Composites

2.3.1.   Reasons for compositing

The MODIS images used for the algorithm (see Section 2.2) have a noticeable angular effect, also called Bi-Directional Reflectance Distribution Function (BRDF), as shown in Figure 6. This is due to the wide field of view (FOV) of the sensor (2330 km). Different filters were tested to smooth the signal, but these filters also smoothed the reflectance drop due to the fires, making them inapplicable.
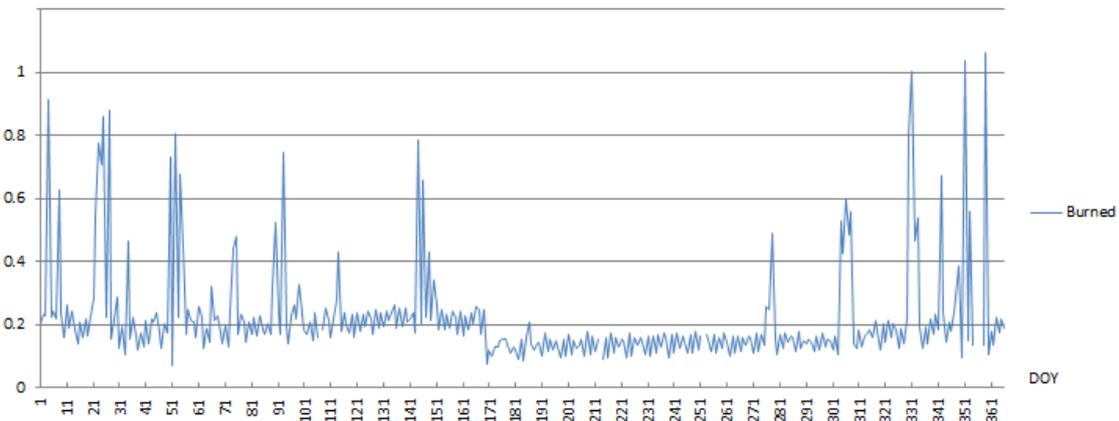
Figure 6: NIR reflectance (Y axis) for a burned pixel of tile h08v05 (California) during 2008 (days of the year in X axis) without applying state mask. Most of the high peaks correspond to clouds, but the continuous oscillation of the reflectance is due to the BRDF effect.

As an alternative to the BRDF correction, monthly composites were used. These composites allow smoothing the BRDF effect and emphasize at the same time the reflectance drop due to the fire (Figure 7).
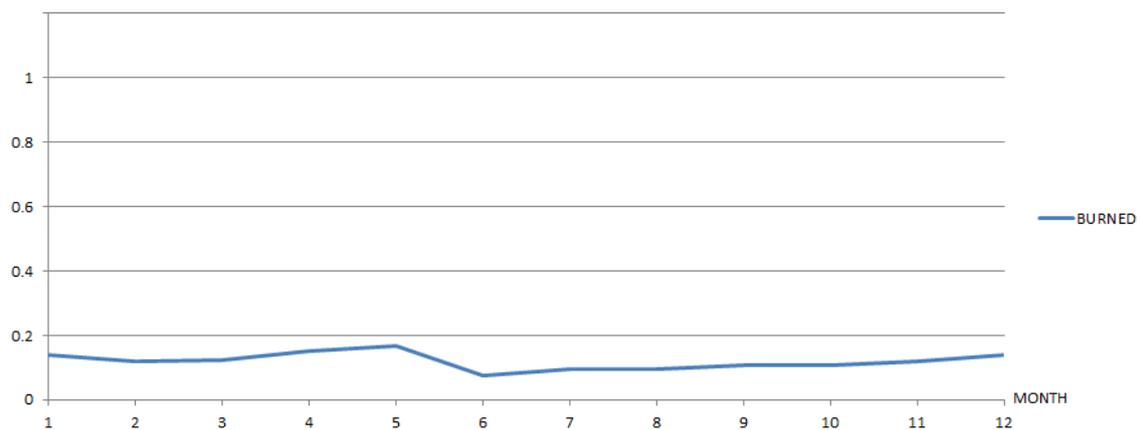
Figure 7: NIR reflectance (Y axis) for the same pixel of Figure 6, as function of the month of the year 2008 (X axis). The smoothing of the BRDF effect due to the composite does not obscure the NIR drop used to detect a burned area (shown in the blue line, for a fire occurred in day 169: 17 June 2008).

2.3.2. Compositing procedure

To create the monthly composites the HS and the daily MOD09GQ images (masked using the MOD09GA state QAs and the burnable mask) were used. It is well known that HS do not provide a full description of fire-affected areas, as satellite sensors only detect those fires that are active when the satellite overpasses the fires. However, the high thermal contrast between burning and background pixels and the sensitivity of MODIS thermal channels, ensures a high confidence in detecting authentic fires, avoiding commission errors. Hantson et al. (2013) performed an exploratory analysis of HS performance to detect burned perimeters by comparing HS with fire reference data extracted from Landsat TM/ETM+. Commission errors found were very low (<3 %) for all study sites, but omission errors (burned patches undetected) were relatively high (>25%) particularly for small BA patches. The HS were used to establish the most appropriate date for the post-fire temporal compositing. The technique of using HS for labelling fire dates has been proposed by other authors (Boschetti et al. 2010).

Since the unit of processing is the standard MODIS tile, the global HS product was cut using the extension of the tile. To account for fires occurring in the boundary between two tiles, a 50-km buffer was added to each tile to avoid boundary effects, and to improve the continuity between tiles. Using the hotspots corresponding to the tile and the month being processed, Voronoi polygons (Iri et al. 2005) were created to assign to each pixel the date of the closest HS. The rationale for this was that the closest the date to a HS, the better sensitivity the algorithm should have to discriminate the burned signal in case a pixel is burned (Figure 8).

(a)	(b)
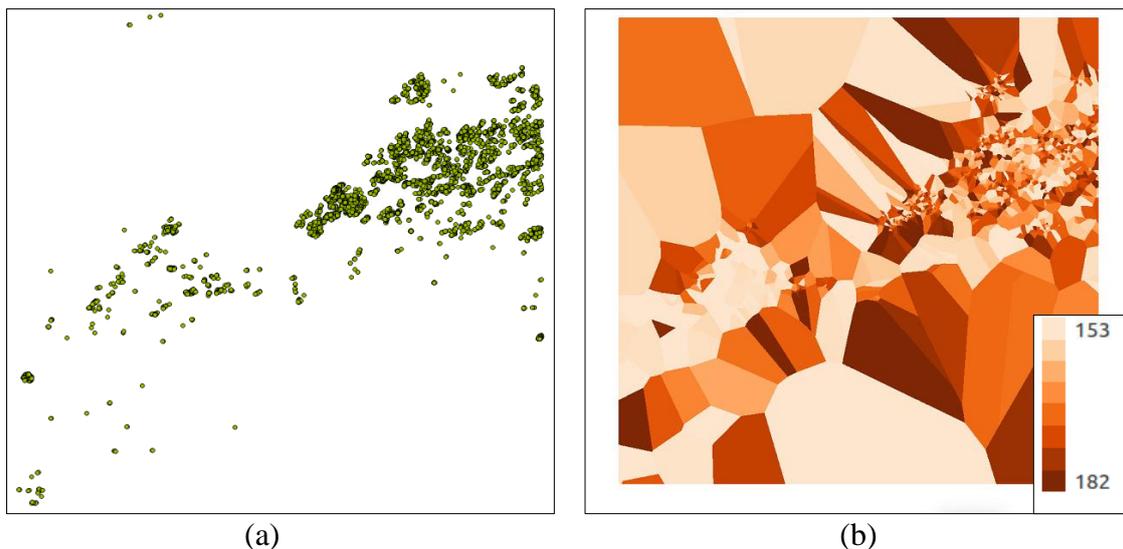
Figure 8: Example of Voronoi polygons created for June 2008 in tile h30v10. (a) Hotspots during that month, including a 50-km buffer. (b) Voronoi polygons with their corresponding date.

The "likely burned date" (LBD) was used to create the monthly NIR reflectance composites by finding the three NIR minima acquired through the month (m). Considering the potential noise in the time series (see next paragraphs), as in the case of the MERIS algorithm, 3 NIR minima were used instead of a single NIR one. Since HS may occur at the end of the month, the monthly composites include also images from the following month to select the 3 NIR minima. Unlike the MERIS algorithm, where a fixed window of 45 days was used (including 15 days from the m+1), in this case the images of the m+1 were used only when LBD was within the last 10 days of each month. In addition, images from the m+1 month were added to the m month to obtain at least 10 post LBD days. In this way, it was assured a minimum number of images to create the monthly composites throughout the month (Figure 9).

The new MODIS algorithm decreases the probability that low NIR values of unburned pixels (floods, shadows, etc.) may be selected, while reducing dating errors potentially caused by selecting images from the following month.

Once the three minimum NIR values were selected, their dates were compared with the date of the HS, and the closest post LBD minimum NIR value was selected. For unburned pixels, the 3 NIR minima may occur before the LBD. In those cases, the second minimum NIR was selected instead the first one, as the latter may be more likely to occur from shadows or other artefacts. The pixels that have no valid observations during the study period are considered not observed (Figure 10).

JUNE										JULY									
153	154	…	172	173	174	175	…	180	181	182	183	184	185	186	187	188	189	190	191	192
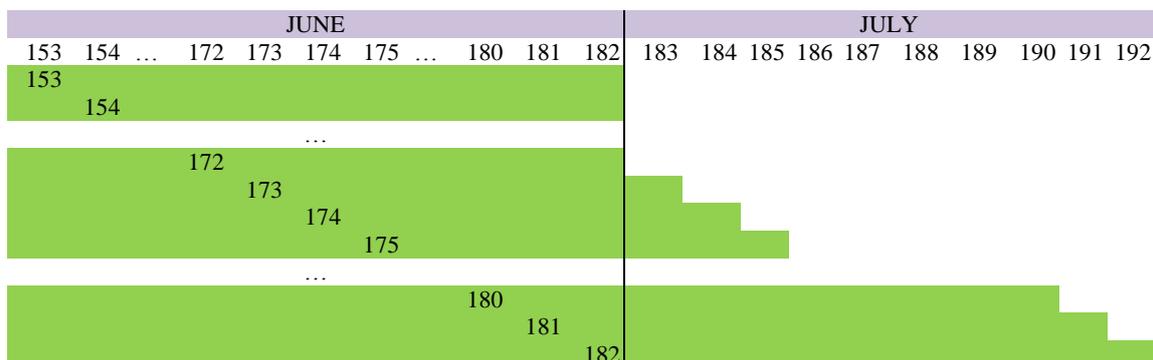
Figure 9: Example of temporal window used for June 2008. The day of the year (DoY) of each month is shown at the top. Each line corresponds to a pixel with a hotspot in the day indicated in the line, and the period marked in green indicates the days when the three minimum NIRs are searched. As shown, until day 172 of (the first 20 days of June 2008) the temporal window is the 30 days of the month of June. After that, the number of days increases, including some days in July, to assure to have at least 10 days posterior to the HS.
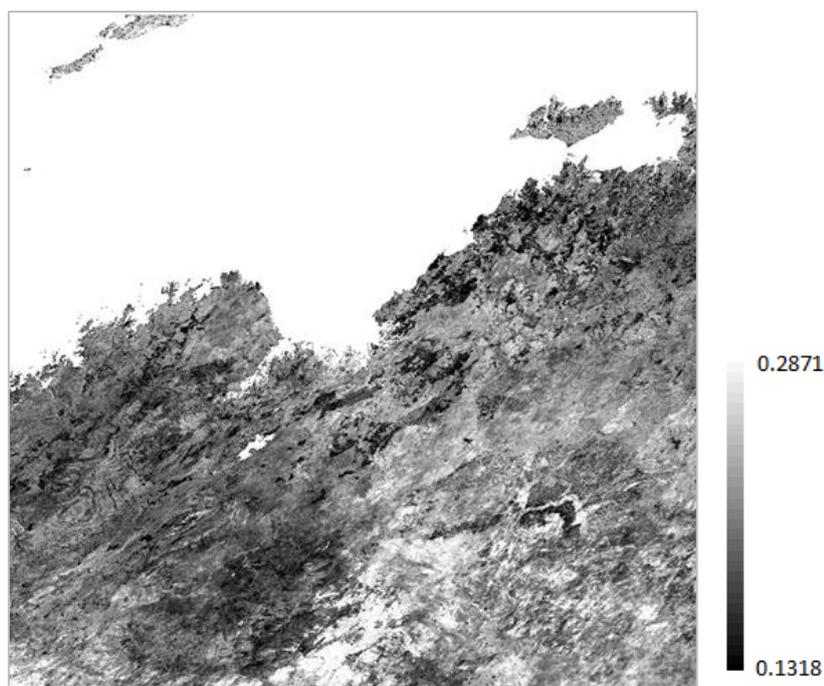
Figure 10: Composite of June 2008 for the tile h30v10 (Australia). This composite has no "*incomplete fires*" or "*eliminate noise*" corrections mentioned in this section.

Some further corrections were applied trying to solve regional errors that were detected in the resulting composites.

- Incomplete fires: it was observed that many burned patches were incomplete, and sometimes the closest posterior date of the NIR with respect to the HS proved not to be the most adequate value (Figure 11). To solve this, the following rules were applied:
  - If the three NIR minima occurred in the 10 days posterior to the HS, the first minimum was selected.
  - If two NIR minima (including always the first minimum) occurred in the first 5 days posterior to the HS, the first minimum was selected.

- <u>Eliminate noise</u>: the resulting composites showed in many cases a small group of pixels with low NIR, which were unlikely to be burned, as they were not associated to HS. To solve this, a compositing rule was added: if the difference of reflectance between the second and third minimum was lower than 0.01, and the difference between the first and second minimum was higher than 0.05, the first minimum was considered to be noise, and the composite used the second minimum NIR of the pixel. These thresholds were empirically obtained based on temporal trends of a set of pixels.

(|Min2 – Min3| < 0.01) & (|Min1 – Min2| > 0.05) →use Min2 for composite

These cases were found most usually due to shadows, where the first NIR minimum was much lower than the second and third ones. With this correction most of these anomalies were eliminated (Figure 12).
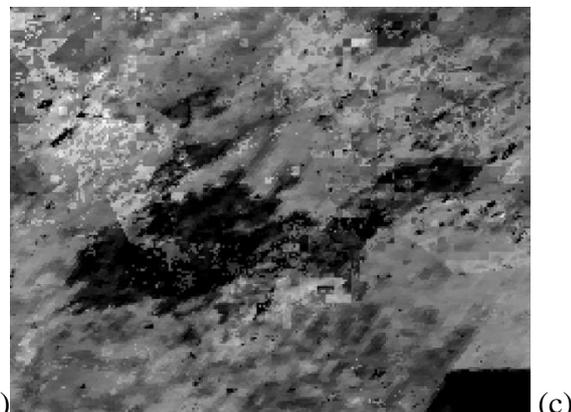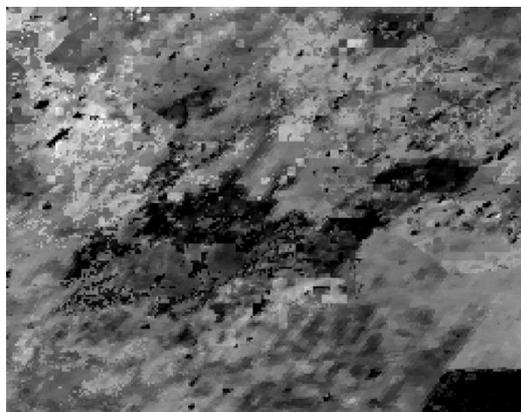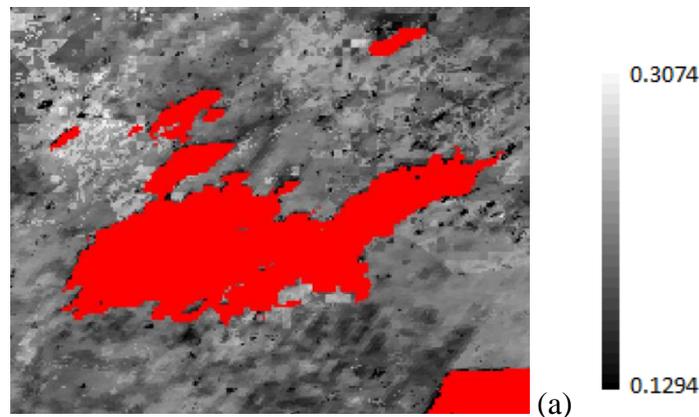
Figure 11: Composite example of November 2008 for tile h30v10. (a) The red areas are the fire perimeters of November 2008 extracted from the North Australian Fire Information (NAFI) database. (b) Original composite without the correction. (c) Corrected composite, where the burned area is more clearly identified.
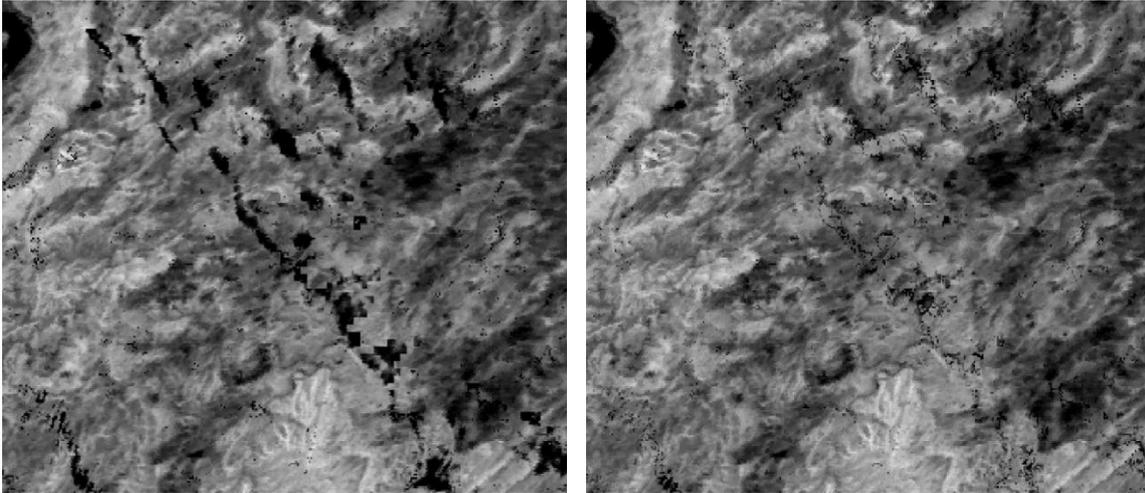
Figure 12: Detail of the composite of June 2008 for the tile h30v10. The left image shows the original composite without the noise correction. The right image shows the composite after the correction. There are no HS in this area during this time period, which shows that the black areas in the left image are due to shadows.

2.3.3. GEMI index

As in the case of the MERIS algorithm, the Global Environmental Monitoring Index (GEMI) was used to help detect burned pixels. This index, along with the NIR reflectance were extensively used for BA mapping in different ecosystems (Barbosa et al, 1999 in Africa; Pereira, 1999 in a comparative study with AVHRR data; Chuvieco et al, 2002 in Mediterranean areas; Martin et al., 2005 in the Mediterranean; Chuvieco et al., 2008 in Boreal forest).

Furthermore, the GEMI index was specifically designed to diminish the influence of atmospheric and soil effects (Pinty and Vestraete 1992). The GEMI expression is:

$$GEMI = \eta * (1 - 0,25\eta) - \frac{\rho_R - 0,125}{1 - \rho_R}$$

Where

$$\eta = \frac{2 * (\rho_{IR}^2 - \rho_R^2) + (1,5 * \rho_{IR}) + (0,5 * \rho_R)}{\rho_{IR} + \rho_R + 0,5}$$

Being ρ_R and ρ_{IR} the reflectivity in the red and infra-red bands, respectively.

When monthly composites were created, they also selected the GEMI value of the same day of the NIR selected to create the composite. In addition, a monthly maximum GEMI index was created (see Figure 13).
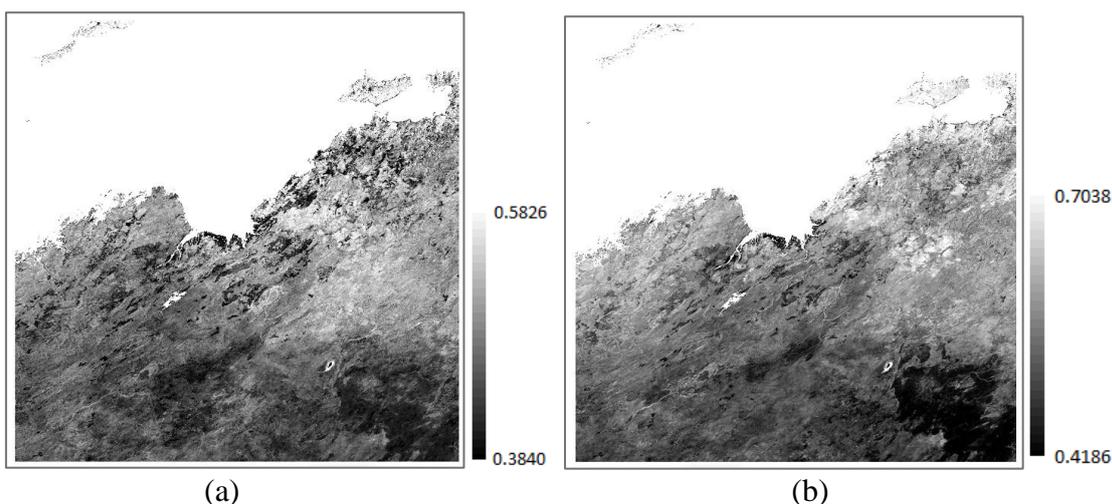
Figure 13: Detail of the month GEMI (a) and month maximum GEMI (b) of June 2008 for the tile h30v10 (Australia).

2.3.4. Voronoi polygon exceptions

There were some cases in which the amount of HS prevented the Voronoi polygon function from performing correctly. For this reason, the following exceptions were included in the algorithm:

- 0 hotspots: when a tile had 0 hotspots in a month, there were two options. One would be to not create a composite for that month and use the composite of the previous month (NIR_{t-1}) for the analysis of the following month. But with this approach the NIR information of the month in study would be lost, which provided useful information for the analysis of the following month. For this reason, the second option was adopted, which was to assign the entire image a HS date of the first day of the month, and calculate the composite accordingly.
- 1 hotspot: all the image was assigned the date of that HS.
- 2 hotspots: in this case there were to possible options: to divide the image in two according to HS location and assign to each zone the date of its HS, or to assign to the whole image the date of the first HS. It was decided to use the second option because only two HS are too few to provide a significant difference between the first and second processing options, and it also saved processing time.
- Geometric anomalies: these are the situations in which the spatial distribution of the HS makes it impossible to create polygons (the HS are in one line, for example). In those cases, the procedure of the case with 0 HS was used.

2.3.5. Improving the Non-burned mask

Generally speaking, pre-fire NIR reflectance should be high (green vegetation), although mid values may also be found due to natural senescence or intense dry periods. When significant low NIR values occurred before the LBD, it indicated that it was not due to a fire event but to other phenomena (shadows, water, wetlands, a previous fire already detected, etc.). For this reason, the unburnable mask was updated using decision rules that were empirically obtained by analysing the NIR minima values and number of observations of a set of pixels related to those events:

- Number of observations > 16, and all NIR minima < 0.1, and at least one NIR minimum prior to the HS date.
- Number of observations > 10, and all NIR minima < 0.07, and at least one NIR minimum prior to the HS date.
- All NIR minima < 0.05, and at least one NIR minimum prior to the HS date.

An example of the need of the non-burned mask is shown in Figure 14.

(a)   (b)
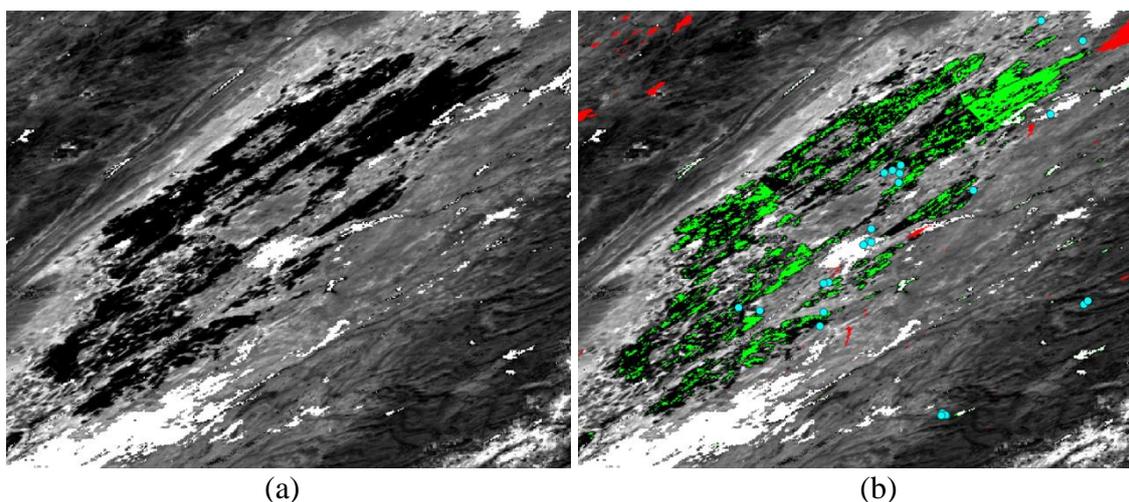
Figure 14: Example of tile h08v05 (California) for May 2008. (a) Composite in a wetland area. (b) The same area with the following complementary information: blue dots corresponding to HS during that month, red polygons corresponding to fires during 2008 according to the California's Fire and Resource Assessment Program (FRAP), green polygons corresponding to the non-burned mask. The improved unburnable mask could detect the wetland so that the HS close to it could be filtered using the criteria that it is explained later (section 2.4.3.).

2.4.   Seed phase

As previously indicated, the seed phase of the algorithm tries to detect those pixels with a high confidence of being burned. The bottom line is to minimize commission errors (avoiding false detections), while keeping low commission rates. Different steps were followed within this phase.

2.4.1.   Positioning of hotspots

Since the MCD14ML HS had 1 km^2 resolution, while the MODIS RNIR product had 250m, a proper method to locate the active fire within each RNIR pixel was needed. For this reason, the HS of the MCD14ML product were positioned within the composite by selecting the lowest NIR value within a 5x5-pixel window around the HS coordinates. (see Figure 15).

(a)   (b)
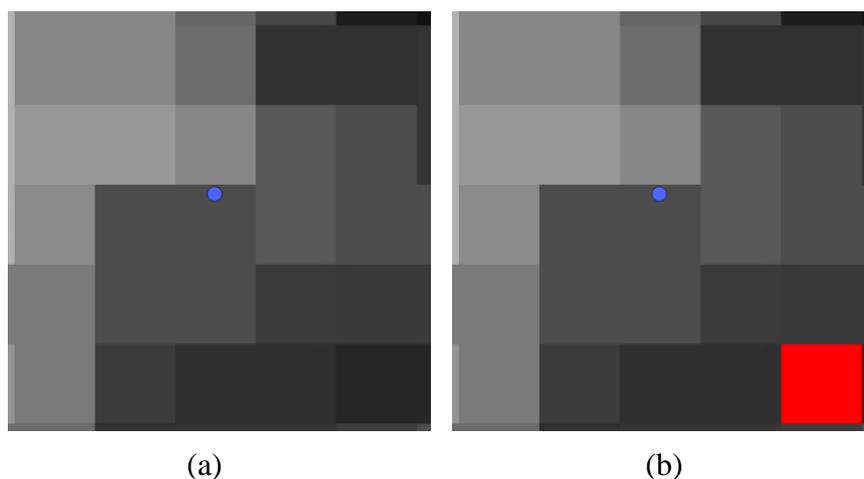
Figure 15: (a) 5x5 pixel window around the pixel where the hotspot is located (blue dot), showing the NIR values in the composite, from higher (brighter) to lower (darker) values. (b) The red pixel is the one assigned to the HS since it has the lowest NIR value of the window.

2.4.2. NIR statistics and potential active fires (PAF) selection

A threshold value was needed to identify which pixels would be classified as seeds. These thresholds were obtained using a Cumulative Distribution Function (CDF), which offered information about the distribution of the burned and unburned class in the study area. The calculation of thresholds was based on deciles. In order to calculate the CDF of the burned and not-burned classes, the most representative pixels of each class were selected.

To select the non-burned pixels used for the threshold statistics calculation, two conditions were simultaneously applied:

- The pixel did not have a HS in a vicinity of 41x41 pixels (≈5 km buffer) when there were less than 15,000 HS in the tile, or 21x21 pixels (≈2.5 km buffer) when the total number of HS was higher.
- The pixel had not been classified as burned in the previous 6 months.

These conditions were similar to the MERIS algorithm (Alonso-Canas and Chuvieco 2015), but with a smaller window, due to the higher number of HS in the MCD14ML Collection 6. But this window could also be too wide in tiles with many HS, leaving the non-burned area underrepresented to calculate the CDF. An example of this situation is shown in Figure 16, where the pixels in cyan show the ones that comply with the conditions, and were used for the CDF calculation. The tile of that example has 35,407 HS, so the 21x21 pixel vicinity was used; if the bigger window had been used, the non-burned pixels would not have enough representation. Using the pixels that comply with the above criteria, the CDF of the NIR reflectance for non-burned pixels was created, and the Growing Threshold (TH_G) was established as 10% of the CDF, using the same criterion of the MERIS algorithm (Figure 17).

For the creation of the burned pixels' CDF, HS information was used, as the pixels with HS have the highest probability of being burned. According to Randerson et al. (2012) the MODIS active fire product can detect fires from 2-6 ha depending on the affected vegetation, so it can detect HS that are sometimes not identified as such in the MODIS 250m surface reflectance product (where a pixel has an area of 6,25 ha). To confirm that a HS is a Potential Active Fire (PAF), the following conditions had to be met:

- There was a drop in the NIR value of the pixel compared to the previous month ($NIR_{t-1} > NIR_t$).
- The NIR value of the pixel was lower than TH_G.
- At least 5 of the 8 neighbouring pixels in a 3x3 window comply with the two previous conditions. This last condition performed a contextual analysis, and confirmed that the tendency showed by the HS was coherent with the neighbouring region.

Using the resulting PAFs, the burned CDF was created.

(a)   (b)
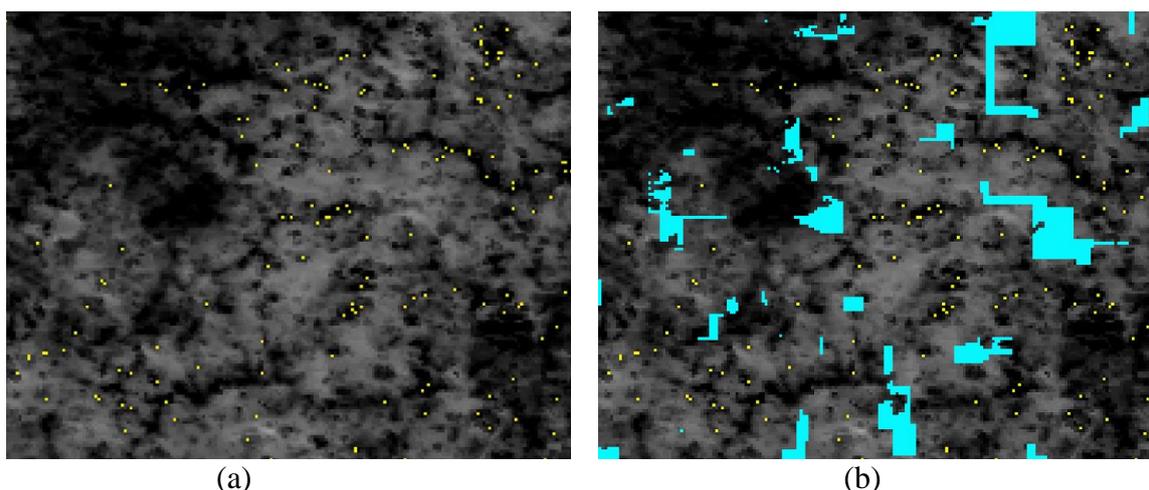
Figure 16: Example of tile h19v10 (Angola) for July 2008. (a) NIR composite showing in yellow the PAFs. (b) The pixels complying with the non-burned criteria, and used for the CDF are shown in cyan.
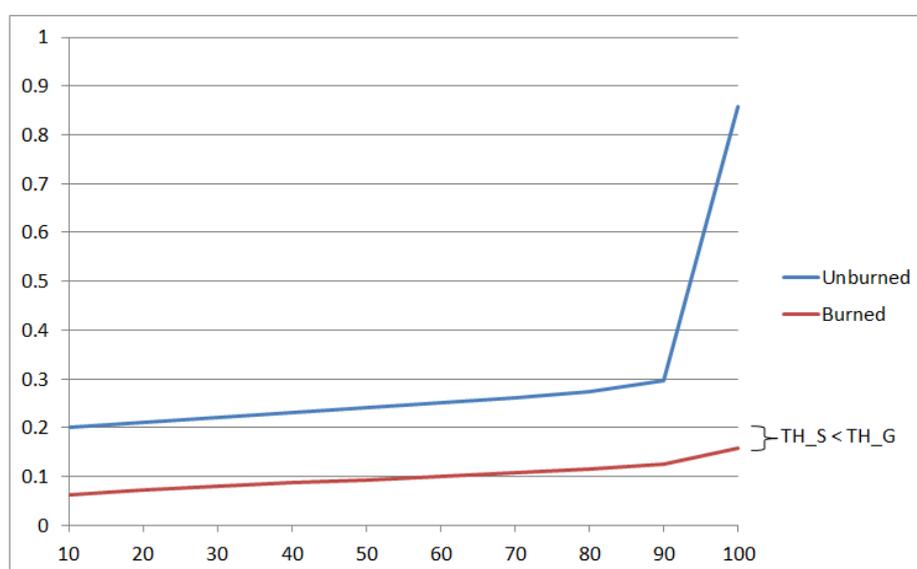
Figure 17: Graph of the CDFs for burned and unburned pixels. The 10% of the unburned is the TH_G, and the 100% of burned is the TH_S. TH_S is always lower than TH_G. The sudden increase of the last 10% of the unburned CDF represents anomalies of the composite, such as clouds not masked, snow, etc.

2.4.3. Filtering of PAFs

After the PAF selection, a contextual analysis around them was performed, to filter those which could be related to other phenomena different from fires. The non-burned

mask (Section 2.3.5) and the density of HS were used for the filtering. For a PAF to be filtered and discarded the two following conditions had to be met:

- The HS density was lower than 10 around a 41x41 window. This was based on the observation that areas with low NIR values but not associated with fires usually have a low HS density (see Figure 18).
- The proportion of non-burned pixels in a 41x41 window around the PAF was higher than 5%. The non-burned pixels are the combination of the masks from the 6 previous months (including the month being analysed), as the masking process was very restrictive.

This filtering process discarded a small number of PAFs, which had no significant impact on the burned CDF, but they did on the final result.
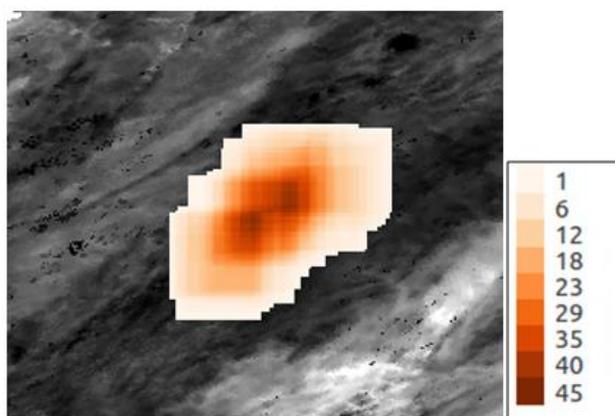
Figure 18: Example of tile h08v05 (California) for May 2008, showing the HS density. If a pixel has a value of 35 it means that there are 35 PAFs in a 41x41 window around it.

2.4.4. Seed identification

Once the burned CDF was calculated, two thresholds were used to classify a pixel as a burned seed. The Seed Threshold (TH_S) was defined as 100% of the burned CDF. As TH_G was used to filter the HS, TH_S will always be lower than TH_G, which ensures that all PAFs are converted to seeds and that only the pixels with NIR values equal or smaller than it are classified as seeds.

The conditions for a pixel to be classified as a seed were (Figure 19):

- There is a drop in the NIR value of the pixel compared to the previous month ($NIR_{t-1} > NIR_t$).
- The NIR value of the pixel is lower than TH_S.
- At least one of the neighbour pixels in a 3x3 window is a PAF.

With this last condition, the seed selection is kept within the limits of the original HS resolution (1000x1000 m).
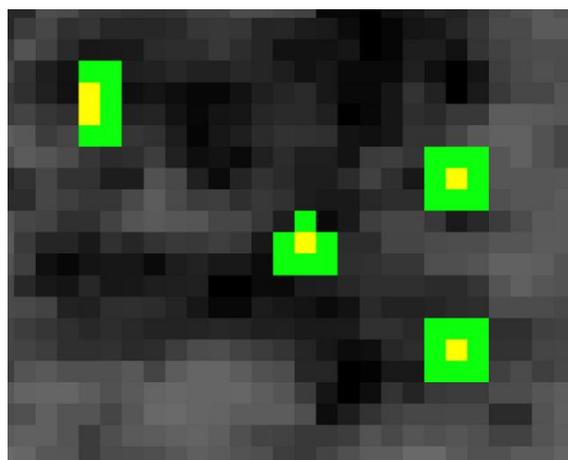
Figure 19: Example of seeds. The PAFs are shown in yellow, and the rest of the seeds in green.

2.5.   Growing phase

The growing phase of the algorithm aimed to improve the delimitation of BA patches from the previously detected seeds. It is based on applying contextual algorithms around seed pixels, as BA has lower NIR values than the surroundings unburned pixels (Figure 20). Two criteria were selected for the growing phase. The NIR values were used for the burned patch core, while the less affected zones were discriminated using both the NIR and the GEMI difference.
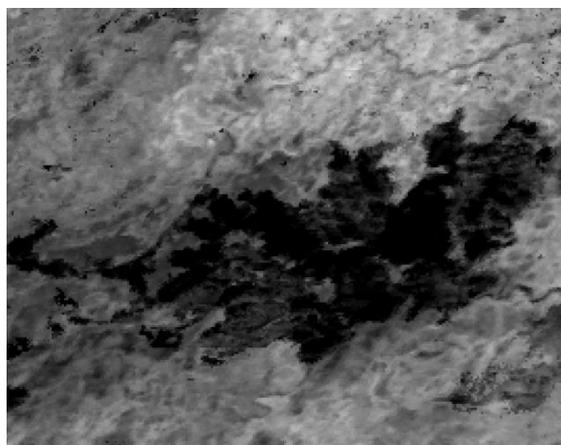
Figure 20: Example of tile h30v10 (Australia) for April 2008. NIR composite where the black areas are those considered as "centre" of the fire, while the dark grey are the zones of low intensity.

2.5.1. Growing threshold using NIR

To perform the growth in the centre of the fire a new threshold called Burned Threshold (TH_B) was applied, which established the limit between the centre of the fire and the zones with low intensity (Figure 21). This threshold needed to be lower than TH_S. The analysis over the calibration study areas (Australia, Canada and California, which include vegetation in tropical savannas, boreal forests and Mediterranean woodlands) determined that this TH_B threshold should be the CDF decile that complied with the following criteria:

- The NIR value of the decile was lower than 0.16 (empirical value resulting from the analysis), and
- The decile corresponded to 90% of the CDF or lower.
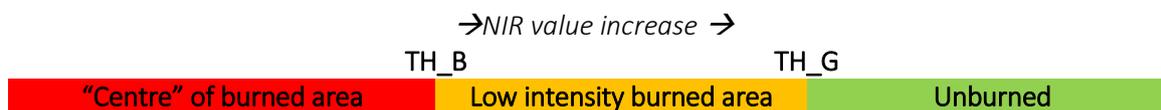
Figure 21: Graph of the classification of pixels as function of their NIR values

2.5.2. Growing threshold using GEMI difference

The other variable used for the growing phase was the GEMI index (Section 2.3.3), calculating a per-pixel difGEMI value as:

$$difGEMI = maximum\ monthly\ GEMI_{m-1} - monthly\ GEMI_m$$

being m the month being analysed. The value of maximum monthly $GEMI_{m-1}$ represents the pre-fire dates that had the maximum vegetation greenness, while the monthly $GEMI_m$ shows the current greenness. The difference should be quite high for burned pixels. Hence, the higher the value of difGEMI, the higher the loss of greenness (Figure 22).

As in the case of the NIR reflectance, CDFs of burned and unburned difGEMI were obtained, and the pixels that best represent each class were selected. In the case of the burned CDF, as it was necessary to find a decrease in vegetation greenness, all seeds with a positive difGEMI were used. In the case of the non-burned CDF, the following requisites were necessary to include a pixel as unburned:

- it should be one of the pixels used for the non-burned NIR CDF,
- it should have a positive difGEMI, and
- its NIR value should be higher than TH_G.

With the distribution function obtained from the data above, the GEMI threshold was calculated as:

$$TH\_GEMI = \frac{(10\%CDF\_B + 90\%CDF\_NB)}{2}$$

where 10%CDF_B is the value of the first burned difGEMI CDF decile, and 90%CDF_NB is the ninth decile of the non-burned difGEMI CDF. An example of the values of the CDF is shown in Table 4. This TH_GEMI threshold is used in the areas identified as "low intensity burned areas" to confirm that the pixel is burned.

Table 4: Example of CDF decile values for the tile h10v08 (Colombia) of January 2008. In green is the 10%CDF_B and in orange the 90%CDF_NB.

difGEMI	10	20	30	40	50	60	70	80	90	100
Burned	0.16587	0.20657	0.23601	0.25912	0.28011	0.30131	0.32244	0.34472	0.37535	0.62897
Unburned	0.04696	0.08062	0.10619	0.12618	0.14147	0.15500	0.16908	0.18559	0.21102	6.75031
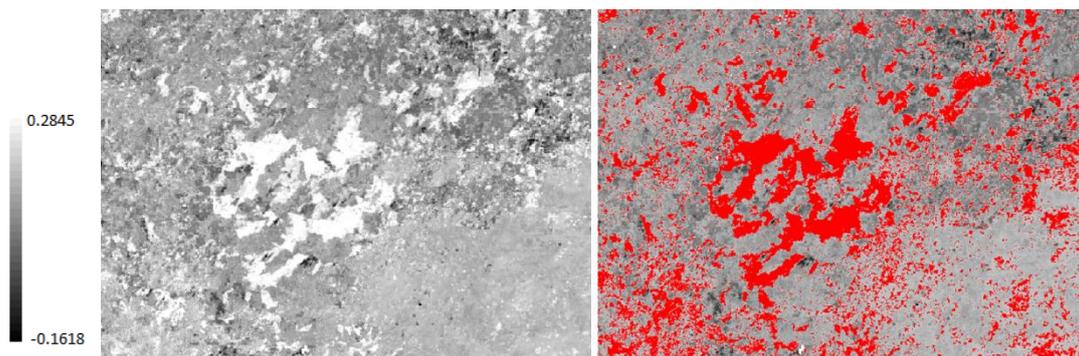
Figure 22: Example of tile h10v08 (Colombia) for January 2008, showing difGEMI in an area of the Orinoco Basin with fires (left). A negative value shows that the greenness of the pixel has increased, while a positive value indicates a loss of greenness. In the right image, the pixels in red are the ones >TH_Gemi = 0.10845

2.5.3. Growing method

The growing is performed using contextual growth, starting from the seeds and classifying the neighbouring pixels. In this way, a pixel can be classified as burned only if a neighbour pixel is either a seed or a pixel previously classified as burned. The growth is done in branches, starting with the seeds and growing in the same row to the east and west until the next pixel doesn't comply with the burned criteria. In the next iteration the growth is done to the north and south from those burned pixels, using the same criteria. In this way, the iterations are performed alternating east-west and north-south until the whole burned patch is identified. An example is shown in Figure 23.

(a)   (b)
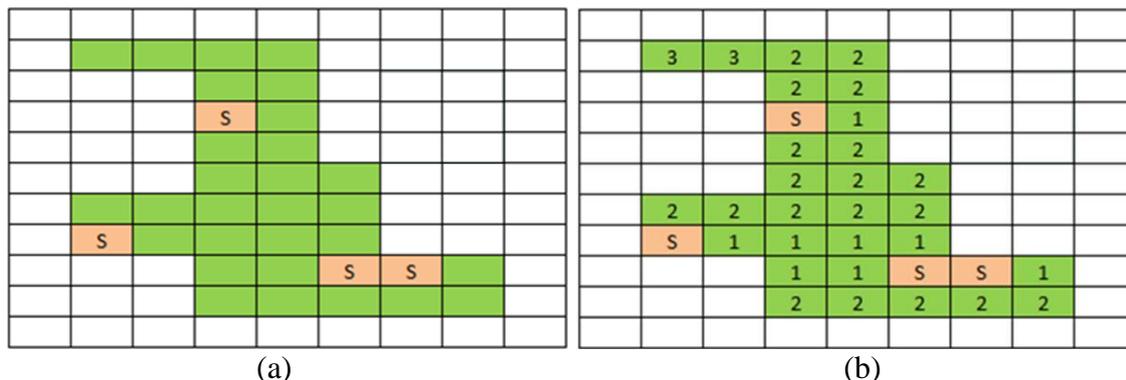
Figure 23: Example of the growing procedure. (a) The orange pixels are the seeds, and the green pixels correspond to the burned area to be detected. (b) The green pixels show the iteration in which they were detected.

A limit to the growth was established at 20 km (81x81 pixel window) around each PAF. This distance is the same applied by Giglio et al. (2009) in the MCD64 algorithm. A more restrictive limit was applied for forest fires, as several studies show that the density of HS in forest fires is higher than in other land covers (Giglio et al. 2006, Hantson et al. 2013). Seeds corresponding to forest fires were identified as those having more than 60% of pixels classified as class 3 (Section 2.2.4) in a 41x41 window (10 km). In these cases, a growing limit of 31x31 pixels was applied, stablished based on the performance of the algorithm in the forest calibration areas.

The growing phase classifies a pixel as burned when the following criteria are met:

- the NIR value was lower than TH_G,
- if the NIR value was higher than TH_B, its difGEMI is higher than TH_GEMI.

- at least a neighbouring pixel was a seed or has been classified as burned in a previous iteration, and
- it had a NIR drop compared to the composite of the previous month ($NIR_{t-1} > NIR_t$).

2.5.4. Morphological filter

To improve the results of the growing phase, a morphological filter was applied, which performed two operations: erosion to contract the burned patch, and dilation to expand it[1]. Two phases were applied in the algorithm: first it performs an erosion followed by a dilation to eliminate the small burned patches. Then it performed a dilation followed by an erosion to fill the patches (Figure 24).

| (a) | (b) |
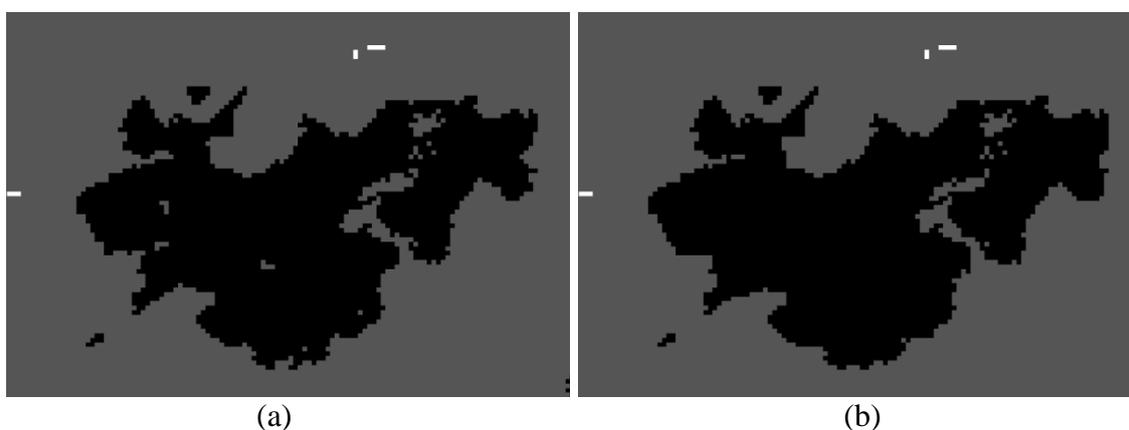
Figure 24: Example of the application of the morphological filter to tile h30v10 (Australia) in June 2008. (a) Burned patch without the filter. (b) Burned patch after the filter. The black pixels are burned, the grey are not burned, and the white are non-burnable.

2.6. Uncertainty

The Fire_cci v5.0 computed the uncertainty of the burned classification of both the burned and unburned pixels. This uncertainty was calculated using four variables related to the probability of each pixel of being burned or not.

The first variable was the number of valid observations available to create the composite, which was calculated during the composite creation (Figure 25). The rationale behind this variables was that the higher the number of observations, the lower the uncertainty in the observation. This variable had values between 0 and 30.

---

[1] See https://docs.opencv.org/trunk/d9/d61/tutorial_py_morphological_ops.html, accessed March 2018.
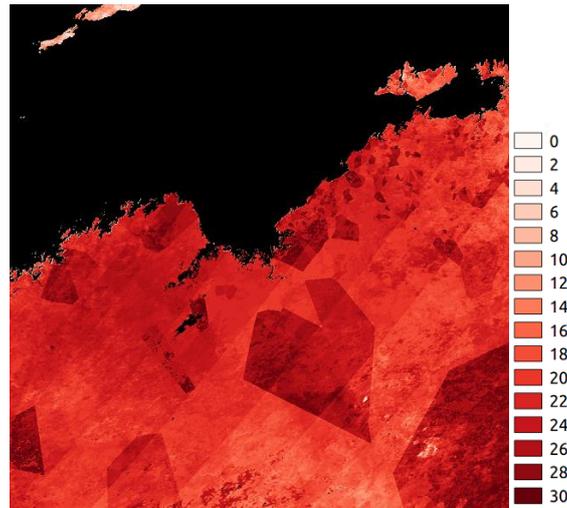
Figure 25: Number of valid observations of the tile h30v10 (Australia) for June 2008.

The second variable was the NIR reflectance, particularly the position within the burned and unburned CDFs (Figure 26). In this case, the lower the NIR value, the higher the probability of the pixel to be burned, and the lower the uncertainty in the burned classification. The variable had values between 0 (highest NIR value and lowest probability of burn) and 19 (lowest NIR value and highest probability of burn).
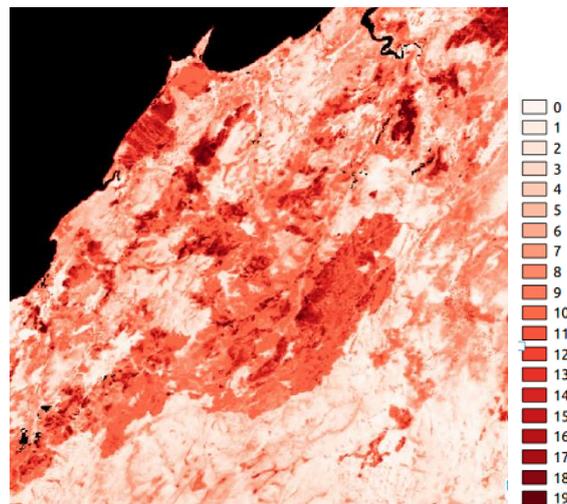
Figure 26: Probability of burn value related to the NIR reflectances of the tile h30v10 (Australia) for June 2008. The higher the value, the higher the probability of the pixel being burned. The meaning of these classes is explained in **Figure 27**.
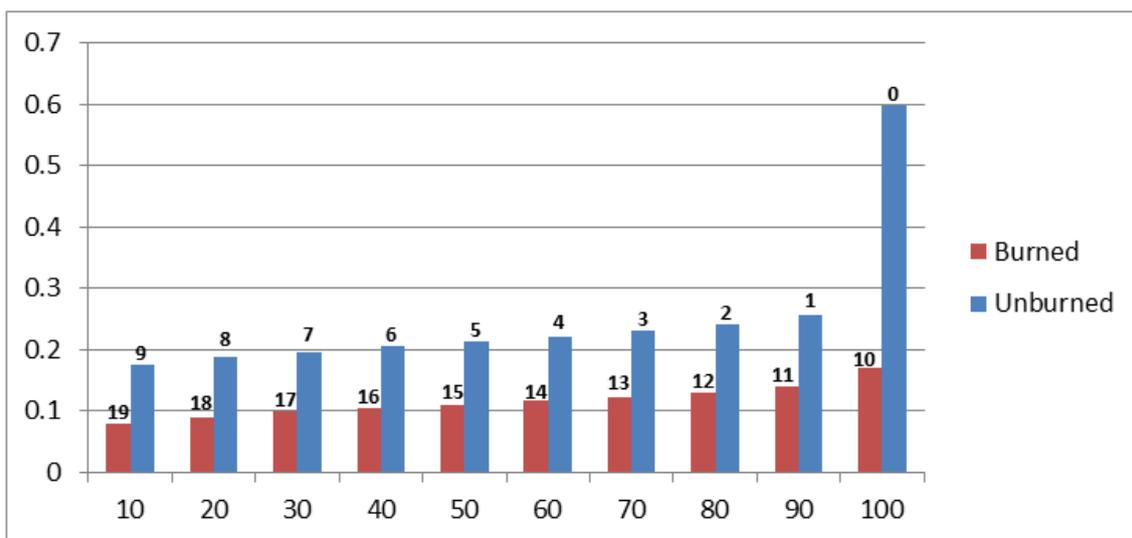
Figure 27: NIR CDF deciles of the image of Figure 26, with the assignation of the values of burn probability above the bars. The lower the NIR value of the decile, the higher the probability of the pixel being burned.

The third variable was similar to the second, but considering the difGEMI value, which had a behaviour inverse to the NIR (Figure 28): the higher the difGEMI, the higher the probability of the pixel being burned. The values also ranged between 0 and 19, where 19 was assigned to a pixel with difGEMI higher than the tenth burned decile, taking into account that only the seeds were used to calculate the deciles, and so other pixels not classified as seeds but with higher difGEMI may appear in the image (Figure 29).
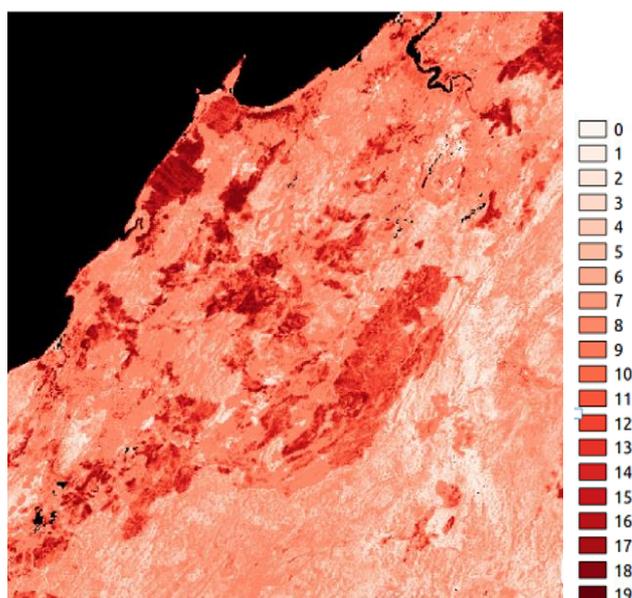
Figure 28: Probability of burn value related to the difGEMI values of the tile h30v10 (Australia) for June 2008. The higher the value, the higher the probability of the pixel being burned.
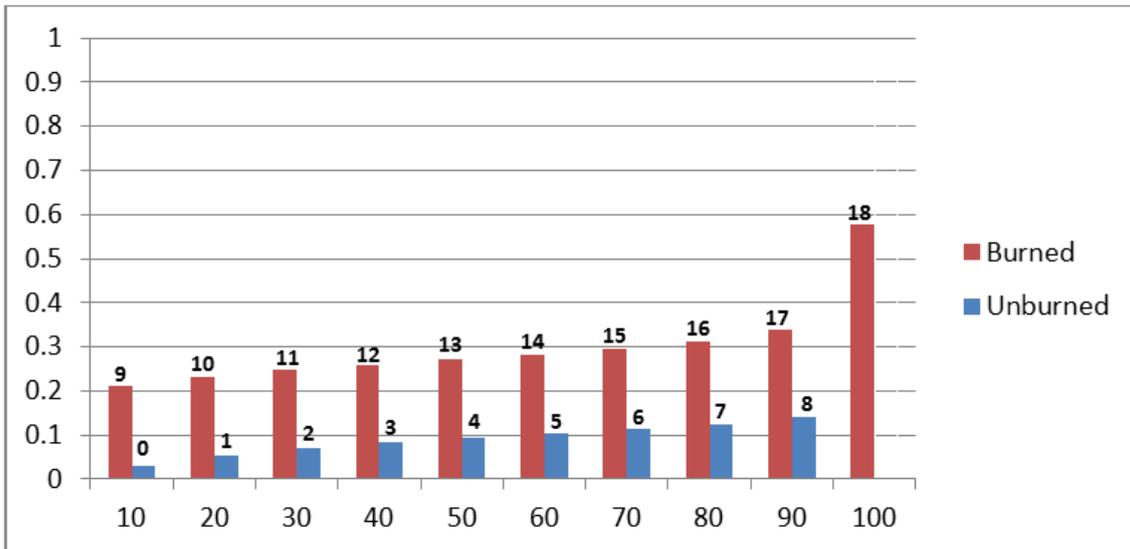
Figure 29: difGEMI CDF deciles of the image of Figure 28, with the assignation of the values of burn probability above the bars. The higher the difGEMI value of the decile, the higher the probability of the pixel being burned. A value of 19 is assigned to pixels with difGEMI higher than the 100% burned seeds.

The last variable used to calculate the uncertainty was the distance to the nearest PAF. The value was calculated after the burned areas were obtained, using a two-phase process (Figure 30). First, the distance within the burned areas was calculated, assigning the value of 240 (an arbitrary number higher than the maximum possible iterations o distance number) to the PAFs, and calculating the distances from the PAF decreasing by one unit each pixel farther from it. In the second phase, and starting from the last value obtained from the burned pixels, the values decreased from the border of the fire to the outside, completing 20 iterations (about 5 km). In this way, the first iteration would always be 240 but the last iteration number would depend on the number of iterations needed to fill all the burned patches and the rest 5 km. This method decreased significantly the computation cost of this variable. An example of the result is shown in Figure 31.
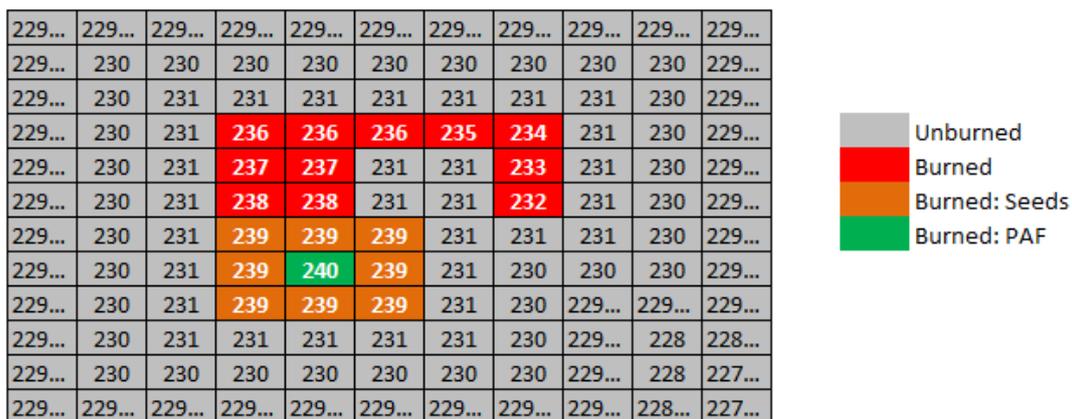
Figure 30: Example of computation of PAF distances.
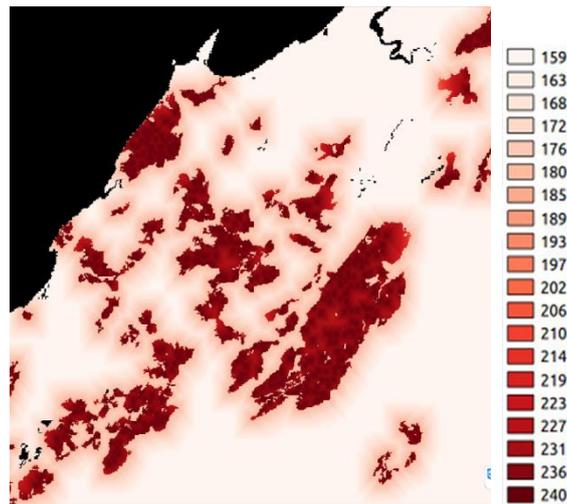
Figure 31: Probability of burn value related to the distance to the PAF of the tile h30v10 (Australia) for June 2008. The higher the value, the higher the probability of the pixel being burned.

Once each variable was calculated, they were normalized using the following formula:

$$V_{nor} = \frac{V - V_{min}}{V_{max} - V_{min}}$$

Vmax – Vmin show the range of values of each variable. The number of observations will always range between 0 and 30, the NIR and difGEMI values will range between 0 and 19, and in the case of the distance to the PAFs the maximum value will always be 240, but the minimum will vary according to the value of the last iteration.

As a last step, the final uncertainty was obtained as the mean value of the four normalized variables (Figure 32), and multiplying the value by 100 to obtain percentages. As such, the uncertainty is expressed by a value between 0 and 100 representing the probability of each pixel to be burned.
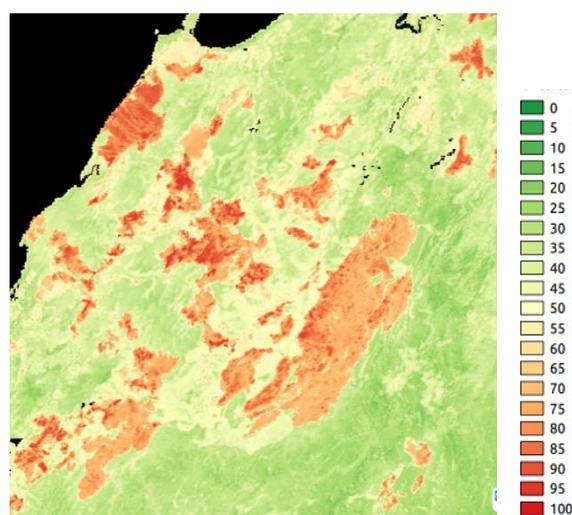
Figure 32: Uncertainty of the tile h30v10 (Australia) for June 2008. The values represent the probability of each pixel being burned.

2.7.   Date of first detection

To assign to each burned pixel the date of the first fire detection (also called day of the year – DoY, or Julian Day – JD), the date of the composite was used, because the algorithm was based on the NIR reflectances of the MODIS composite. When there was a good coverage of the area, this date differed only slightly from the date of the hotspot, but in areas with a low number of valid observations, the difference between the dates could exceed 20 days. An example is shown in Figure 33.
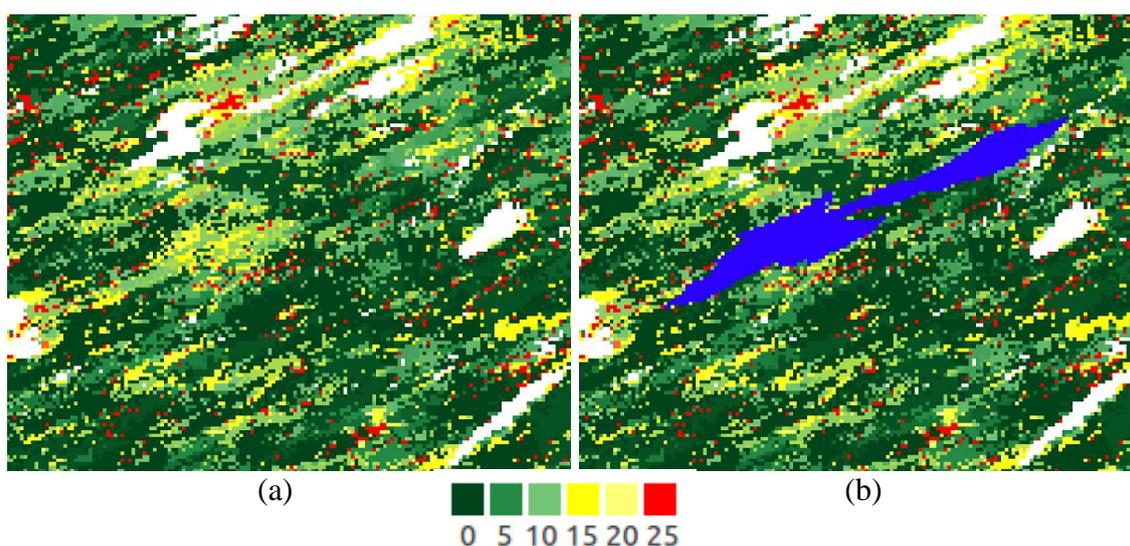
Figure 33: (a) Temporal precision of the composite, calculated as the difference in days between the day of each pixel of the composite and the closest HS. If a pixel has a value of 10 it could mean that the day of the composite is 10 days before or after the HS. The image corresponds to the tile h11v03 (Canada) for June 2008. (b) shows in blue a fire in that area and time period.

3.   Formatting MODIS-based BA data to PSD-compliant products

This piece of software was designed to obtain the final pixel and grid products following the specifications of the Product Specification Document (PSD, Chuvieco et al., 2017).

3.1.   Pixel product

3.1.1. Binning

The pixel products are stitched products which consist of the result of the BA algorithm described in the previous section. According to the PSD, the pixel products are created as subsets of 6 zones: North and Central America, South America, Europe, Africa, Asia, and Australia. They have the spatial resolution of the MODIS data (approx. 250m), but the coordinate system differ, so the pixel values cannot be directly copied from source to target, as it was done for the MERIS product. Instead, the pixel products were created using the binning technique, which has been defined by NASA and is described in Hooker et al. (1995). This technique is a fundamental part of the Calvalus system, and allows the aggregation of multiple input images spatially and temporally, and thus creates aggregated (so-called Level 3 - L3) maps.

Basically, the pixel products are composites computed by temporally aggregating BA pixels into spatial equal-area bins (binning). The bin cells in a pixel product are arranged in an integerised, sinusoidal grid (ISIN) which is compatible to the one used for ESA MERIS L3 products and the MODIS L3 products generated by the NASA Ocean Biology Processing Group (OBPG). After the aggregation is done, the data is re-projected onto a configurable target grid – in the Fire_cci case, this is a geographic lat/lon grid of 0.0022457331 degrees (approximately 250 m at the Equator), as defined in the PSD. The complete description of the L3 generation methodology can be found in Hooker et al. (1995). Figure 34 depicts the processing scenario.
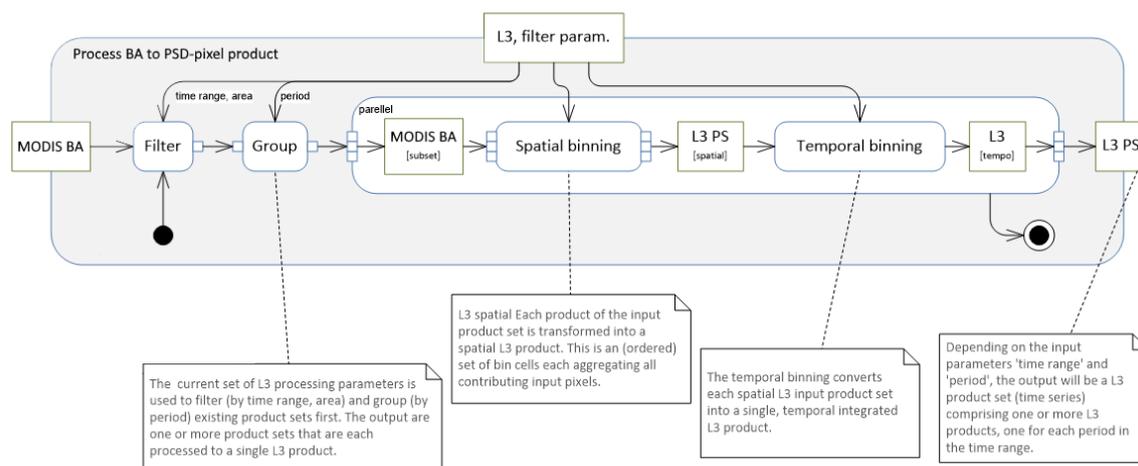
Figure 34: Calvalus L3 processing

It is noteworthy that the spatial binning and the temporal binning employ a specific piece of software, the so-called aggregator. Basically, an aggregator provides the strategies for spatial and temporal binning. Operating on single bin cells, an aggregator provides the answers for:

- Spatial binning: how are input samples of a single observation (swath) aggregated to spatial bins?
- Temporal binning: how are spatial bins aggregated to temporal bins?
- Final statistics: how are final statistics computed?

3.1.2. The ModisJDAggregator

There are multiple built-in aggregators in Calvalus, such as an aggregator that computes a mean value of the input data, or another one that computes the min and max values of the input data. For the aggregation of the MODIS BA data, a dedicated aggregator has been newly developed, the "ModisJDAggregator". This aggregator produces two of the three image layers the pixel products consist of: JD (the day of the year of first detection of the fire) and CL (the level of confidence of the observation). The third layer, LC (which is the land cover class, according to the PSD), is computed in a dedicated finalisation step which is described in Section 3.1.3.

The ModisJDAggregator does only a spatial aggregation, as the processing is configured to use only those BA products from the month which is aggregated; so if the processing task is to create the pixel product for January 2015, only input data from January 2015 is considered. The spatial aggregation is done as shown in the pseudo-code below:

```
validJdSet = previousJDValue >= 0
             && previousJDValue >= first day of month
             && previousJDValue <= last day of month;
```

A valid JD value has already been set if it is positive, and within the considered month.

```
inTimeBounds = true if jd >= minDoy and jd <= maxDoy
```

It is checked if the JD value from the current observation is within the considered month, in order to prepare the check if the value should be preferred to the previous value. This only happens when multiple pixels contribute to a single bin cell, and is done according to the following rule set:

```
preferToPreviousValue = true if:
    no validJdSet or the new value is earlier than old value
    and the new value is >= 0
    and the new value is within the month
```

If the value should be preferred over the previously set value, it will be set to the target bin cell as well as the CL value:

```
if preferToPreviousValue:
    set new JD to current JD value
    set new CL to current CL value
```

3.1.3. Finalisation

In order to create the LC image layer and the metadata, a finalisation step is run after the binning step. This step takes as input the result of the binning step. For each pixel, the process checks if the pixel is in a burnable class and does the respective re-mapping: if the pixel is reported to have burned, but the respective LC class of that pixel is not burnable, the JD value of the pixel is changed to 0 (not burned). The same is done for the confidence level: if the pixel is reported to have burned, but the respective LC class of that pixel is not burnable, the CL value of the pixel is changed to 0. These cases could happen due to the re-projection from the sinusoidal projection of the BA algorithm to the geographical coordinates of the PSD-compliant product. Also, the LC class is set to each burned pixel; in other pixels, it is set to 0.

The result files are written to disk; it is configurable if all image layers (JD, CL, and LC) shall be written into a single GeoTIFF-file, or if three independent files shall be written (as has been done for v5.0). Finally, the metadata is produced by simple pattern replacement of a template XML, and also written to disk.

3.2. Grid product

The grid products are global products which aggregate the result of the BA algorithm described in the previous section. They are mapped onto a 0.25 degree spatial resolution grid. Hence, the pixels in the result of the BA algorithm cannot simply be copied, but need some aggregation. This aggregation is performed from the outputs of the BA algorithm in sinusoidal projection.

In order to find the correct input pixels for each grid cell in a performant way, dedicated lookup-tables have been generated. These lookup-tables provide the positions of the input pixels for all MODIS tiles and for each grid cell, which are constant for the whole

time series. Without these lookup-tables, it would have been necessary to compare the geo-position of each pixel in the BA results with the geo-boundaries of each grid cell, which would have been a very costly operation. Instead, this has been only done once, precisely for the generation of the lookup-tables.

The lookup-tables are provided as JSON files and have the following structure:

```
{"h18v02":["3349,983","3349,982","3349,985","3374,1069","3349,
984","3374,1068","3349,981", ...],
"h19v02":["3149,3548","1568,1641", ...], ...}
```

The target grid cell is given by the file name; the example above is an excerpt from the file "`modis-geo-lut-794-88.json`".

These lookup-tables have been generated by the following algorithm, given in pseudo-code:

```
for each target grid cell c, do:
    result file = open new json file for writing
    for each MODIS tile t, do:
        for each pixel p in t, do:
            g = geo-position g of p;
            if g is inside bounds of t:
                write p into result file
```

The layers created for the grid product are: the sum of BA, the standard error of the burned area, the fraction of burnable area, the fraction of the observed area, the number of patches, and the sum of burned area of each land cover class. In the following sections, it is explained how these layers are derived from the BA algorithm result and from LC data.

3.2.1. Sum of burned area

The sum of burned area is given in m², and computed by identifying for each target grid cell the pixels of the source grid.

```
for each grid cell c in the target grid, do:
    find the respective lookup table l
    for each tile t and source pixel sp in l, do:
        fetch the BA product belonging to t
        if sp is burned, in time bounds & in valid LC class:
            burned_area(c) = burned_area (c) + area(sp)
```

Note that `area(sp)` is a constant value (~53664 m²), as the MODIS grid is an equal-area grid.

3.2.2. Standard error

The standard error also is given in unit m². It is computed after the burned area has been identified, as an aggregation of the confidence level of each source pixel. The basis of the aggregation is explained in Annex 2.

```
for each grid cell c in the target grid, do:
    find the respective lookup table l
    for each tile t and source pixel sp in l, do:
        fetch the BA product belonging to t
        probabilityOfBurnFirstHalf(c) = confidence_level(sp)
    initialise sum_pb = 0.0
    for each probability p, do:
        if p is valid:
            sum_pb += p;
    S = numberOfBurnedPixels / sum_pb
    initialise array pb_i_star
    for each probability pb_i, do:
        pb_i_star[i] = pb_i * S;
    initialise var_c = 0.0
    initialise count = 0
    for each scaled probability pb_s in pb_i_star, do:
        var_c = var_c + pb_s * (1 - pb_s);
        if pb_s is valid:
            count = count + 1
    se(c) = sqrt(var_c * (count / (count - 1.0))) * area;
```

Note: confidence_level = the value of the CL variable in the BA product, and area = ~53664 m².

3.2.3. Fraction of burnable area

The fraction of burnable area is given as a unitless value between 0 and 1, where a cell value of 1 indicates that the whole area covered by the cell consists of burnable pixels, as indicated by LC_cci, while a value of 0 indicates that no source pixels of the area covered by the cell are burnable.

```
for each grid cell c in the target grid, do:
    initialise burnable_temp = 0
    find the respective lookup table l
    for each tile t and source pixel sp in l, do:
        fetch the LC product p belonging to t
        if p(sp) is burnable:
            burnable_temp = burnable _temp + area(sp)
    burnable(c) = burnable_temp / area(c)
```

Note that here, `area(c)` denotes the area of the target grid cell.

3.2.4. Fraction of observed area

The fraction of observed area is given as a unitless value between 0 and 1, where a cell value of 1 indicates that the whole burnable area covered by the cell has been observed, while a value of 0 indicates that no source pixels of the burnable area covered by the cell have been observed.

```
for each grid cell c in the target grid, do:
    initialise OAF_temp = 0
    find the respective lookup table l
    for each tile t and source pixel sp in l, do:
        fetch the BA product p belonging to t
        if p(ps) has been observed and sp is burnable:
            OAF_temp = OAF_temp + area(sp)
    OAF(c) = OAF_temp / burnable(c)
```

Note that `burnable(c)` denotes the burnable area of the target grid cell, as calculated in Section 3.2.3.

3.2.5. Number of patches

The number of patches provides for each grid cell the count of different patches within the cell. A patch is a contiguous group of burned pixels. The algorithm to find the patches is a modified standard graph traversing algorithm called depth-first-search[2].

```
for each grid cell c in the target grid, do:
    find the respective lookup table l
    for each tile t and source pixel sp in l, do:
    run depth-first-search
        if burned pixel is encountered:
            mark pixel as burned
            increment count of patches
    patch_number(c) = count of patches
```

3.2.6. Sum of BA of each LC class

The sum of BA in each land cover class allows discriminating the different land covers affected by the fire in each grid cell. For each LC class, it is computed as follows:

```
for each LC class lc, do:
    for each grid cell c in the target grid, do:
        find the respective lookup table l
        for each tile t and source pixel sp in l, do:
            if sp is burned and corresponds to lc:
                BA (lc, sp) = BA(lc, sp) + area
```

---

[2] See https://en.wikipedia.org/wiki/Depth–first_search, accessed March 2018.

Note: area = ~53664 m².

4. References

Alonso-Canas, I., & Chuvieco, E. (2015). Global Burned Area Mapping from ENVISAT-MERIS data Remote Sensing of Environment, Remote Sensing of Environment, 163, 140-152. http://dx.doi.org/10.1016/j.rse.2015.03.011.

Alonso-Canas I., Chuvieco E., Kirches G., Storm T. (2016) ESA CCI ECV Fire Disturbance: Algorithm Theoretical Basis Document-MERIS, version 1.1. Available at: http://www.esa-fire-cci.org/documents.

Barbosa, P.M., Stroppiana, D., Gregoire, J.M., & Pereira, J.M.C. (1999). An assessment of vegetation fire in Africa (1981-1991): Burned areas, burned biomass, and atmospheric emissions. Global Biogeochemical Cycles, 13, 933-950.

Boschetti, L., Roy, D.P., Justice, C.O., & Giglio, L. (2010). Global assessment of the temporal reporting accuracy and precision of the MODIS burned area product. International Journal of Wildland Fire, 19, 705-709.

Chuvieco, 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.

Chuvieco, E., Englefield, P., Trishchenko, A.P., & Luo, Y. (2008). Generation of long time series of burn area maps of the boreal forest from NOAA–AVHRR composite data. Remote Sensing of Environment, vol. 112, 2381-2396.

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.

Chuvieco, E., Pettinari, M.L., Heil, A., Storm, T. (2017) ESA CCI ECV Fire Disturbance: D1.2 Product Specification Document, version 6.3. Available at: http://www.esa-fire-cci.org/documents.

Fraser, R.H., & Li, Z. (2002). Estimating fire-related parameters in boreal forest using SPOT VEGETATION. Remote Sensing of Environment, 82, 95-110

Fraser, R.H., Li, Z., & Cihlar, J. (2000). Hotspot and NDVI Differencing Synergy (HANDS): a new technique for burned area mapping over boreal forest. Remote Sensing of Environment, 74, 362-376

Giglio, L. Csiszar, I., Justice, C.O. (2006) Global distribution and seasonality of active fires as observed with the Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) sensors. Journal of Geophysical Research-Biogeosciences 111, G2. doi: 10.1029/2005JG000142

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., Schroeder, W., Justice, C.O. (2016) The collection 6 MODIS active fire detection algorithm and fire products. Remote Sensing of Environment 178, 31-41.

Gong, P., Pu, R.L., Li, Z.Q., Scarborough, J., Clinton, N., & Levien, L.M. (2006). An integrated approach to wildland fire mapping of California, USA using NOAA/AVHRR data. Photogrammetric Engineering and Remote Sensing, 72, 139-

150.

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.

Hooker, S. B., Firestone, E.R., Acker, J. G., Campbell, J. W. Blaisdell, J. M., Darzi, M. (1995) NASA Technical Memorandum 104566, Volume 32: Level-3 SeaWIFS data product – Spatial and temporal bining algorithms. SeaWIFS technical report series. NASA Goddard Space Flight Center; Greenbelt, MD, United States. Available at: https://oceancolor.gsfc.nasa.gov/SeaWiFS/TECH_REPORTS/ PreLPDF/PreLVol32.pdf

Iri, M., Murota, K., Ohya, T. (1984) A fast Voronoi-diagram algorithm with applications to geographical optimization problems. In: Thoft-Christensen P. (Eds.) System Modelling and Optimization. Lecture Notes in Control and Information Sciences, Vol. 59, pp 273-288. Springer, Berlin, Heidelberg.

Kirches G., Krueger O., Boettcher M., Bontemps S., Lamarche C., Verheggen A., Lembrée C., Radoux J. and Defourny P. (2013). "Land Cover CCI: Algorithm Theoretical Basis Document Version 2." Land_Cover_CCI_ATBDv2_2.3. Louvain, Belgium, 191 pp. Available at http://www.esa-landcover-cci.org/?q=documents#.

Martín, M.P., Gómez, I., & Chuvieco, E. (2005). Performance of a burned-area index (BAIM) for mapping Mediterranean burned scars from MODIS data. In J. Riva, F. Pérez-Cabello, & E. Chuvieco (Eds.), Proceedings of the 5th International Workshop on Remote Sensing and GIS applications to Forest Fire Management: Fire Effects Assessment (pp. 193-198). Paris: Universidad de Zaragoza, GOFC-GOLD, EARSeL.

Pereira, J.M.C. (1999). A Comparative Evaluation of NOAA/AVHRR Vegetation Indexes for Burned Surface Detection and Mapping. IEEE Transactions on Geoscience and Remote Sensing, 37, 217-226

Pinty, B., Verstraete, M.M. (1992). GEMI: a non-linear index to monitor global vegetation from satellites. Vegetatio, 101, 15-20.

Pu, R.L., Li, Z.Q., Gong, P., Csiszar, I., Fraser, R., Hao , W.-M., Kondragunta, S., & Weng, F. (2007). Development and analysis of a 12-year daily 1-km forest fire North America from NOAA/AVHRR data. Remote Sensing of Environment, 108, 198-208.

Randerson, J.T., Chen, Y., van der Werf, G.R., Rogers, B.M., Morton, D.C. (2012) Global burned area and biomass burning emissions from small fires. Journal of Geophysical Research: Biogeosciences, 117 (G04012), 1-23.

Roy, D.P., Giglio, L., Kendall, J.D., & Justice, C.O. (1999). Multi-temporal active-fire based burn scar detection algorithm. International Journal of Remote Sensing, 20, 1031-1038.

Santoro, M, Kirches, G., Wevers, J., Boettcher, M., Brockmann, C., Lamarche, C. Bontemps, S., Defourny, P. (2017) Land Cover Product User Guide Version 2.0 (CCI-LC-PUGv2). Available at: https://www.esa-landcover-cci.org/?q=webfm_ send/112.

Annex 1: Acronyms and abbreviations

BA	Burned Area
BDRF	Bidirectional Reflectance Distribution Function
CCI	Climate Change Initiative
CDF	Cumulative Distribution Function
CL	Confidence level
DoY	Day of the Year
ECV	Essential Climate Variables
ESA	European Space Agency
ETM+	Enhanced Thematic Mapper Plus
FOV	Field of view
FRAP	Fire Resource Assessment Program
GEMI	Global Environmental Monitoring Index
HS	Hotspot
IPCC	Intergovernmental Panel on Climate Change
ISIN	Integerised sinusoidal grid
JD	Julian Day, also day of the year of first detection of a fire
JSON	JavaScript Object Notation
L3	Level 3
LBD	Likely burned date
LC	Land Cover
LC_cci	Land Cover CCI
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
NA	Not applicable
NAFI	North Australian Fire Information
NIR	Near InfraRed
OBPG	Ocean Biology Processing Group
PAF	Potential Active Fire
Pb	Probability of burn
PSD	Product Specification Document
PDF	Probability Distribution Function
QA	Quality assessment
R	Red Band
TH_G	Growing Threshold
TH_S	Seed Threshold
TM	Thematic Mapper
XML	eXtensible Markup Language

Annex 2: Aggregation of the MODIS BA confidence level to the grid product

The outputs of the MODIS BA algorithm provide a binary indicator of the presence of a fire within a certain time period (a month). This information is extended with the most likely date of the fire taken place if a fire has been detected. This information is available at high spatial resolution (10s-100s of m grid size), and is too detailed for climate users, where spatial resolution is usually of the order of 100s of km. In order to aggregate the high resolution information to the coarser climate grid of 0.25 degrees, the procedure starts by defining what high resolution pixels fall within a given grid cell, and then counting the pixels in that set that have burned. This is either reported as an area (multiplying the sum of burned pixels by the individual high resolution pixel area), or as an area fraction.

A major development in the ESA Fire CCI is the addition of uncertainty information. This means that each pixel is qualified by some uncertainty, a so-called probability of burn, p_b. This is a number between 0 (absolute certainty that the pixel **did not** burn in the temporal interval considered) to 1 (absolute certainty that the pixel **did** burn in the temporal interval considered). This metric should reflect the degree in which factors limit the detection of fires (e.g. the observational opportunity, the inevitable limitations of the pre-processing chain, such as residual atmospheric effects not fully corrected by atmospheric correction, gridding artefacts, the properties of the fire and its effect in the remote sensing signal, etc.). These developments affect how one does aggregation to a coarser resolution.

A2.1.   Aggregation basics

From the point of view of the ESA Fire CCI pixel-level product, there are two layers which are relevant: the date of first detection (JD), and confidence level (CL). Additionally, some pixels will be labelled as non-burnable (e.g. ocean or lakes, deserts, etc.), or may be deemed unobservable (insufficient number of observations). Those with insufficient observations form an important aspect of the grid cell information, either for use as a quality control measure or for attempting a correction for missing values. For this reason all pixels considered 'non burnable' should be labelled as such, even if they are not observed. Generally speaking, the burned area inside a grid cell can be determined as the sum of pixels where the first date of detection is between 0 and 366 (both inclusive), multiplied by the area of the pixel. This is intuitive and in line with previous estimates. However, if the confidence layer is interpreted as a probability of burn, p_b, (and in consequence, a probability of not being burned of 1- p_b), then this information would need to be scaled up to the grid cell, as a form of standard error. There are two common definitions relating to standard error[3]: (i) the square root of the estimated error variance (standard deviation); (ii) the standard error of a sample of sample size n is the sample standard deviation divided by $\sqrt{n}$. It is necessary to consider then which would be appropriate in this context.

The sample variance σ^2 of a sample set of size n is given by:

---

[3] Weisstein, E.W. Standard Error. From: MathWorld – A Wolfram Web Resource, available at http://mathworld.wolfram.com/StandardError.html (accessed March 2018).

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2,$$

where y_i is sample i and $\bar{y}$ is the sample mean, given by

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

The sample terms $\bar{y}$ and σ^2 are random variables, and the expected value of the variance $\hat{\sigma}^2$ is given by

$$\hat{\sigma}^2 = \frac{n}{n-1}\sigma^2$$

Often, σ^2 is the *biased sample variance* and $\hat{\sigma}^2$ is the *unbiased sample variance.* Going back to the earlier discussion of definitions of standard error, the first definition $\hat{\sigma}_1$ (square root of the estimated error variance) is thus

$$\hat{\sigma}_1 = \frac{n}{n-1}\sigma$$

where σ is the sample standard deviation. Using the second definition (sample standard deviation divided by $\sqrt{n}$), then

$$\hat{\sigma}_2 = \frac{1}{\sqrt{n}}\sigma_1$$

The first definition is more consistent with many uses of standard error in the physical sciences, where it takes the role of an unbiased estimate of the standard deviation of a distribution. If the distribution is assumed Normal and y is continuous (or effectively continuous if n is large), then the estimate of the mean ($\bar{y}$) and standard deviation σ_1 fully define the Probability Distribution Function (PDF) for BA.

The second definition is more directly related to the uncertainty of the mean and is used in the definition of probable error. The standard error of the mean is given by σ^2. So, with more samples (greater n) the mean of the distribution can be better estimated.

In the light of this, the formula used is:

$$\hat{\sigma} = \frac{1}{\sqrt{n-1}}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

which is a unbiased estimate of the likely variability in burned area.

Assuming that each pixel has an independent probability of burn p_b, which can be different for every pixel, then the sum of these independent probabilities is given by a Poisson Binomial distribution. This distribution is only defined over positive integer numbers, and has first and second order statistics given by

$$\overline{N_b} = \sum_{i=1}^{N_p} p_{b,i}$$

$$\sigma_b^2 = \sum_{i=1}^{N_p} p_{b,i} \left(1 - p_{b,i}\right)$$

In Figure 35, the full PDF derived from a set of samples is shown, each characterised by a different p_b. The PDF was calculated as a Poisson binomial, as well as the mean and variance using the equations above, and the normal approximation to the PDF was plotted. For a large number of samples, the skewness of the PDF was very low, and the PDF was acceptably approximated by a Gaussian distribution. This is of importance, as it means that one can parametrize the full PDF of BA using only the mean and the "standard error" (defined as the standard deviation in the discussion above), and in accordance to the product specification.
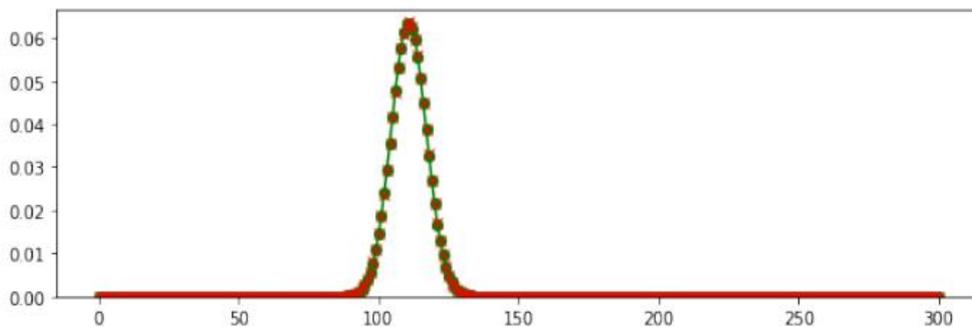

Figure 35: The Poisson binomial PDF (green line) derived from a simulated set of independent samples (300, 100 with probabilities between 0.7 and 0.9, 100 with probabilities between 0.2-0.3 and 100 with probabilities between 0-0.1). A Gaussian approximation (red line) derived from calculating the mean (~110) and standard deviation (~39) is also shown. Skewness was ~0.01.
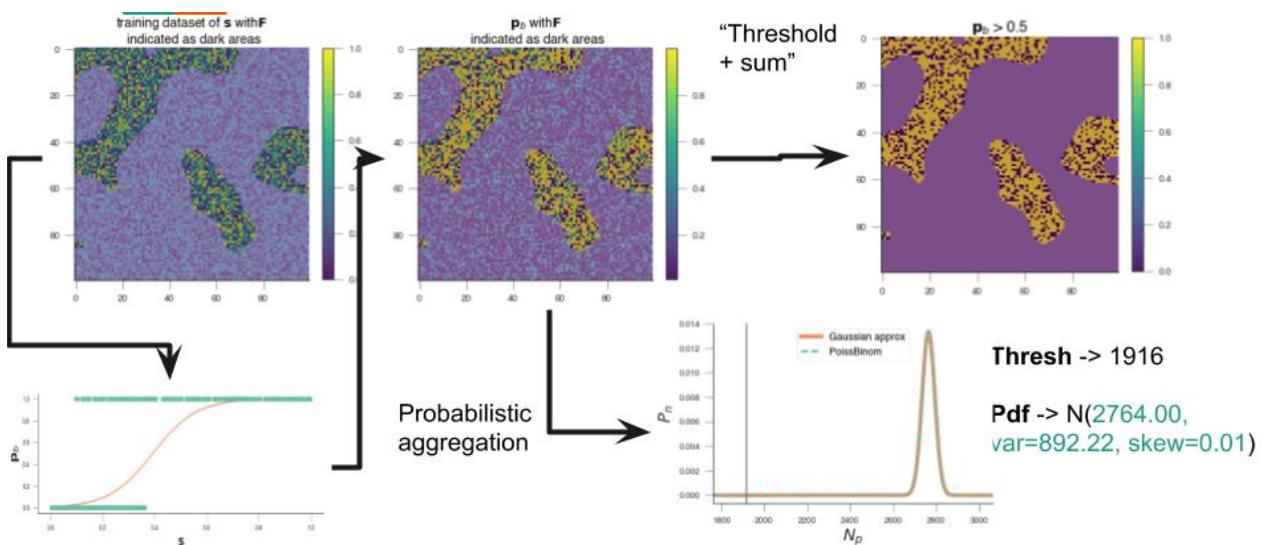

Figure 36: Example of aggregation. See text for more details.

Figure 36 shows a typical workflow. The top left panel shows the spatial distribution of some metric *s* that is somehow related to burned area (e.g. reflectance in some band, or a band combination). The burn scars are characterised by larger values in this space, and it is clear that there is some randomness. On the panel below that one, a mapping from the *s* to probability of burn is shown (in this case, it was done using a simple logistic regression, but this is general). The middle top panel shows the spatial distribution of the probability of burn, which shows a speckly but visually clear distribution of burn area. If a threshold is applied to the probability of burn map (so that e.g. any pixel with $p_b > 0.5$ is assumed burned), the result is a binary map with a value of 0 for values below 0.5 and 1 for values above. Finally, the aggregation from both approaches can be displayed. The thresholded approach results in a single value, the sum of pixels above the threshold (in this example, 1916). The proposed aggregation using a Poisson binomial results in a PDF (again the Poisson Binomial and the Gaussian approximation are shown), with a mean of ~2760 and a standard deviation of ~30. The actual number of burned pixels in this case was around 2780.

A2.2. Unreliable probability of burn estimates

Since the quantification of probability of burn per pixel is still fairly new, and both users and product developers are more used to using the sum of pixels aggregation, it might be useful to consider how to make both approaches compatible. One simple approach might be to re-scale p_b based on the sum of pixels: if the mean of the Poisson binomial (or Gaussian approximation) is given by the sum of pixels (rather than by the sum of individual p_b), then the individual values of p_b can be scaled so that the mean is identical to the sum of pixels (Figure 37), and then use that rescaled p_b to calculate the standard deviation and thus provide some form of uncertainty estimate.
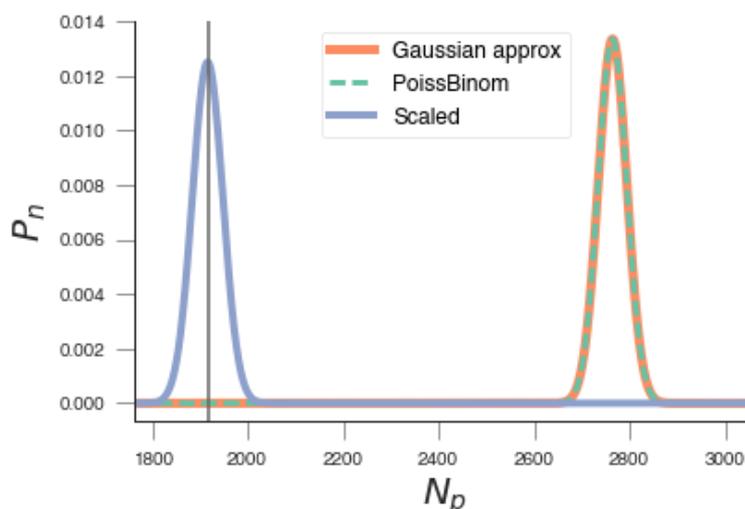

Figure 37: Example of applying the rescale approach to the example shown in Figure 36.