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 Nº	4000115006/15/I-NB
Issue Date	20/11/2018
Version	2.0
Author	Joshua Lizundia-Loiola, M. Lucrecia Pettinari, Emilio Chuvieco, Thomas Storm, José Gómez-Dans
Document Ref.	Fire_cci_D2.1.3_ ATBD-MODIS_v2.0
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 v5.1, 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	12/11/2018
Reviewed	UAH – Project Manager	Lucrecia Pettinari	20/11/2018
Authorized	UAH - Science Leader	Emilio Chuvieco	20/11/2018

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
2.0	20/11/2018	Update to cover FireCCI51

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.
			Figures 6, 10, 14	Title expanded
			Figures 17, 22	Figure updated
2.0	20/11/2018	UAH	Sections 1, 2.2.3, 2.3.1, 2.7	Minor updates to the text
			Section 2.1, 2.3.2, 3.2.3, 2.4.2, 2.4.3, 2.5 (and sub-sections), 2.6	Text updated
			Section 2.3.5	Section removed

Table of Contents

<u>**List of Tables**</u>

<u>**List of Figures**</u>

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.

The first algorithm developed within the project 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 (FireCCI41) product only covers the years 2005-2011 (Chuvieco et al., 2016). However, a new algorithm based on the MODIS sensor (MODIS Fire_cci v5.0 or FireCCI50) extended that time series to the period 2001-2016 (Chuvieco et al., 2018). The algorithm was developed for the two highest-spatial resolution bands of the MODIS sensor (Red and Near Infrared, NIR) and followed a similar approach to the MERIS algorithm, as it combined 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 FireCCI41 product advised to extend previous developments and introduce some adaptations for that MODIS BA algorithm.

Nevertheless, posterior analysis of the products showed that some issues related to the input variables and thresholding approach could be significantly improved. This document explains the theoretical basis of the improved global burned area (BA) algorithm MODIS Fire_cci v5.1 (FireCCI51), which make use of all the knowledge acquired through the years within the Fire_cci project.

2. BA Algorithm description

2.1. General scheme

Following our previous experience (Alonso-Canas and Chuvieco, 2015; Chuvieco et al., 2018) 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 FireCCI50 BA algorithm (Lizundia-Loiola et al., 2018; Chuvieco et al., 2018), the general scheme of the FireCCI51 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. Once the composites are obtained the relative NIR difference is calculated between two consecutive months. These two variables are the basis to perform the seed identification phase, but also the growing phase, which is performed on the composite and NIR difference images to obtain the BA maps.
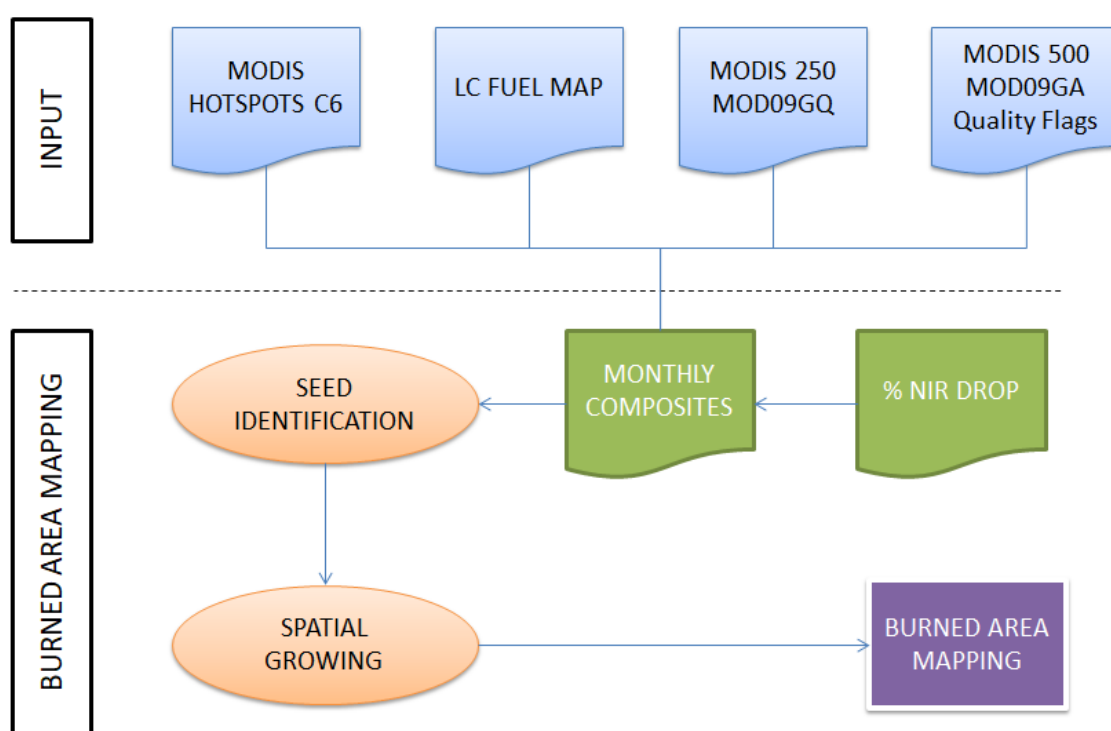
Figure 1: Algorithm main scheme definition

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 September 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
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.
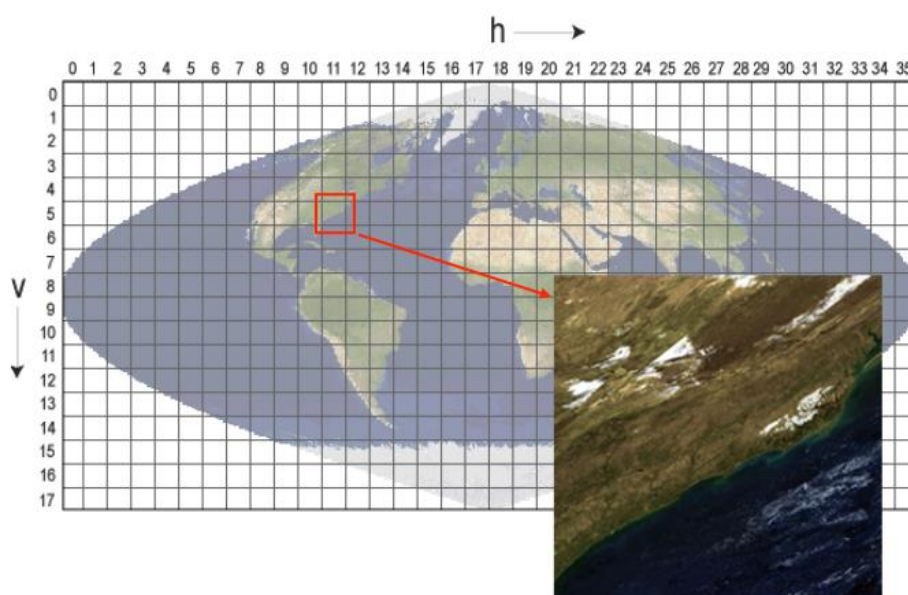
Figure 2: MODIS Sinusoidal Grid, as shown in Vermote et al. (2015).

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 September 2018), which in Band 2 (state_1km_1: Reflectance Data State) contains the State Quality Assessment Description (QA) that offers 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
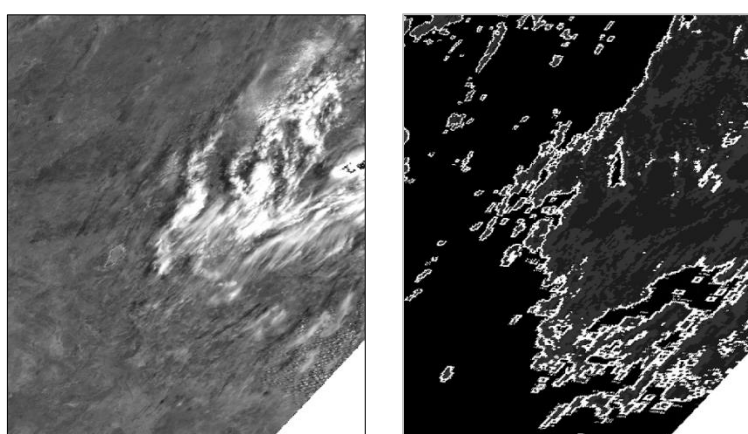
(a)   (b)

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 of MOD09GA.

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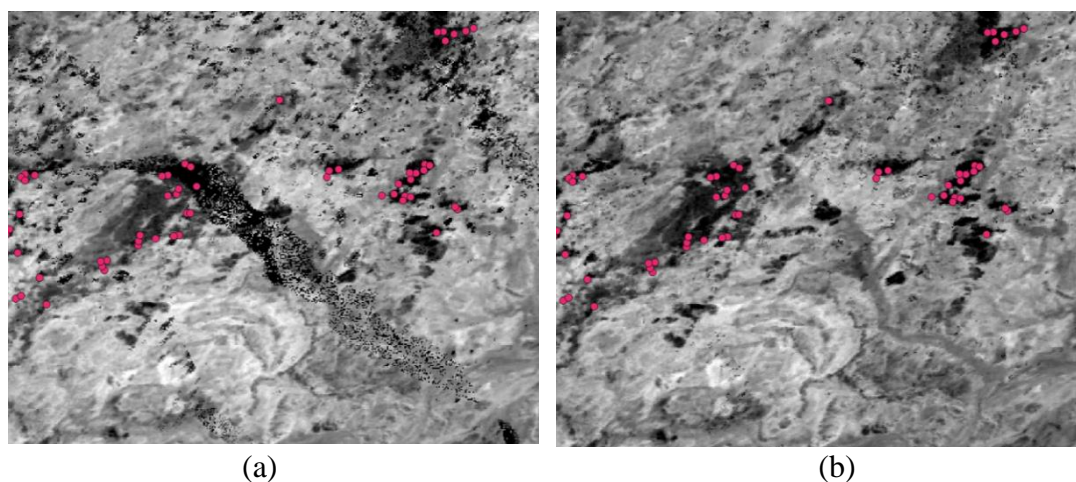
Figure 4: Example of June 2008 for tile h30v10. (a) Composite (see Section 2.3) without QA correction. (b) Composite with QA applied to the images that create it. The red dots are the hotspots for that period. The QA eliminates the dark pixels in the middle of the image that correspond to cloud shadows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

2.2.3. Hotspots

The HS information was derived from the MCD14ML Collection 6 product, which improves the detection of active fires from C5. 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 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, ESA, 2017), which provides information on global land cover. The collection used for this second global MODIS BA processing is v2.0.7, containing annual LC information between 1992 and 2015. 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 prior to the BA analysis, i.e. LC_cci 2000 was used for the BA calculation of 2001, the LC_cci 2001 for the 2002 BA product, and so on. For 2017 LC_cci 2015 was used. This version of the LC_cci product allows a better characterization of the LC fuel mask used by the FireCCI51 than the previous version, as FireCCI50 used the LC_cci v1.6.1.

The LC_cci product has 37 classes (including no data), which were reclassified into 2 general classes to create a burnable mask (see

Table 3 and Figure 5):

- 0: non burnable
- 1: burnable
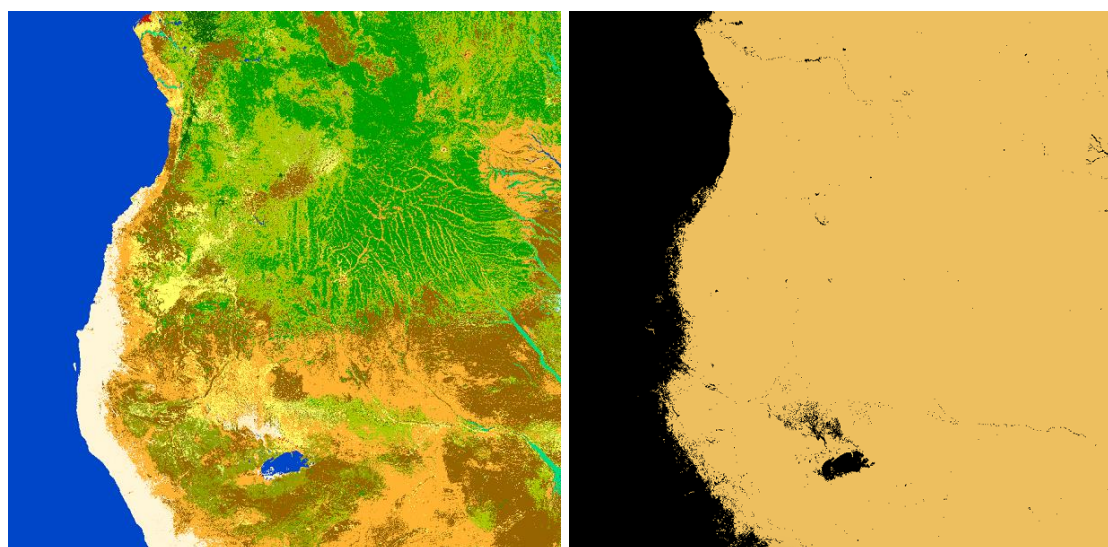
(a)   (b)

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

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	1
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	1
Tree cover, broadleaved, deciduous, closed to open (>15%)	60	1
Tree cover, broadleaved, deciduous, closed (>40%)	61	1
Tree cover, broadleaved, deciduous, open (15-40%)	62	1
Tree cover, needleleaved, evergreen, closed to open (>15%)	70	1
Tree cover, needleleaved, evergreen, closed (>40%)	71	1
Tree cover, needleleaved, evergreen, open (15-40%)	72	1
Tree cover, needleleaved, deciduous, closed to open (>15%)	80	1
Tree cover, needleleaved, deciduous, closed (>40%)	81	1
Tree cover, needleleaved, deciduous, open (15-40%)	82	1
Tree cover, mixed leaf type (broadleaved and needleleaved)	90	1
Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	100	1
Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	110	1

Label	LC_cci Value	Fire_cci class
Shrubland	120	1
Evergreen shrubland	121	1
Deciduous shrubland	122	1
Grassland	130	1
Lichens and mosses	140	1
Sparse vegetation (tree, shrub, herbaceous cover) (<15%)	150	1
Sparse shrub (<15%)	152	1
Sparse herbaceous cover (<15%)	153	1
Tree cover, flooded, fresh or brackish water	160	1
Tree cover, flooded, saline water	170	1
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 2.2.1) 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.
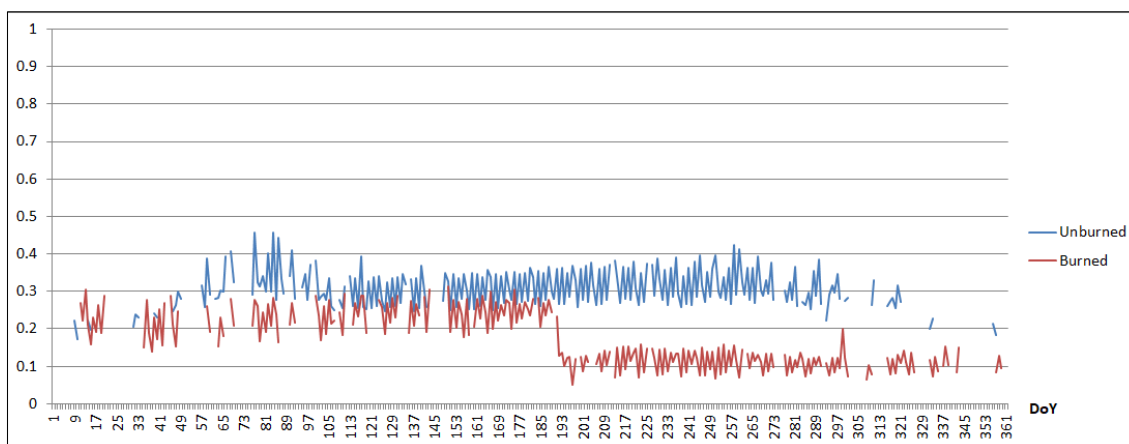
Figure 6: NIR reflectance (Y axis) for a burned and unburned pixel of tile h08v05 (California) during 2008 (days of the year in X axis) after applying state mask. The continuous oscillation of the reflectance is due to the BRDF effect.

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).
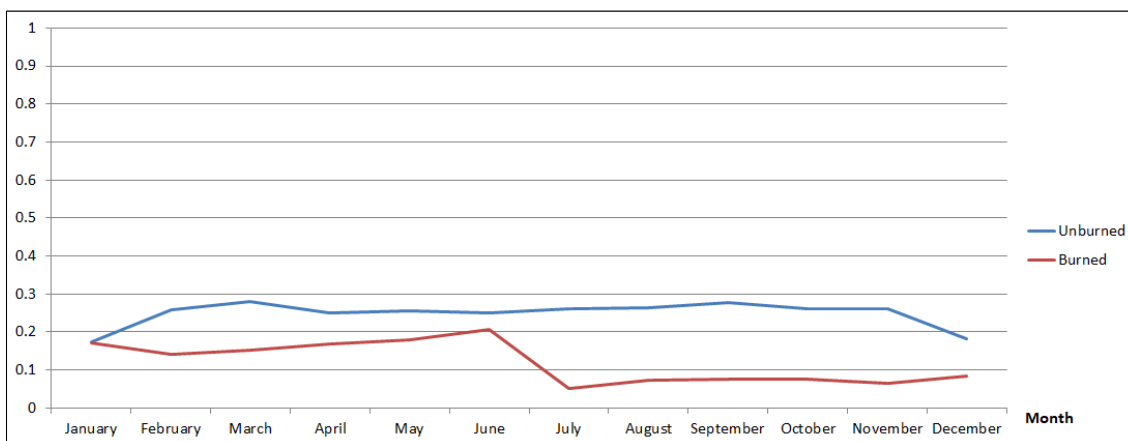
Figure 7: NIR reflectance (Y axis) for the same pixels of Figure 6, as function of the month of the year 2008 (X axis). The smoothing of the BRDF effect due to the composite does not obscure the NIR drop used to detect a burned area (shown in the red line, for a fire occurred in day 169: 17 June 2008).

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 ensure 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. 1984) were created to assign to each pixel the date of the closest HS (CHSD). 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).

The "likely burned date" (LBD) was used to create the monthly NIR reflectance composites by finding the three NIR minima acquired through a moving window around the CHSD. Considering the potential noise in the time series (see next paragraphs), as in the case of the MERIS algorithm, three NIR minima were used instead of a single NIR one. Since HS may occur at the end of the month, the moving window allows a good transition between consecutive months, as it includes images from the following month. Unlike the FireCCI50 algorithm, where the moving window was only applied for the last 10 days of the month generating very irregular searching windows through a same region, in this case FireCCI51 uses 10 days pre- and 10 days post-fire for all the cases. The quantity of post days used by the algorithm could be extended up to 25 days if no more

than three valid observations were found in the first 10 days. This is very useful in regions where persistent clouds are found (Giglio 2009). In this way, it was assured the use of only the necessary images to create the monthly composites or NIRt (Figure 9).

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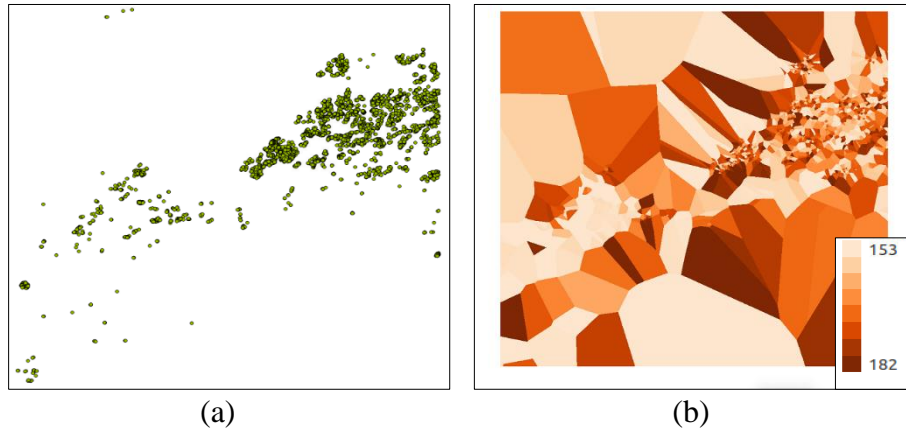
Figure 8: Example of Voronoi polygons created for June 2008 in tile h30v10. (a) Hotspots during that month, including a 50-km buffer. (b) Voronoi polygons with their corresponding date.

Thanks to the stricter searching window around CHSD, 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 for each pixel, their dates were compared with the CHSD, and the closest post LBD minimum NIR value was selected only in cases where at least two minima were found after the CHSD. In the other cases, the second minimum NIR was selected instead of 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).
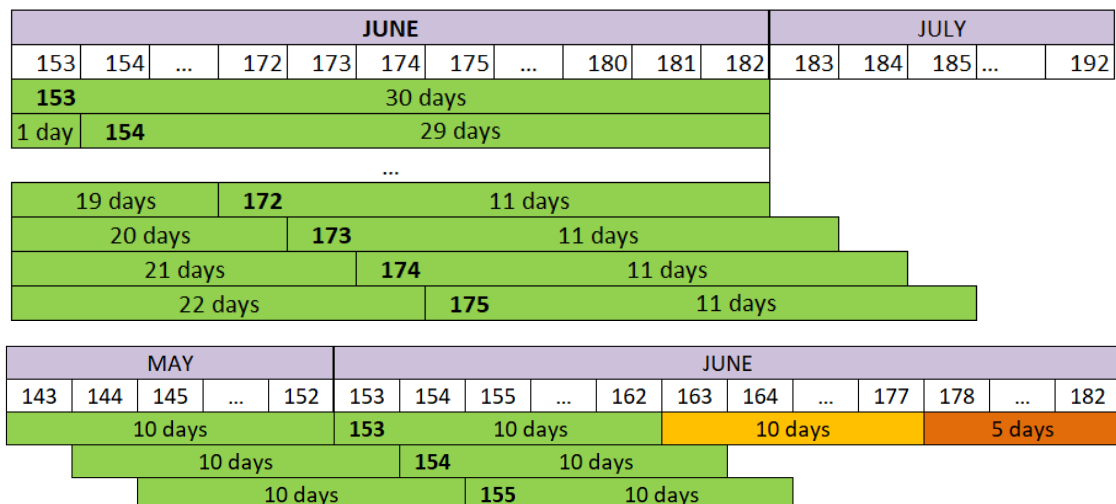
Figure 9: Example of temporal window used for June 2008 by the FireCCI50 (top) and the FireCCI51 (bottom). The day of the year (DoY) of each month is shown at the top. Each line corresponds to a pixel with a hotspot in the day indicated in bold, and the period marked in green indicates the days when the three minimum NIRs are searched. The vertical line before the day shows the break between the pre and post days. The new algorithm FireCCI51 only adds 10 extra post days (yellow) in case three or less valid observations are found in the first 10 post days (green). Another 5 post days are added if this situation persists. The temporal windows of the new algorithm are more stable than in the old one.
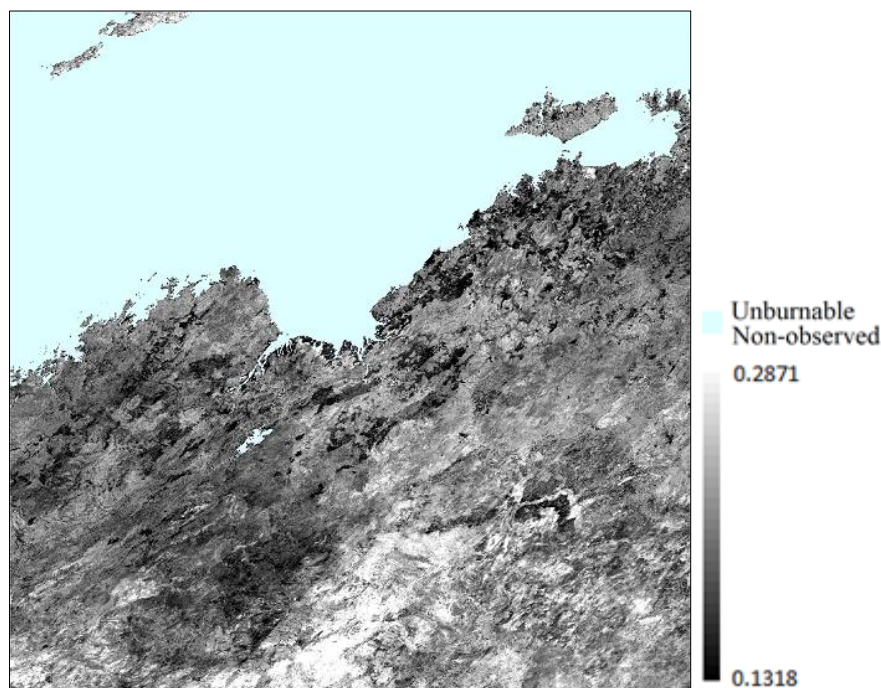
Figure 10: Composite of June 2008 for the tile h30v10 (Australia). This composite has no "*incomplete fires*" or "*eliminate noise*" corrections mentioned in this section.

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:
    - o If the three NIR minima occurred in the 10 days posterior to the HS, the first minimum was selected.
    - o If two NIR minima (including always the first minimum) occurred in the first 5 days posterior to the HS, the first minimum was selected.
- Eliminate noise: 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).
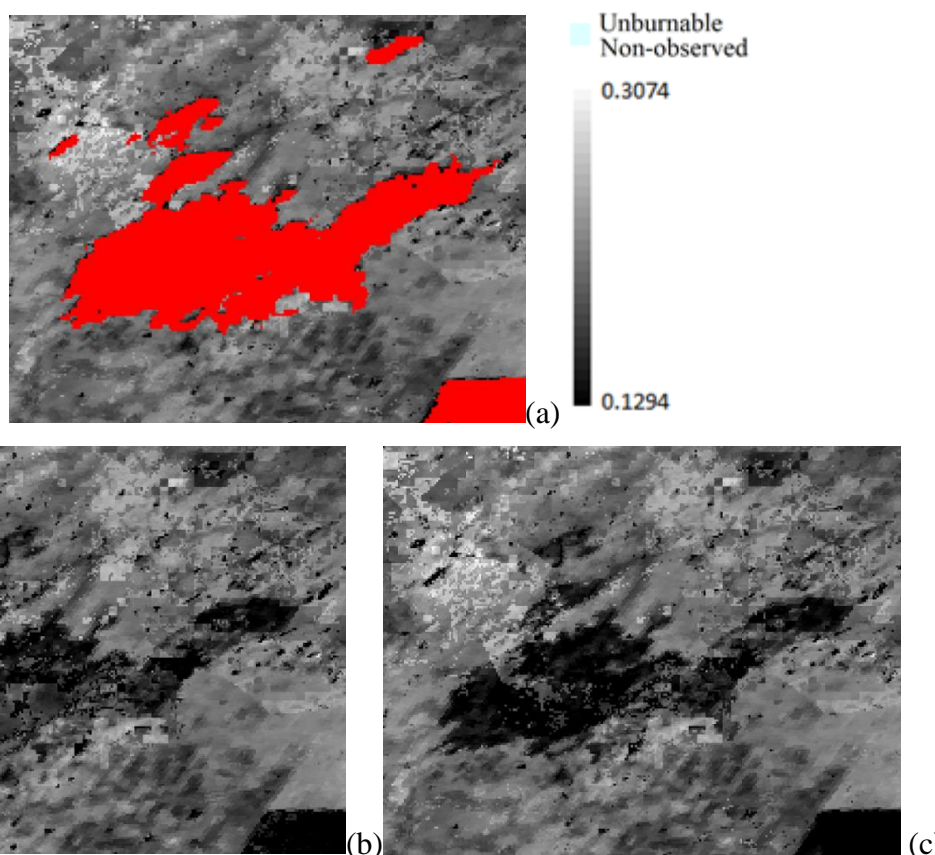
Figure 11: Composite example of November 2008 for tile h30v10. (a) The red areas are the fire perimeters of November 2008 extracted from the North Australian Fire Information (NAFI) database. (b) Original composite without the correction. (c) Corrected composite, where the burned area is more clearly identified.
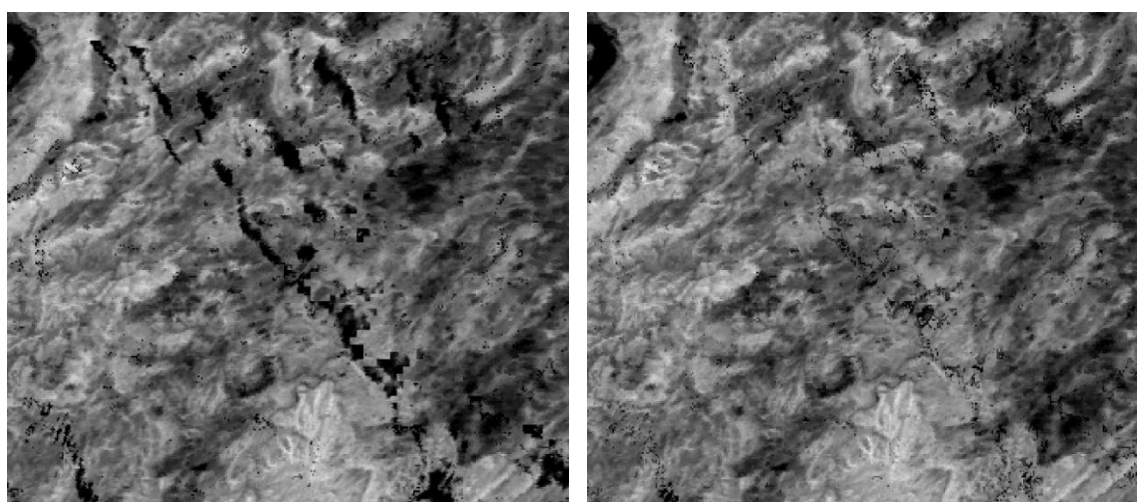
Figure 12: Detail of the composite of June 2008 for the tile h30v10. The left image shows the original composite without the noise correction. The right image shows the composite after the correction. There are no HS in this area during this time period, which shows that the black areas in the left image are due to shadows.

2.3.3. Relative NIR drop

One of the main differences between the two MODIS versions is that FireCCI51 uses only the NIR band, instead of using also the RED band to compute the Global

Environmental Monitoring Index (GEMI). The FireCCI50 used the absolute difference between the maximum pre-fire GEMI and the monthly composite GEMI. However, it was found that the relative NIR drop (RelΔNIR) between two consecutive composites was able to give the same information and, in some cases, a better separability between burned and unburned classes (Figure 13). The RelΔNIR expression is:

$$Rel\Delta NIR_t = \left(1 - \frac{NIR_t}{NIR_{t-1}}\right) \cdot 1000$$

Being NIR_t and NIR_{t-1} the values of two consecutive NIR composites. As this variable is represented by integers the processing and memory issues related to this layer have improved. The obtained values show the relative drop in percentage suffered by the pre-fire value after the fire event i. e. a RelΔNIR = 232 means that the NIR value decrease in 23.2%.
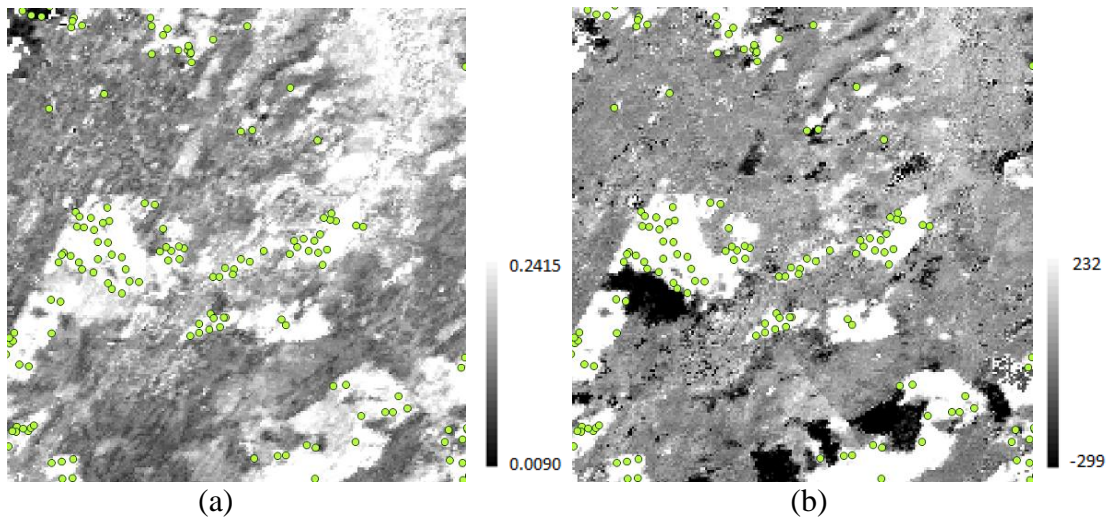
Figure 13: Detail of the GEMI difference (a) and RelΔNIR (b) of June 2008 for the tile h30v10 (Australia). The HS are coloured in green.

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.

- <u>Geometric anomalies</u>: 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.

2.4. 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 NIR product had 250m, a proper method to locate the active fire within each NIR 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 14).

(a)   (b)
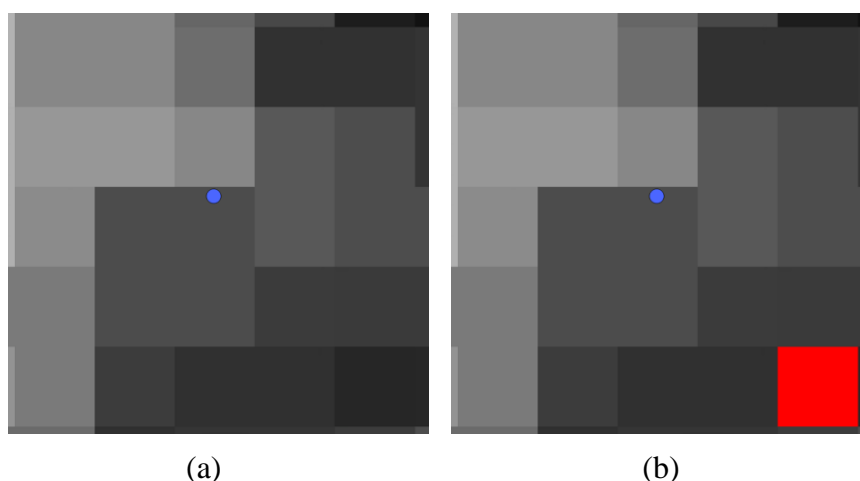
Figure 14: (a) 5x5 pixel window around the pixel where the hotspot is located (blue dot), showing the NIR values in the composite, from higher (brighter) to lower (darker) values. (b) The red pixel is the one assigned to the HS since it has the lowest NIR value of the window.

2.4.2. Potential active fires selection

A threshold value was needed to identify which pixels would be classified as potential active fires (PAF). This threshold was obtained using a Cumulative Distribution Function (CDF) of the unburned class, 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 unburned class, the most representative pixels of it were selected.

To select the unburned pixels used for the threshold statistics calculation, a unique condition was applied:

- The pixel did not have a HS in a vicinity of a 10 km (R_{IN}) buffer distance (circular raster window).

This condition was similar to the FireCCI50 algorithm (Chuvieco et al., 2018), but with a bigger window. In the previous version there was no additional filtering after this phase,

as the thresholding was tile-based, so this first threshold had to be able to discard the majority of the false alarms and HS related to fires too small to be detected at our resolution. However, this smaller window might generate too restrictive thresholds for some land covers, as it did not take into account the internal variability of the tile. So in the case of the FireCCI51 the objective of this first filtering was to eliminate the clearest cases of false alarms. Besides, the use of circular raster windows instead of a square one allowed being more precise in the application of the buffer distances, although the sinusoidal projection is not equidistant. The selected distance R_{IN} has been previously used by other authors (Giglio et al., 2009 and Giglio et al., 2018). So using the pixels that comply with the above criterion, the CDF of the NIR reflectance for unburned pixels was created, and the Seed Threshold (TH_S) was established as 10% of the CDF, the same used in the previous versions.

To confirm that a HS is a PAF, the following conditions had to be met:

- The RelΔNIR > 10%
- The NIR value of the pixel was lower than TH_S.
- At least 3 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.

The third condition introduced a considerable change between the two versions. In the oldest version the number of pixels that had to comply with the conditions were 5 of the 8 surrounding pixels. However, this stricter condition filtered lots of HS related to small fires. So the condition was reduced to three in favour of a reduction of the omission error. The 10% relative NIR drop has been empirically established based on low intensity fires from African savannah and southern Sahara.

2.4.3. Thresholding

Another main change introduced in this new version is the method used to compute the thresholds in each tile. The approach used by the FireCCI50 was inherited from FireCCI41 which stablished that the tiles of MODIS were a good statistical unit to compute thresholds. But the analysis carried out with FireCCI50 showed that in several tiles this approach was not performing adequately. Tile h19v10 (Angola) is a clear example of that. In this case the northern part is composed of deciduous broadleaved forests and the southern part shrublands and grasslands are the main land cover, generating two areas with a high contrast in NIR reflectance values within the same tile (Figure 15).

So in FireCCI51 the thresholding process was strictly related to the context of each fire, instead of being a unique tile-based threshold. In each of the following three steps the initial PAFs were filtered to get the final seeds and a threshold adjusted to each spectrally similar region was estimated.
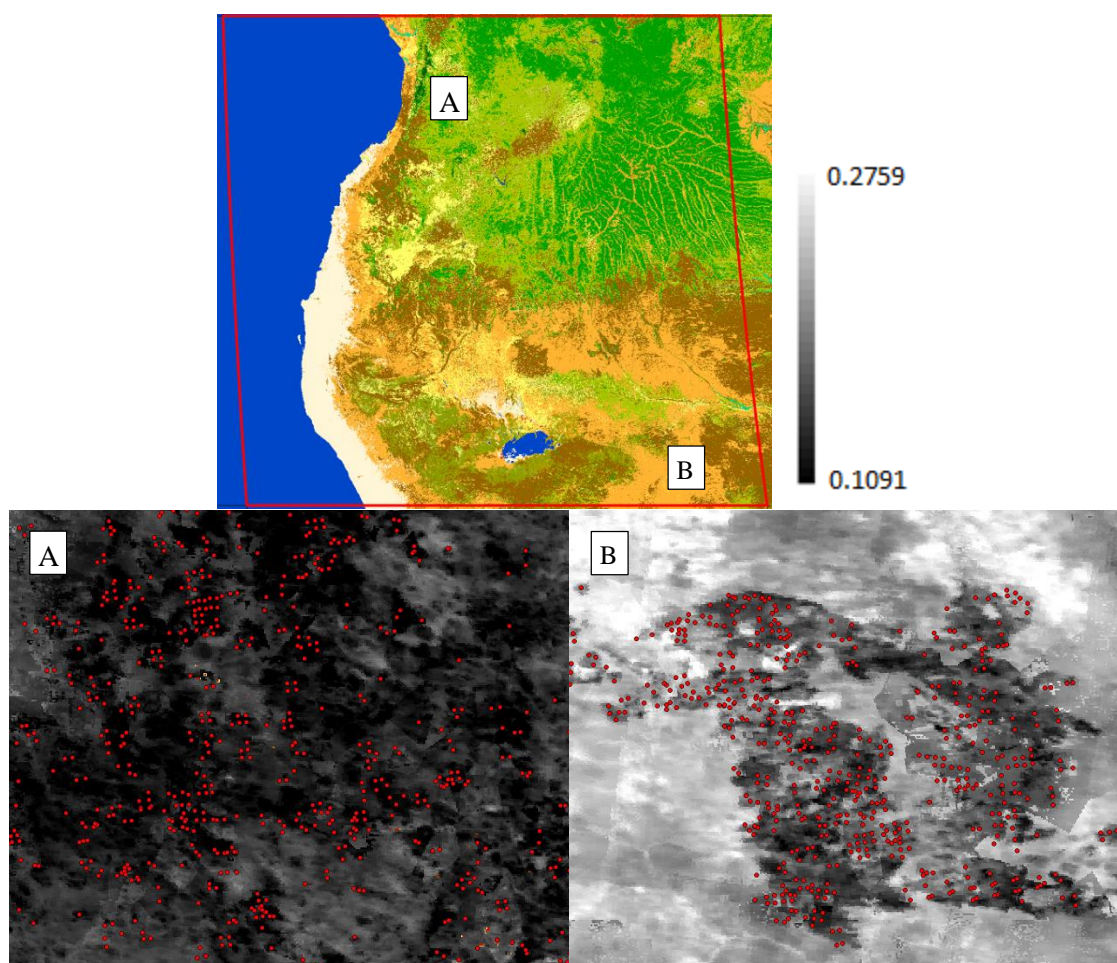
Figure 15: Example of tile h19v10 (Angola) for July 2008, showing the NIR reflectance contrast between two land covers: A represents an area of deciduous broadleaved forest (LC code 60 and 62) and B represents a grassland fire (LC code 130). Both images show the NIR composite coloured with the same grey scale. The top image shows the LC_cci corresponding to year 2008.

2.4.3.1. Spatio-temporal HS clusters

In order to obtain an adequate threshold for each area of the tile with a similar spectral behaviour, HS were first clustered using spatial and temporal conditions. The rationale behind this clustering is that HS that are near to each other and have a similar date are likely to be part of the same fire. So two conditions were applied to group HS in a same spatio-temporal cluster (STC):

   - the distance between the HS must be < 3750 m.

   - the difference of detection between the HS must be <= 4 days.

The first condition is derived from a study carried out by Hantson et al. (2013) where the commission and omission relations of MODIS HS with fire reference perimeters were analysed. He found that an important proportion of HS was found in a 1500 m buffer distance set around the reference perimeters. So, after adding 375 m (1.5 pixels) to reduce problems related to geo-location, an area of influence of 1875 (R_{AI}) m was established for each HS. HS with overlapping R_{AI} were also included in the same STC, therefore the 3750 m threshold. The second condition was derived from an analysis done by Archibald

and Roy (2009), where they used a threshold of 4 days to consider if two adjacent burned pixels were part of the same fire or not.

Once the STC were created burned and unburned samples were obtained in order to find a representative threshold for each STC. The former class is represented by the PAFs of the STC while the latter is formed by pixels that are between 10 (R_{IN}) and 20 (R_{OUT}) kilometres from the nearest HS of the STC being analysed. Pixels that are within the R_{AI} of any PAF of any STC are removed to reduce the presence of possible burned areas in the unburned sample. The R_{IN} has been previously used in section 2.4.2 and the R_{OUT} was set as the farthest distance from a HS that a burned pixel can be found based on the calibration areas. Then the following mathematical expression was applied in order to obtain the threshold of each STC:

$$TH_{1,i} = \frac{1}{3} \cdot Med_{B,i} + \frac{2}{3} \cdot Mo_{U,i}$$

Where i represents each STC, $Med_{B,i}$ is the median of the burned class and $Mo_{U,i}$ is the mode of the unburned class. Based on this simple lineal relation, different metrics were applied before the definitive expression was achieved. Mean (M), Median (Med) and Mode (Mo) metrics were used to find a representative value for each class. In the case of burned samples only M and Med were used, as the small sample size in some cases makes the Mo too unstable. The analysis showed that the best combination of metrics was given by the Med for the burned class and Mo for the unburned one (see Annex 2).

Once these contextual thresholds were obtained spatial relations described in the following two sections were used to refine them.

2.4.3.2. Spatial HS clusters

In the following step the STCs were merged in spatial HS clusters (SC) attending only to the distance condition mentioned in section 2.4.3.1 (Figure 16). That previous step ensured a good burned sampling, especially in regions with a high HS density where too large groups are created if only a distance relation is used. So, based on the closeness of near STC, a new threshold is obtained for each SC:

$$TH_{2,j} = \frac{\sum_{i=1}^{N} w_i \cdot TH_{1,i}}{\sum_{i=1}^{N} w_i}$$

Where j represents each SC, N represents the total number of STC within j, i represents each STC within j, $TH_{1,i}$ is the threshold of i obtained in section 2.4.3.1 and w_i is the weight of each STC given by the number of PAFs in it.
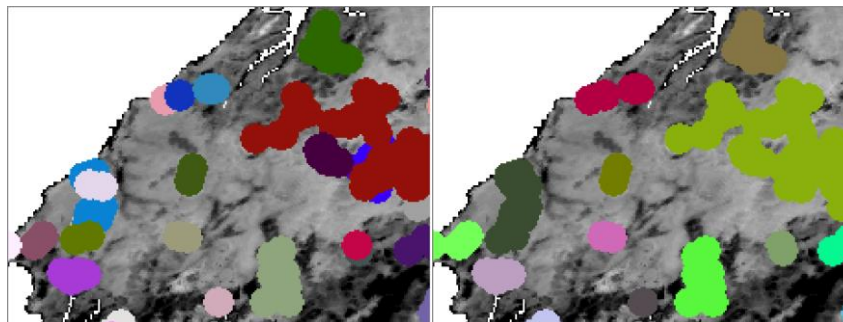
Figure 16: Example of the thresholding process in h30v10 (North Australia) of July 2008. In the image on the left spatio-temporal clusters have been coloured and on the right those STC have been merged in spatial clusters (SC).

2.4.3.3. Regional HS clusters

The last adjustment that was applied to each $TH_{2,j}$ is based on the existing influence among SC groups that are close to each other. The final objective of the thresholding is to obtain thresholds for each area with a similar spectral behaviour. So thanks to the regional influence more stable thresholds are obtained for each area, since the influence of anomalous thresholds is reduced. Therefore, the area of influence of each SC is set as R_{OUT}. A similar weighted average of the previous section is again applied:

$$TH_k = \frac{\sum_{j=1}^{M} w_j \cdot TH_{2,j}}{\sum_{j=1}^{M} w_j}$$

Where k represents each SC, M represents the total number of SC within R_{OUT} from k, j represents each SC within R_{OUT} from k (including k), $TH_{2,j}$ is the threshold of j obtained in section 2.4.3.2 and w_j is the weight of each SC given by the number of PAFs in it.

All the process is done for both NIR and RelΔNIR variables, hence two thresholds are obtained for each SC, one per variable (TH_NIR_k and TH_RelΔNIR_k). However, in many cases SC belonging to a same spectral region will have a very similar or equal final TH_k. Besides, the initial PAFs are filtered in each level using the corresponding TH. Those filtered PAFs will be identified as seeds in the rest of the document.

Figure 17 shows an example of the thresholds calculated for September of 2008 in Angola (h19v10). It is a clear example of how a context-based thresholding is able to better adapt the existing variety of reflectance in the MODIS standard tiles.
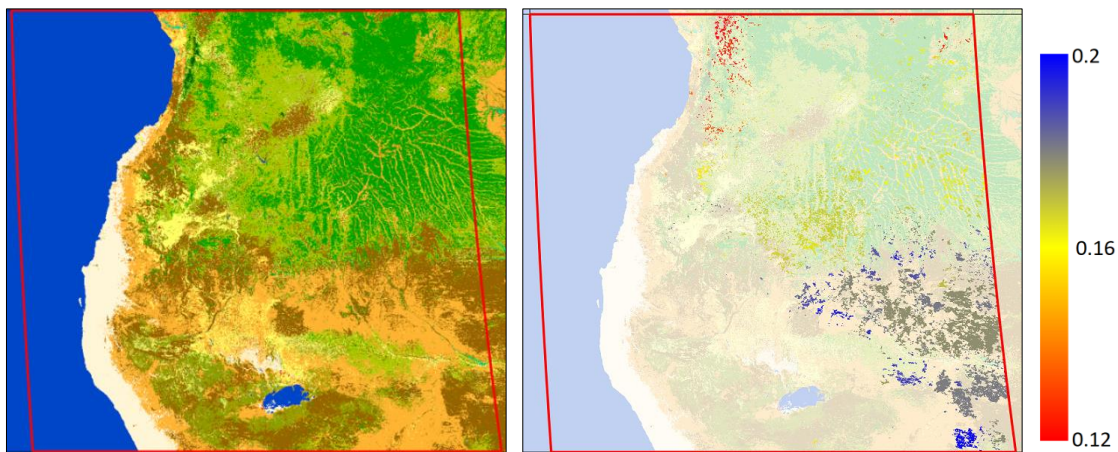
Figure 17: Example of tile h19v10 (Angola) for September 2008. On the left LC_cci of 2008 and on the right burned patches coloured by their final threshold used in the growing phase. The contrast between different spectral regions is clear and also the relationship with the LC.

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. The growing phase has been computed over the NIR composite image and the RelΔNIR image.

2.5.1. Growing method

The growing was performed separately for each SC 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 18.

(a)   (b)
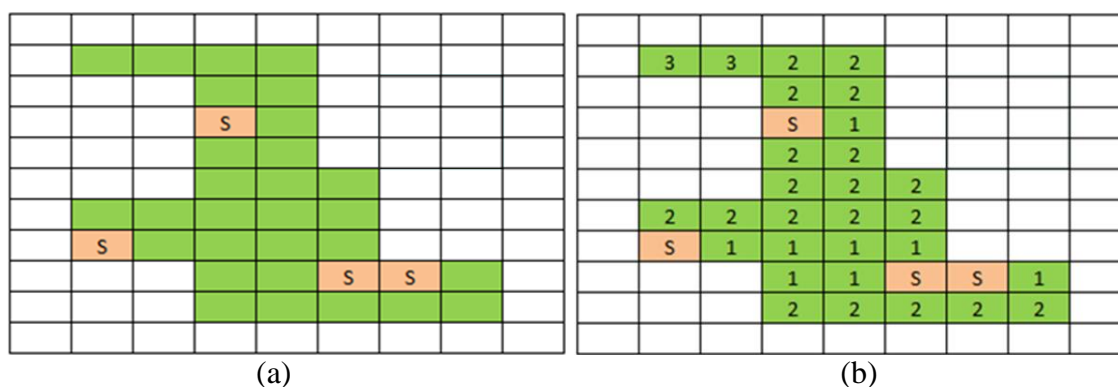
Figure 18: Example of the growing procedure. (a) The orange pixels are the seeds, and the green pixels correspond to the burned area to be detected. (b) The numbers in the green pixels show the iteration in which they were detected.

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

- the NIR value was lower than TH_NIR_k,
- the RelΔNIR value was lower than $TH\_Rel\Delta NIR_k$, and
- at least a neighbouring pixel was a seed or has been classified as burned in a previous iteration.

Immediately after the growing phase of a specific SC was finished two criteria were applied in order to detect possible excessive growing:
- The relation TOTAL_BA / Seeds > 1000 (62.5 km^2)
- BA within R_{AI} / TOTAL_BA < 10%

The first condition is related to the total amount of pixels that have been generated per seed. The second analyses the proportion of the total BA that is found in the R_{AI} of any seed of the SC. Both conditions are quite unrestrictive so if any of them is fulfilled the burned patch is entirely eliminated.

2.5.2. Spatial filter

Once the growing phase was finished burned patches were refined using two spatial filters. The first one is oriented to reduce commission errors that might be generated by excessive growing that have not been detected by the conditions applied in the previous section. In some cases, unburned areas with a similar spectral behaviour than that of burned areas might create "bridges" between the burned area and other regions,

originating commission errors near the burned area (Figure 19). So in this first filtering a burned pixel will be considered unburned if:

- It is out of R_{AI}
- In a 3x3 window, the northern and southern pixels or the eastern and western pixels are unburned.

In a next step another spatial filter was applied to fill possible gaps that can be found within the burned patches. So the inverse of the previous second condition was applied to label an unburned pixel as burned:

- In a 3x3 window, the northern and southern pixels or the eastern and western pixels are burned.

(a)   (b)
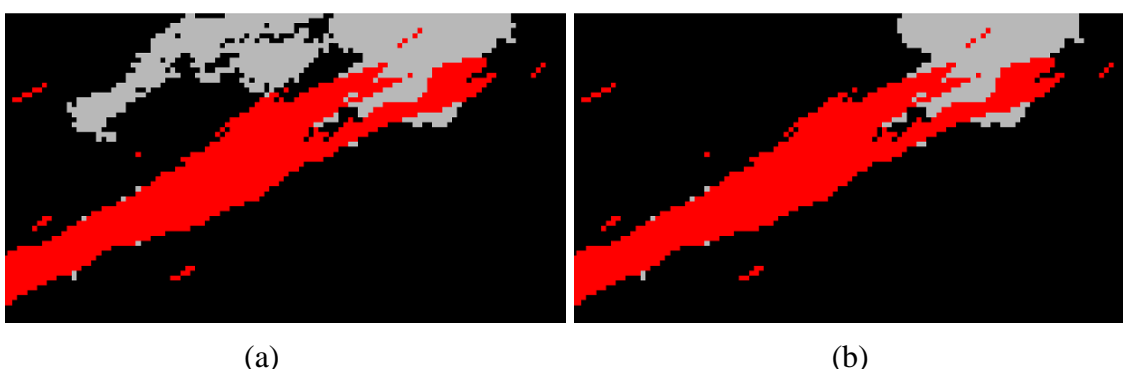
Figure 19: Example of the application of the "breaking bridges" filter to a burned patch in the Northern border of tile h08v05 (California) in June 2008. (a) Burned patch without the filter. (b) Burned patch after the filter. Red colour shows the FRAP perimeters, black the unburned pixels and grey the commission errors.

2.6.   Uncertainty

FireCCI51 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 in the first 10 post – fire days. In the previous version total number of observations were used, but this does not take into account the position of the valid observation with respect to the HS date. So, the use of the first 10 post – fire valid observations was considered to be a better approach to estimate uncertainty (Figure 20). The rationale behind this variable was that the higher the number of observations, the lower the uncertainty in the observation. This variable had values between 0 and 10.

The second variable was the NIR reflectance (Figure 21). 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 10000 (highest NIR value and lowest probability of burn) and 0 (lowest NIR value and highest probability of burn). The reason for these values was the original scale factor of 0.0001 of MOD09GQ product.
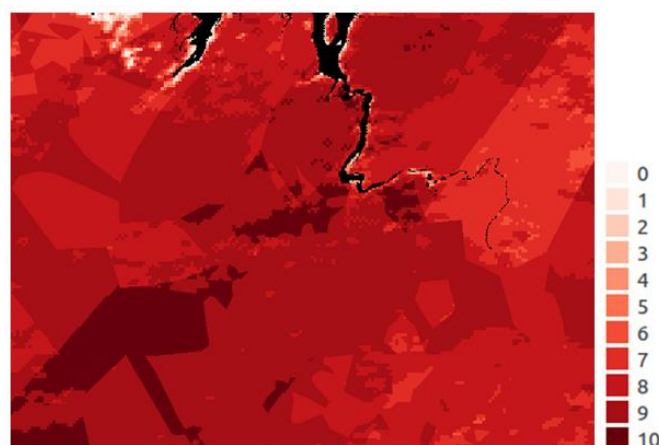
Figure 20: Number of post-fire valid observations of an area of the tile h30v10 (Australia) for June 2008.
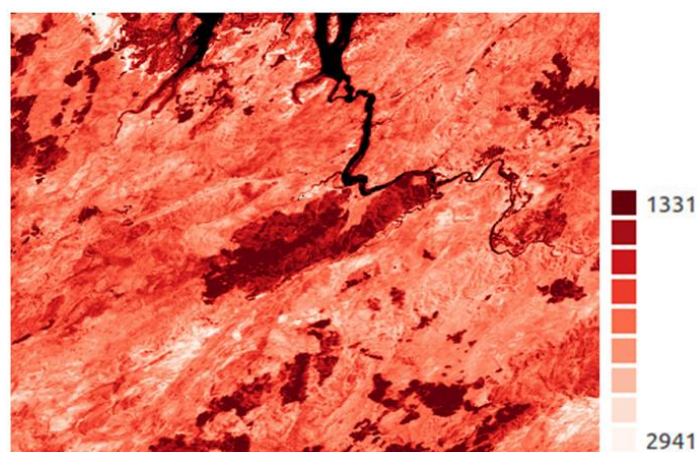
Figure 21: Probability of burn variable related to the NIR reflectance of an area of the tile h30v10 (Australia) for June 2008. The lower the value, the higher the probability of the pixel being burned.

The third variable was similar to the second, but considering the percentage of NIR decrease between t-1 and t. This variable had a behaviour inverse to the NIR: the higher the RelΔNIR, the higher the probability of the pixel being burned. The values ranged between -∞ and 1000, where 1000 represent 100% of relative NIR decrease (Figure 22). This variable values have a 0.001 scale factor.
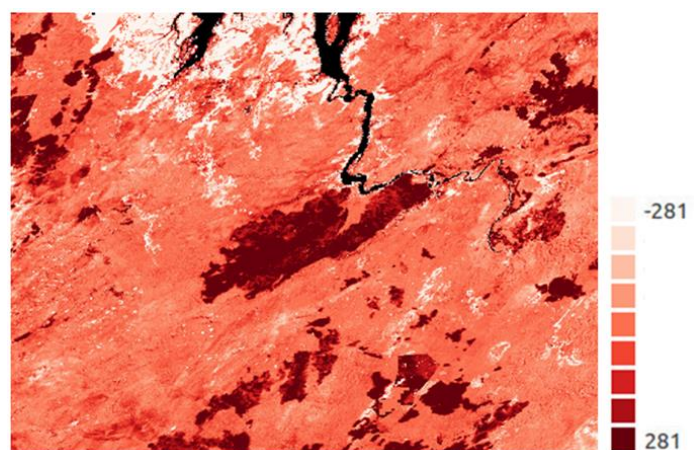
Figure 22: Probability of burn variable related to the RelΔNIR values of an area of the tile h30v10 (Australia) for June 2008. The higher the value, the higher the probability of the pixel being burned.

The last variable used to calculate the uncertainty was the distance to the nearest PAF. The variable was calculated after the final PAFs and, consequently, burned area, were obtained. The distance unit is metre and it ranges from 0 to N metres; in this case N, the maximum distance, will be the diagonal of a MODIS standard tile (Figure 23). A lower distance means that the pixel is near a PAF (seed), so the lower the value, the higher the probability of burn.
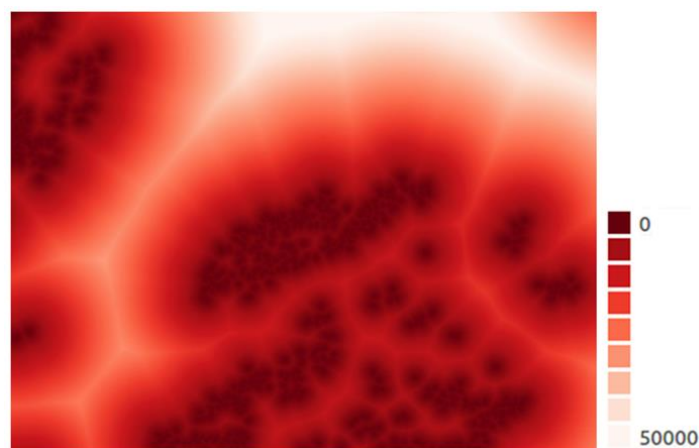
Figure 23: Probability of burn variable related to the distance to the PAF of an area of the tile h30v10 (Australia) for June 2008. The lower the value, the higher the probability of the pixel being burned.

The approach used to estimate the uncertainty was through a logistic regression model. To develop the model five global zones were used to generate unburned and burned samples from the FireCCI51 classification: Australia (h30v10), Canada (h11v03, h12v03), California (h08v05) and Angola (h19v10). These tiles were selected based on the diversity of vegetation found in them. 2008 was selected because is the reference year used to calibrate the algorithm.

The logistic regression was performed following a k-Fold validation (k = 10). The results are presented in the next formula:

$$p = \frac{1}{1 + e^c}$$

$$c = -(3.533 - 0.01175 * obs - 0.001996 * NIR + 0.01417 * NIR_{reldrop} - 0.0009282 * distance)$$

The rationale behind the NIR and distance variables sign is that the lower the value, the higher the probability of being burned. Also the relative NIR drop keeps the same relation with burned probability. However, the number of observations had a coefficient that means the inverse sense of the expected: the lower the number of observations, the higher the probability of burned.

The uncertainty is expressed by a value between 0 and 100 representing the probability of each pixel to being burned (Figure 24).
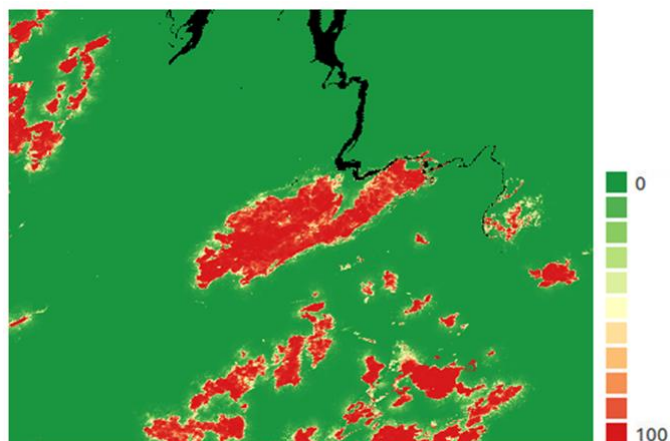
Figure 24: Uncertainty of an area of the tile h30v10 (Australia) for June 2008. The values represent the probability of each pixel being burned.

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 (Figure 25), 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.

(a)   (b)
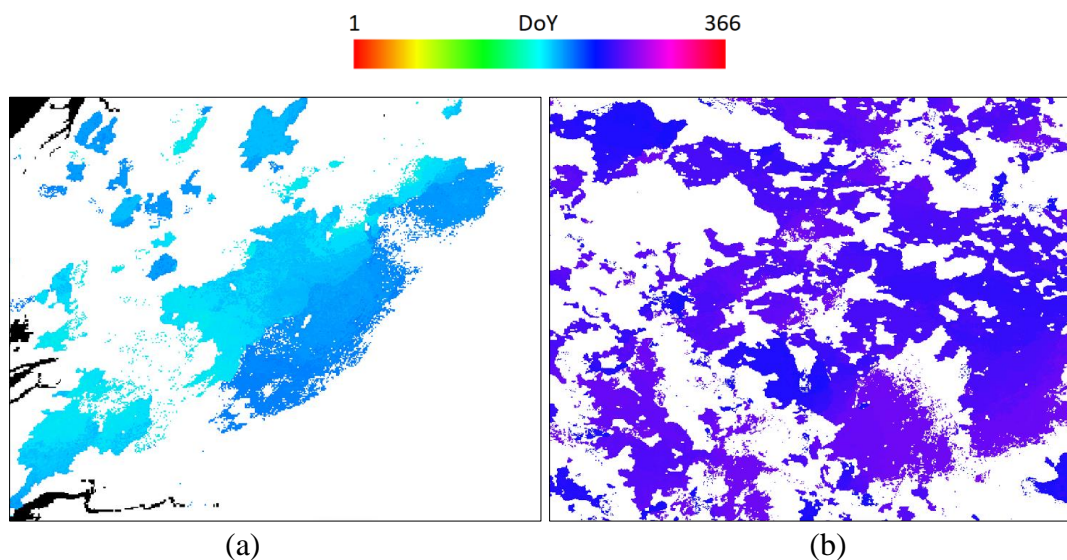
Figure 25: (a) Dates of detection of an area of tile h30v10 (North Australia) of July 2008. (b) Dates of detection of an area of tile h19v10 (Angola) of September 2008.

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 differs, 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 26 depicts the processing scenario.
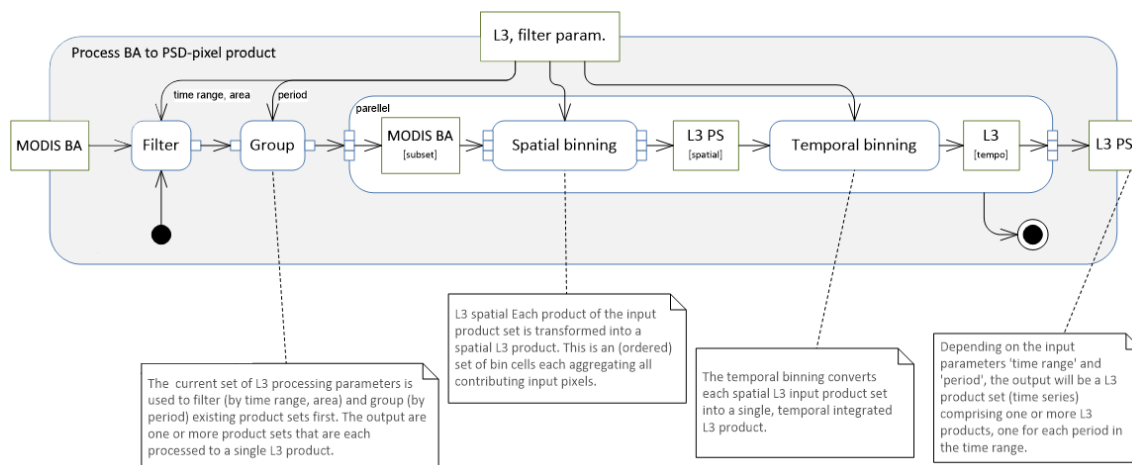
Figure 26: Calvalus L3 processing

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 and v5.1). 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 Section 2. 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 c:
                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 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 3.

```
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[1].

```
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:
```

---

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

```
        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
```

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.

Archibald, S., & Roy, D.P. (2009). Identifying individual fires from satellite-derived burned area data. In, 2009 IEEE International Geoscience and Remote Sensing Symposium (pp. III-160-III-163)

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., 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.

Chuvieco, E., Lizundia-Loiola, J., Pettinari, M.L., Ramo, R., Padilla, M., Tansey, K., Mouillot, F., Laurent, P., Storm, T., Heil, A., Plummer, S. (2018) Generation and analysis of a new global burned area product based on MODIS 250 m reflectance bands and thermal anomalies. Earth System Science Data 10, 2015-2031, https://doi.org/10.5194/essd-10-2015-2018

ESA. Land Cover CCI Product User Guide Version 2. Tech. Rep. (2017). Available at: http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

Giglio, L., Boschetti, L., Roy, D.P., Humber, M.L., & Justice, C.O. (2018). The Collection 6 MODIS burned area mapping algorithm and product. Remote Sensing of Environment, 217, 72-85

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.

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.

Lizundia-Loiola J., Pettinari M.L., Chuvieco E., Storm T., Gómez-Dans J. (2018) ESA CCI ECV Fire Disturbance: D2.1.3 Algorithm Theoretical Basis Document-MODIS, version 1.1. Available at: http://www.esa-fire-cci.org/documents

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.

Vermote, E.F., Roger, J.C., Ray, J.P. (2015) MODIS Surface Reflectance User's Guide, Collection 6. 35 pp. Available at: http://modis-sr.ltdri.org/guide/MOD09_ UserGuide_v1.4.pdf

Annex 1: Acronyms and abbreviations

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

Annex 2: Analysis of metrics for spatio-temporal HS clusters

Table 4: Commission error (Ce), omission error (Oe) and Dice Coefficient (DC) of different metric combinations. B refers to burned and U to unburned.

Tile	Error	MedB MeanU	MedB MedU	MedB MoU	MeanB MeanU	MeanB MedU	MeanB MoU
h30v10	Ce	12%	12.7%	13.6%	12.2%	13%	13.8%
	Oe	25.8%	23.1%	21.1%	25.4%	22.7%	20.7%
	DC	0.805	0.817	0.825	0.807	0.819	0.826
h11v03	Ce	84.6%	87.1%	86.9%	85.1%	86.5%	86.6%
	Oe	64.4%	63.9%	62.9%	63.7%	63.3%	62.4%
	DC	0.215	0.19	0.194	0.211	0.198	0.198
h12v03	Ce	10.8%	11.4%	11.9%	11%	11.5%	12.0%
	Oe	35.1%	31.4%	26.9%	34.2%	30.6%	26.0%
	DC	0.751	0.774	0.799	0.757	0.778	0.804
h08v05	Ce	28.2%	29.7%	31.5%	28.7%	30.1%	32.2%
	Oe	25.7%	24.4%	22.9%	25.8%	24.9%	24.3%
	DC	0.731	0.729	0.726	0.727	0.724	0.716
Global	Ce	12.6%	13.4%	14.3%	12.8%	13.6%	14.4%
	Oe	26.5%	23.7%	21.5%	26%	23.3%	21.2%
	DC	0.799	0.811	0.820	0.800	0.813	0.821

Annex 3: 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.

A3.1. 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[2]: (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:

---

[2] 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 27, 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.
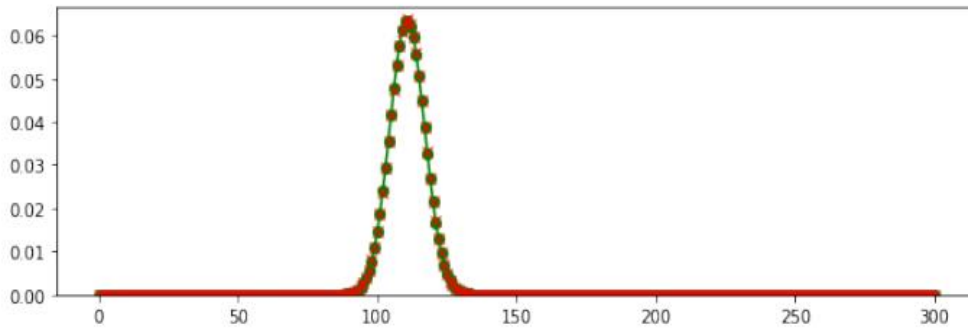

Figure 27: The Poisson binomial PDF (green line) derived from a simulated set of independent samples (300, 100 with probabilities between 0.7 and 0.9, 100 with probabilities between 0.2-0.3 and 100 with probabilities between 0-0.1). A Gaussian approximation (red line) derived from calculating the mean (~110) and standard deviation (~39) is also shown. Skewness was ~0.01.
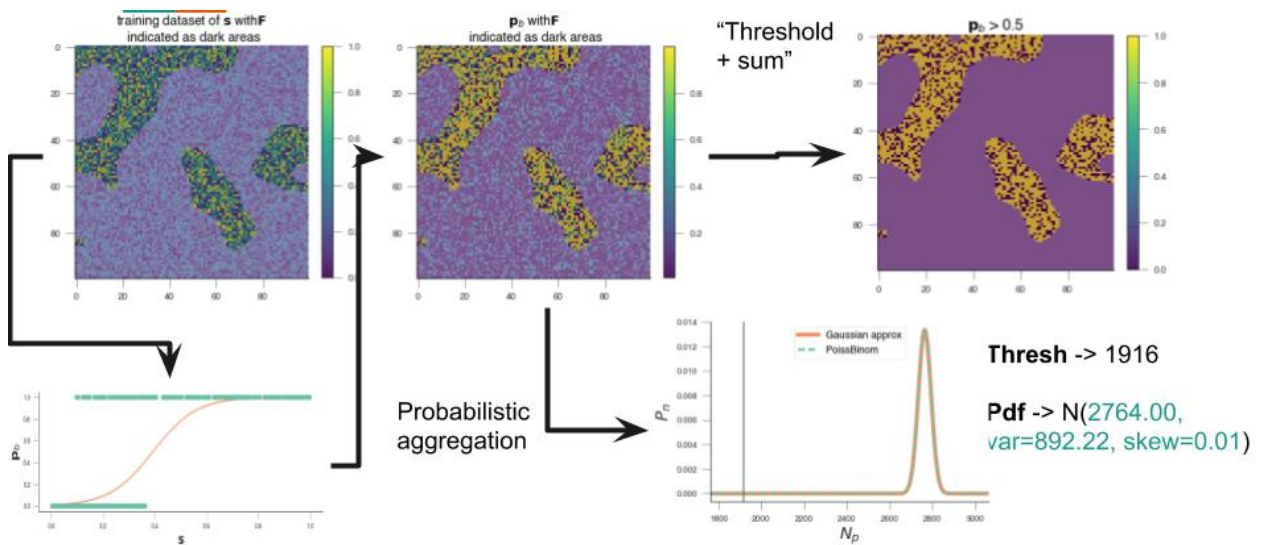

Figure 28: Example of aggregation. See text for more details.

Figure 28 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.

A3.2. 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 29), and then use that rescaled p_b to calculate the standard deviation and thus provide some form of uncertainty estimate.
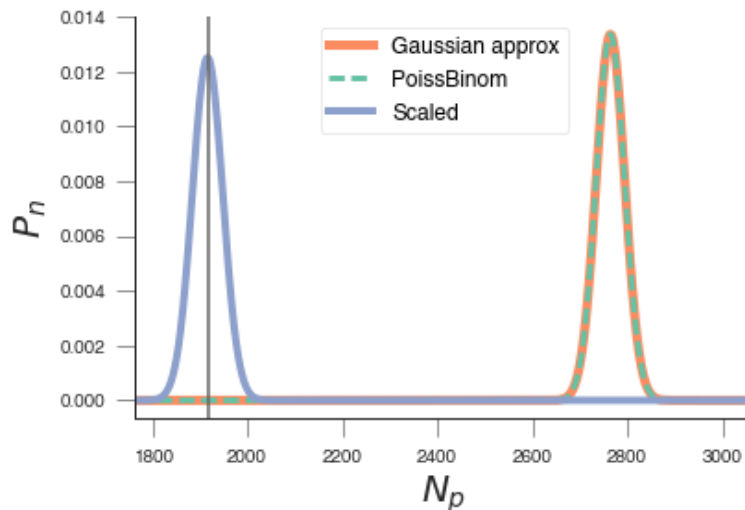

Figure 29: Example of applying the rescale approach to the example shown in Figure 28.