




ESA Climate Change Initiative – Fire_cci

Option 2 - Extending Fire Disturbance Time Series using AVHRR-LTDR

O2.D2 - Algorithm Theoretical Basis Document (ATBD) for AVHRR LTDR data

Project Name	ECV Fire Disturbance: Fire_cci Phase 2
Contract N°	4000115006/15/I-NB
Issue Date	13/06/2018
Version	1.1
Author	Gonzalo Otón, Emilio Chuvieco
Document Ref.	Fire_cci_ O2.D2 - Algorithm Theoretical Basis Document (ATBD) for AVHRR LTDR data
Document type	Public

*To be cited as: G. Otón, E. Chuvieco (2018) ESA CCI ECV Fire Disturbance: O2.D2
Algorithm Theoretical Basis Document (ATBD) for AVHRR LTDR data, version 1.1.
Available from: <http://www.esa-fire-cci.org/documents>*

	Fire_cci	Ref.:	Fire_cci_O2.D1_ATBD-AVHRR_v1.0		
	Algorithm Theoretical Basis Document – AVHRR LTDR data	Issue	1.1	Date	13/06/2018
		Page			2

Project Partners

Prime Contractor/ Scientific Lead & Project Management	UAH – University of Alcalá (Spain)
Earth Observation Team	UAH – University of Alcalá (Spain)
System Engineering	BC – Brockmann Consult (Germany)
Climate Research Group	MPIM – Max Planck Institute for Meteorology (Germany) UCLo – Université Catholique de Louvain (Belgium)



Max-Planck-Institut
für Meteorologie



Distribution

Affiliation	Name	Address	Copies
ESA	Stephen Plummer (ESA)	stephen.plummer@esa.int	electronic copy
Project	Emilio Chuvieco (UAH)	emilio.chuvieco@uah.es	electronic copy
Team	M. Lucrecia Pettinari (UAH)	mlucrecia.pettinari@uah.es	
	Gonzalo Otón (UAH)	gonzalo.oton@uah.es	
	Thomas Storm (BC)	thomas.storm@brockmann-consult.de	
	Guy Brasseur (MPIM)	guy.brasseur@mpimet.mpg.de	
	Idir Bouarar (MPIM)	idir.bouarar@mpimet.mpg.de	
	Angelika Heil (MPIM)	a.heil@mpic.de	
	Pierre Defourny (UCLo)	pierre.defourny@uclouvain.de	
	Céline Lamarche (UCLo)	celine.lamarche@uclouvain.de	

Summary

This document describes the algorithm used to generate the long-term BA time series derived from the AVHRR (LTDR, Long Term Data Record) dataset to extend the BA product of the Fire_cci contract baseline.

	Affiliation/Function	Name	Date
Prepared	UAH UAH	Gonzalo Otón Emilio Chuvieco	12/06/2018
Reviewed	UAH – Project Manager	Lucrecia Pettinari	13/06/2018
Authorized	UAH - Science Leader	Emilio Chuvieco	13/06/2018
Accepted	ESA - Technical Officer	Stephen Plummer	13/06/2018

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

Document Status Sheet

Issue	Date	Details
1.0	13/04/2018	First release of the document.
1.1	13/06/2018	Addressing comments of ESA-CCI-FIRE-EOPS-MEM-18-0146

Document Change Record

Issue	Date	Request	Location	Details
1.1	3/06/2018	UAH	Page 3	Updated list of Project Team
		ESA	Sections 3.1, 3.2, 5, 6	Minor changes in the text to clarify the information.
			Figures 7, 8 and 9	Caption updated.
			Figures 10 and 11	Figures updated.

Table of Contents

1	Executive Summary	5
2	Introduction	5
2.1	Background	5
2.2	Applicable Documents	6
3	Input data	6
3.1	Characteristics of the LTDR dataset	6
3.2	LTDR versions	7
3.3	Auxiliary information	9
4	General scheme of the LTDR BA algorithm	10
5	Creating multitemporal composites	11
6	Developing a synthetic burned area index	12
7	Classification stage	15
8	Calculation of proportion of burned area	17
9	Uncertainty characterization	19
10	Validation	19

11 References21
Annex 1: Acronyms and abbreviations24

List of Tables

Table 1: Cross tabulation of LTDR results and MCD64A1 for January 2008..... 19
Table 2: Cross tabulation of LTDR results and MCD64A1 for July 2008 20

List of Figures

Figure 1: Availability of LTDR v4.0 daily images for the full time series..... 7
Figure 2: Temporal series of NIR values for a stable pixel of the LTDR time series (v4) 8
Figure 3: Temporal series of B4 brightness temperature values for a stable pixel of the LTDR time series (v4)..... 8
Figure 4: availability of LTDR v5.0 daily images for the full time series 9
Figure 5: Temporal series of NIR values for a stable pixel of the LTDR time series (v5) 9
Figure 6: General scheme of the LTDR BA algorithm 10
Figure 7: An example of an LTDR monthly composite from 2008. Left, Day 200; Right, days 183-213. RGB =NIR, RED, RED, white shown non-burnable areas. 11
Figure 8: Box plot distributions of the LTDR BA index for all pixels with different proportions of BA for January 2008..... 13
Figure 9: Same as Figure 8 but for July 2008..... 13
Figure 10: Monthly trajectories of the LTDR BA index for unburned and burned pixels in January in Africa 14
Figure 11: Monthly trajectories of the LTDR BA index for unburned and burned pixels in July in Africa 15
Figure 12: RF BA classification of LTDR data for January 2008 16
Figure 13: RF BA classification of LTDR data for July 2008 17
Figure 14: Percentage BA of LTDR data for January 2008 18
Figure 15: Percentage BA of LTDR data for July 2008..... 18
Figure 16: Uncertainty of LTDR data for January 2008 19
Figure 17: Uncertainty of LTDR data for July 2008 19

	Fire_cci	Ref.:	Fire_cci_O2.D1_ATBD-AVHRR_v1.0		
	Algorithm Theoretical Basis Document – AVHRR LTDR data	Issue	1.1	Date	13/06/2018
		Page	5		

1 Executive Summary


The Fire_cci project aims to obtain the longest possible time series of global BA products from existing satellite data (AD-1). Most climate modellers emphasize the importance of having long time and consistent series of ECV (RD-2). The bases have been European sensors, but others have also been considered to extend the length or the quality of the data series. One of the options approved in Phase 2 of the project aimed to extend the available time series of BA products back to the early 80's, using existing archives of AVHRR-LTDR data (RD-1). This document describes the technical steps to obtain BA information from LTDR. Considering the particular spatial and spectral resolution of this dataset, a new algorithm has to be developed, which is not similar to others used for the Fire_cci product, mainly because the LTDR data cannot be accurately linked to active fire detection. Therefore, we separate from hybrid algorithms, which have been the main basis for those developed in other tasks of the project. Instead, we use multitemporal time series and dedicated spectral indices that maximize the burn signal with the coarse LTDR pixel (5x5 km).

2 Introduction

2.1 Background

The ESA Fire_cci project focuses on developing and validating algorithms to meet GCOS ECV requirements for (consistent, stable, error-characterized) global satellite data products from multi-sensor data archives. Fire modellers require precise information on the timing, size, intensity and spatial distribution of fires (Hantson et al. 2016). Satellite burned area products have been generated in the last years based on existing global observation sensors (Mouillot et al. 2014; RD-2), such as MODIS, VEGETATION, ATSR and MERIS data. These products only cover from 2000 and all show inconsistencies and omission and commission errors (Granier et al. 2011; Padilla et al. 2015). Most papers dealing with fire models emphasize the need to extend the existing time series of burned area information backwards. Although the GFEDv4 includes BA trends from 1995, the authors recognize greater uncertainties for the pre-MODIS era (Giglio et al. 2010). Estimations for previous periods were based on fire ignition and propagation models (Thonicke et al. 2001) or on fire occurrence reports provided by some national fire services (Mouillot and Field 2005). These data were complemented with auxiliary sources, such as journalist news, literature sources or charcoal deposits in lakes (Power et al. 2010).

The alternative to extend the BA datasets before the MODIS era is the NOAA-AVHRR (Advanced Very High Resolution Radiometer) archive, dated back to 1981. Unfortunately, only a degraded version of the AVHRR data was archived globally. It was named GAC (4x4 km), obtained through a spatial sampling of the High Resolution data (HRPT, 1x1 km approximately at nadir). From the GAC data, NASA has released a time series of AVHRR data, with adequate geometric and radiometric corrections to make it useful for time series analysis. Since the HRPT data were archived only at particular regions and need considerable pre-processing effort, it was considered that the LTDR was the most adequate time series of AVHRR datasets to feed a BA algorithm that would extend the BA time series of the Fire_cci baseline contract.

	Fire_cci	Ref.:	Fire_cci_O2.D1_ATBD-AVHRR_v1.0		
	Algorithm Theoretical Basis Document – AVHRR LTDR data	Issue	1.1	Date	13/06/2018
		Page			6

2.2 Applicable Documents

(AD-1)	ESA Climate Change Initiative (CCI) Phase 2 Statement of Work, prepared by ESA Climate Office, Reference CCI-PRGM-EOPS-SW-12-0012, Issue 1.3, date of issue 24 March 2015
(RD-1)	ECV Fire Disturbance Proposal for Option 2: Extending Fire Disturbance Time Series using AVHRR-LTDR.
(RD-2)	Heil A., Yue C., Mouillot F. and Kaiser J.W. (2017) ESA CCI ECV Fire Disturbance: D1.1 User requirement document, version 5.2.
(RD-3)	Pettinari M.L., Otón G., Chuvieco E. (2017) ESA CCI ECV Fire Disturbance: O2.D1 User Requirement Document and Product Specification Document for AVHRR, v1.1.

3 Input data

3.1 Characteristics of the LTDR dataset

NASA created a global-full time series of AVHRR data, at 0.05° resolution (approx. 5 km), named the long-term data record (LTDR) dataset. There are other long-term coarse AVHRR data series (GIMMS or Pathfinder), whose strengths and weaknesses have been cross validated, as reported in the literature (Alcaraz-Segura et al. 2010; Tucker et al. 2005). As a result of those analyses the AVHRR LTDR data series was chosen as the most adequate input for extending the Fire_cci BA datasets back to early 80's. NASA's LTDR also includes data from MODIS instruments to obtain a long-term consistent time series of surface reflectance and temperature.

LTDR Version 5 is the latest product available. The pre-processing includes radiometric in-flight vicarious calibration for the visible and near infrared channels and inverse navigation to relate an Earth location to each sensor instantaneous field of view (IFOV). Atmospheric corrections for Rayleigh scattering, ozone, and water vapour are undertaken, with aerosol correction being implemented.

The AVHRR-LTDR V5 products for the 1981-2018 time period are:

- Surface reflectance:
 - SREFL_CH1 (0.5 – 0.7 μm)
 - SREFL_CH2 (0.7 – 1.0 μm)
 - SREFL_CH3 (3.55 – 3.93 μm)*
- TOA brightness temperature:
 - BT_CH3 (3.55 – 3.93 μm)*
 - BT_CH4 (10.3 – 11.3 μm)
 - BT_CH5 (11.5 – 12.5 μm)
- View zenith angle (VZEN), Solar zenith angle (SZEN).
- Relative azimuth (RELAZ).
- Quality Assessment Field (QA).
- Fill value (No Data) = -9999
- Scale Factor:
 - Surface reflectance = 0.0001
 - TOA brightness temperature = 0.1
 - Zeniths and Azimuth = 0.01

- Spatial reference:
 - GCS datum based upon the Clarke 1866 ellipsoid.
 - EPSG: 4008
 - Datum: Based on Clarke 1866 spheroid
- Satellites used to generate the LTDR dataset are the following:
 - N07: 1981 – 1985
 - N09: 1985 – 1988
 - N11: 1988 – 1994
 - N14: 1995 – 2000
 - N16: 2000 – 2007
 - N18: 2005 – 2009
 - N19: 2009 – 2018

(*) From 2000 to 2003, the channel 3 was changed from 3.55 – 3.93 μm to 1.58 - 1.64 μm for day time acquisitions.

3.2 LTDR versions

The first analyses to create the LTDR algorithm were based on v4 of this product, available when Option 3 started. We found several problems with this version, but since it was the best available we started to work with this dataset. The main issues were related to the gaps in the temporal series, particularly in 1994 and 2000, with many other shorter gap periods (Figure 1).

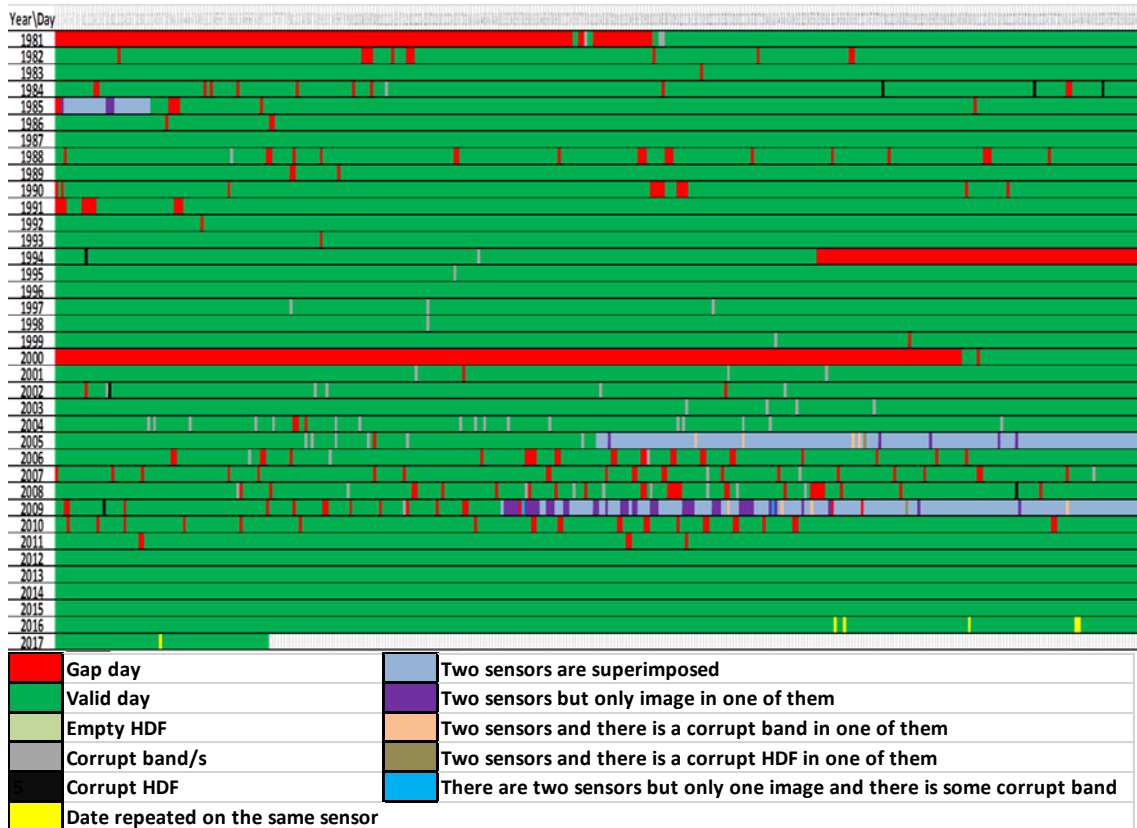


Figure 1: Availability of LTDR v4.0 daily images for the full time series

Another problem was related to the lack of temporal consistency of the time series, particularly in the calibration of the near infrared reflectance (channel 2) and visible (channel 1), which had a discontinuity around the year 2000. Figure 2 includes a time series of a pixel extracted from the Sahara desert after filtering out the cloudy or poor



quality observations. The reduction of NIR values around the 2000 is evident, which may be caused by a lack of proper calibration between versions 2 and 3 of the AVHRR sensor (P. Lewis, personal communication, 2017). The problem was not observable in the brightness temperatures of bands 4 and 5 (Figure 3).

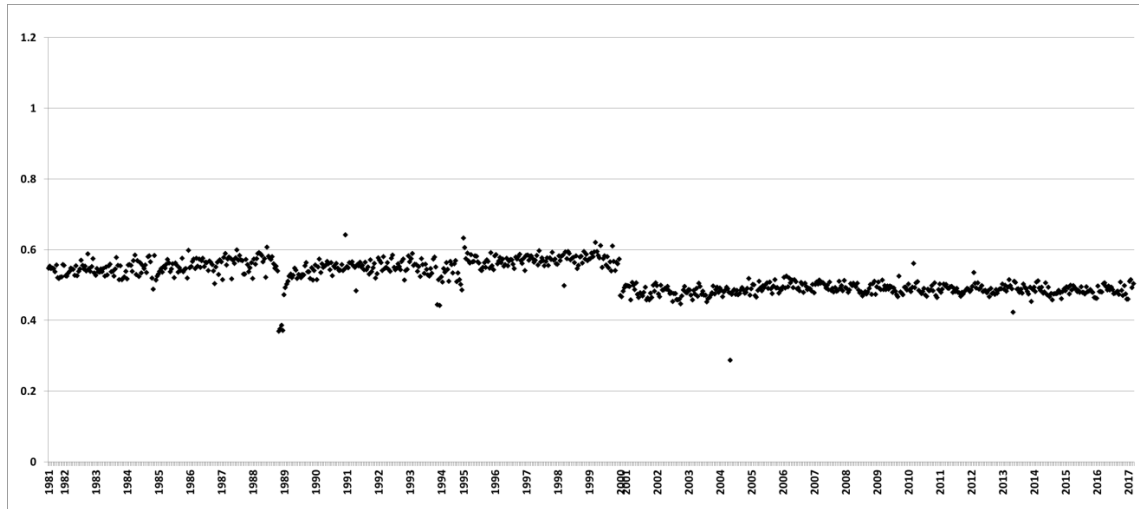


Figure 2: Temporal series of NIR values for a stable pixel of the LTDR time series (v4)

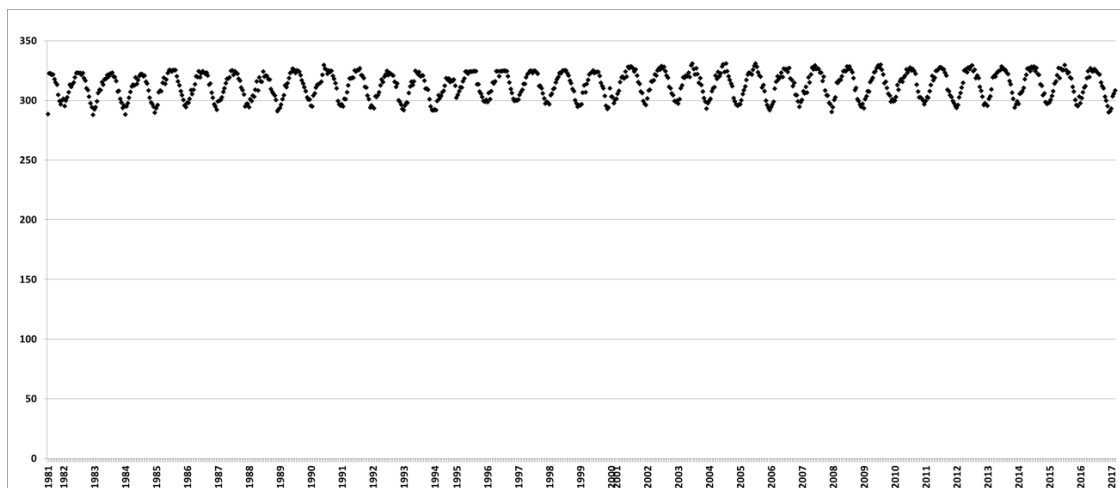


Figure 3: Temporal series of B4 brightness temperature values for a stable pixel of the LTDR time series (v4)

For this reason, we initially decided to use 2008 as a test year to develop the algorithm, as it was not affected by the NIR calibration problem, and had reference datasets derived from Fire_cci Phase 1 for parameterizing the algorithm. However, in August 2017 NASA released a new version of the LTDR product, named v5, which apparently solves the two main issues observed in v4. The temporal gaps were greatly reduced (Figure 4), and now only 1981 and 1994 are really problematic to derive BA products. Therefore, we used LTDR v5 for developing the BA algorithm. The time series for processing will start from 1982 and end up in 2017.

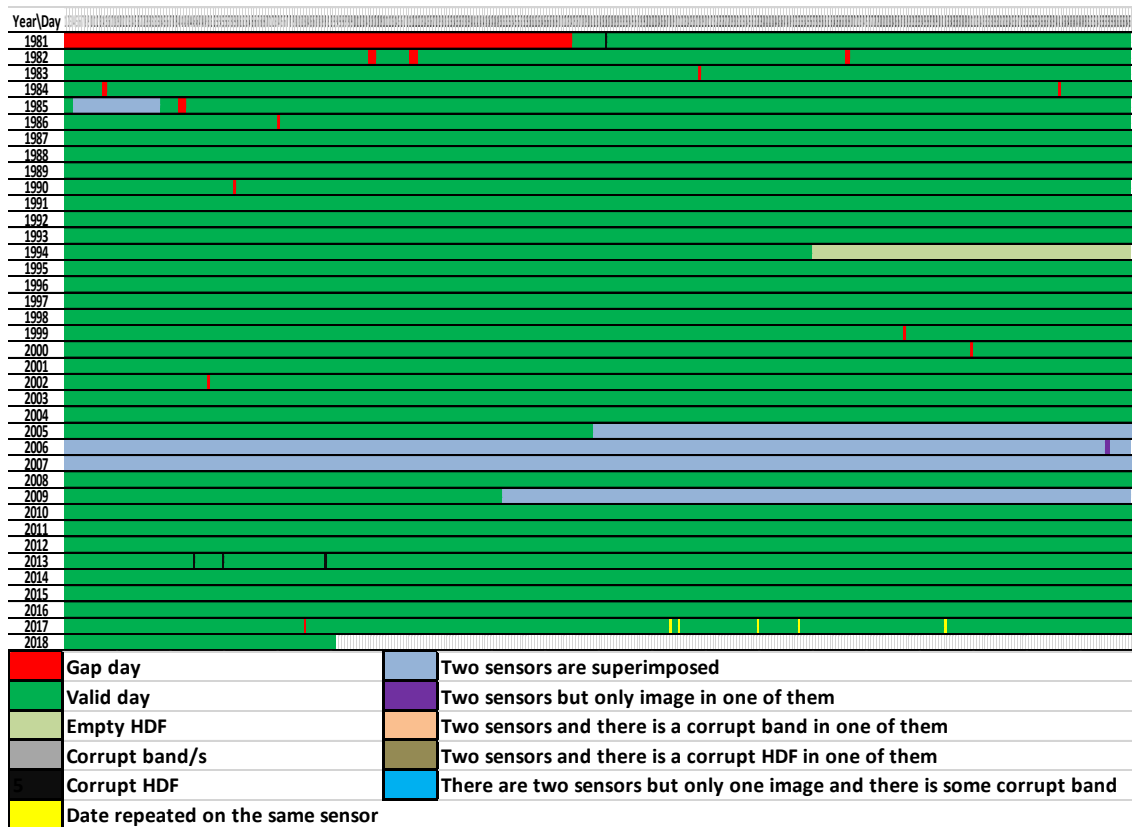


Figure 4: availability of LTDR v5.0 daily images for the full time series

The intercalibration problems with the RED and NIR bands have also been apparently solved as can be observed in the time series of the same pixel used in Figure 2, where the continuity between AVHRR2 and AVHRR3 acquisitions is greatly improved (Figure 5).

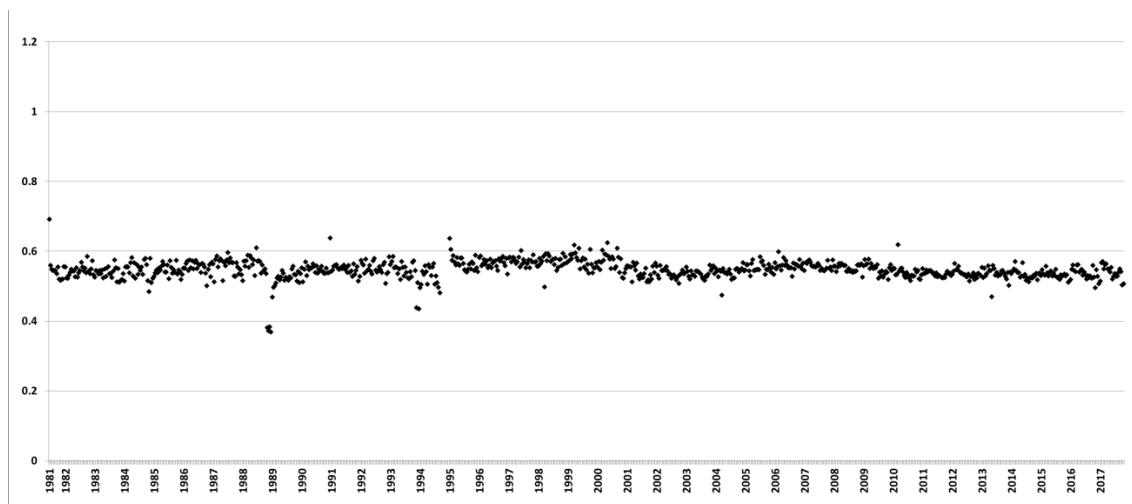


Figure 5: Temporal series of NIR values for a stable pixel of the LTDR time series (v5)

Further information on the product is available at <https://ltdr.modaps.eosdis.nasa.gov/cgi-bin/ltdr/ltdrPage.cgi> (last access April 2017).

3.3 Auxiliary information

In addition to LTDR data, other auxiliary datasets were collected to train and validate the algorithm. A land cover map was used to mask out non burnable classes, as well as

to report burned land cover in the PSD product. The source was ESA’s CCI Land Cover v1.6.1 (LC_cci, Kirches et al. 2013, <http://www.esa-landcover-cci.org/>, last access September 2017) product corresponding to year 2000, as it is the year of the middle of the time series.

For training and validating the algorithm we used the MCD64A1 C6 dataset (Giglio et al. 2009), now the standard BA product of NASA. It has been shown to be the most accurate among the existing global products (Padilla et al. 2015).

4 General scheme of the LTDR BA algorithm

The BA algorithm needed to be adapted to the specific characteristics of LTDR data, which were quite different from those previously used for Fire_cci project. Two main difficulties arose when designing the algorithm: firstly, active fires were not available, and therefore the hybrid algorithms used for MERIS, MODIS and Sentinel 2 could not be adapted to this dataset; secondly, the pixel size is considerable larger (5x5 km, versus 250x250 of MODIS or 300x300 m of MERIS), and therefore much larger confusions with other spectrally similar covers to BA were expected.

The general scheme of the BA algorithm is included in Figure 6. Daily images were converted to monthly temporal composites to obtain from them a list of spectral indices closely related to BA (extracted from literature and our previous experience). Calibration of these indices was helped by locating specific areas where fire perimeters were available (from the Landsat validation archive). Then, a synthetic BA model was developed based on biophysical relations of fire with those spectral indices. Monthly summaries of this variable were computed for a single year (2008) used to calibrate the model, which use a sample of MCD64 c6 data as the target variable (assuming it was the ground truth, which is obviously not the case, but it has been documented as the most accurate BA product from those globally available). The calibration of the model was tested with several classification algorithms: Random Forest and Logistic Regression were the most used. Only the best results of the model are reported in this document.

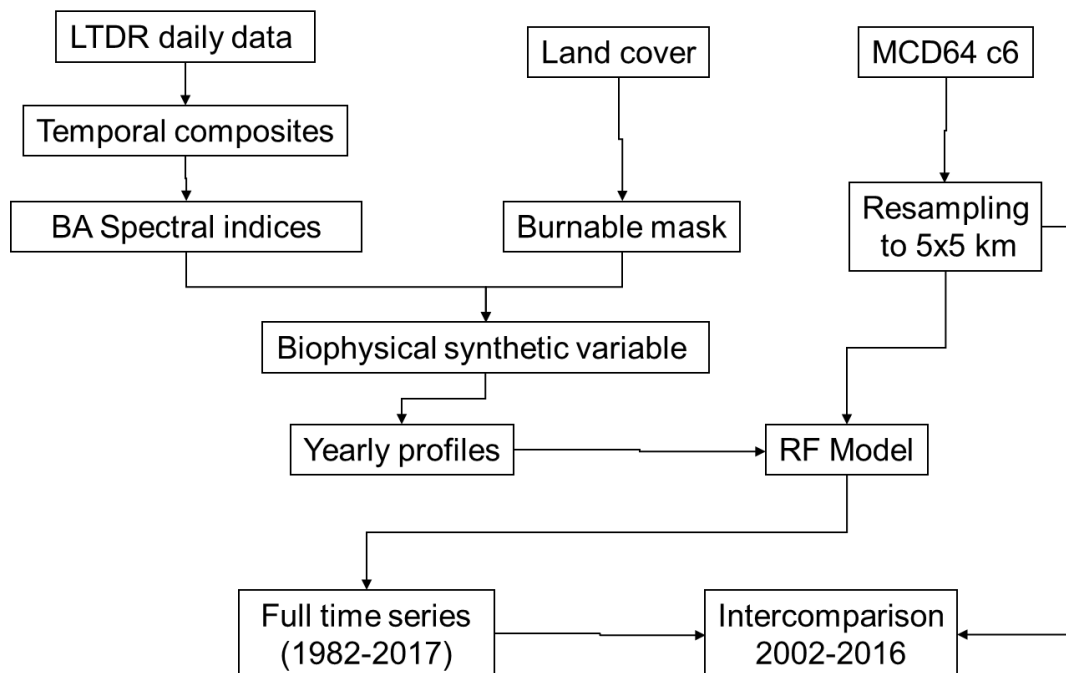


Figure 6: General scheme of the LTDR BA algorithm



5 Creating multitemporal composites

The first step was to check the input images, to assure that they included all required bands and files were not corrupted. Problematic files were discarded from further processing. Whenever more than one file existed in a single date, both were included in the processing. Every daily image had 3600 rows and 7200 columns. The whole LTDR dataset covered the period from 1981 to 2017, but some gaps were detected in 1994.

LTDR data include daily images. However, daily data are difficult to use as input for BA detection as observations may be cloudy, affected by BRDF effects or have incorrect observations. For this reason temporal composites have been commonly used in BA classification (Martín and Chuvieco 1993; Pereira et al. 1994; Razafimpanilo et al. 1995). Several techniques have been tested by different authors (Adair et al. 2002; Barbosa et al. 1998; Cabral et al. 2003; Cihlar et al. 1999; Duchemin et al. 1999), being quite common the maximum value of the NDVI for the target period of compositing. However, even though this criterion has proved to be very useful for vegetation trend monitoring, it is not adequate for BA detection, as the maximum NDVI values tend to select unburned or even cloudy pixels over burned ones within the compositing period (Chuvieco et al. 2005a). For this reason, alternative criteria were tested. When comparing 8 compositing techniques and using 4 indicators of quality, the maximum temperature of AVHRR channel 4 was selected as the most adequate compositing criteria for BA detection from AVHRR (Chuvieco et al. 2005b). Following results from those papers, LTDR monthly composites were created from daily data, by selecting for each pixel the day with maximum monthly temperature, as it provided the best equilibrium between near nadir and recently burned observations. All other bands were extracted from that day to create the composite, and the information of the date was saved for the post-processing stage. Pixels with a very low proportion (<20%) of burnable categories were discarded. They include water, desert and urban areas. The burnable information was taken from the Land Cover CCI product, which was resampled to the LTDR pixel size of 5x5 km. Figure 7 includes an example of a daily image and a monthly composite of July, 2008.

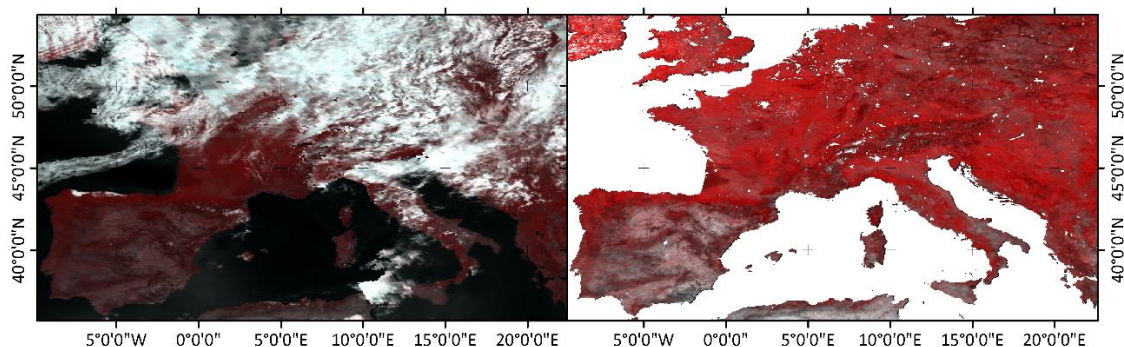


Figure 7: An example of an LTDR monthly composite from 2008. Left, Day 200; Right, days 183-213. RGB =NIR, RED, RED, white shown non-burnable areas.

Once the monthly composites were created, pixels with anomalous data were discarded as well. They were identified as pixels with clouds (R and NIR values > 90%), or with notorious artefacts (for instance when $R > NIR$ reflectance in vegetated areas). Unfortunately, the Quality Flags of the LTDR dataset are not very accurate, and we decided not to use them as they flagged out many good observations. Non observed areas were given a dedicated code to be used for computing the PSD's fraction of observed area layer.



6 Developing a synthetic burned area index

After different tests with the original bands, we decided to base the discrimination of burned areas on a synthetic biophysical index, built on the previous experiences of mapping BA from AVHRR images (Carmona-Moreno et al. 2005; Chuvieco et al. 2008; Chuvieco et al. 2005b; Fraser et al. 2004; Maggi and Stroppiana 2002; Martín et al. 2002; Moreno Ruiz et al. 2014; Pu et al. 2007; Riaño et al. 2007; Sukhinin et al. 2004). The 2008 year was used to calibrate this index, taking the MCD64A1 c6 images as a reference BA set. Input bands were brightness temperatures from band 5, the RED and NIR reflectances, the GEMI and BAI spectral indices and the temporal differences of those variables. The variables that showed clear tendencies among different proportions of BA area (0%, 1-20%, 21-40%, etc.) were finally selected.

All input variables were normalized using the mean and standard deviation of the available pixels in each monthly composite:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Where x is the monthly pixel value, μ is the global mean and σ the global standard deviation of each variable. This was done to balance the different layers and give them a similar importance in the final index.

The final biophysical LTDR BA index was defined as:

$$\begin{aligned} \text{LTDR BA index} = & \frac{T5 - \text{mean } T5}{\text{std } T5} - \frac{T5_{\text{diff}} - \text{mean } T5_{\text{diff}}}{\text{std } T5_{\text{diff}}} - \frac{\text{Red} - \text{mean Red}}{\text{std Red}} + \\ & \frac{\text{Red}_{\text{diff}} - \text{mean Red}_{\text{diff}}}{\text{std Red}_{\text{diff}}} + \frac{\text{NIR} - \text{mean NIR}}{\text{std NIR}} + \frac{\text{NIR}_{\text{diff}} - \text{mean NIR}_{\text{diff}}}{\text{std NIR}_{\text{diff}}} + \frac{\text{GEMI} - \text{mean GEMI}}{\text{std GEMI}} + \\ & \frac{\text{BAI} - \text{mean BAI}}{\text{std BAI}} + \frac{\text{BAI}_{t+1} - \text{mean BAI}_{t+1}}{\text{std BAI}_{t+1}} \end{aligned} \quad (2)$$

where

$T5$ = TOA brightness temperature of Channel 5 (~11.5 – 12.5 μm)

$T5_{\text{diff}} = T5_{t-1} - T5_t$, where t is the monthly composite being analyzed, and $t-1$ is the composite of the previous month.

Red = Surface reflectance of Channel 1 (0.5 – 0.7 μm)

$\text{Red}_{\text{diff}} = \text{Red}_{t-1} - \text{Red}_t$

NIR = Surface reflectance of Channel 2 (0.7 – 1.0 μm)

$\text{NIR}_{\text{diff}} = \text{NIR}_{t-1} - \text{NIR}_t$

GEMI, the Global Environmental Monitoring Index (Pinty and Verstraete 1992) defined as:

$$\text{GEMI} = \eta * (1 - 0,25\eta) - \frac{\rho_R - 0,125}{1 - \rho_R} \quad (3)$$

Where

$$\eta = \frac{2 * (\rho_{IR}^2 - \rho_R^2) + (1,5 * \rho_{IR}) + (0,5 * \rho_R)}{\rho_{IR} + \rho_R + 0,5} \quad (4)$$

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

BAI, the Burned Area Index (Martín and Chuvieco 1998) is a non-linear index specifically defined for burned area discrimination in AVHRR images. It was defined as



the inverse spectral distance to a convergence point, defined by the minimum reflectance of burned vegetation in the near infrared and the maximum in the red bands:

$$BAI=1 /((\rho_{IR}-\rho_{cIR})^2 + (\rho_R-\rho_{cR})^2) \tag{5}$$

where ρ_{cIR} and ρ_{cR} are the convergence values for burned vegetation (defined for AVHRR as 0.06 and 0.1, respectively).

BAI_{t+1} is the BAI of the posterior monthly composite.

The distribution function of the LTDR BA index shows good global relation with the different proportions of BA, considering the MCD64A1 as the reference dataset.

Figure 8 and Figure 9 show for two contrasting months (January and July) the box plot distribution of all pixels having different proportions of BA, from unburned (BA0) to >90% burned (BA9). These distributions show the sensitivity of this index to detect different proportions of burned area, even though great dispersion exists in all ranges.

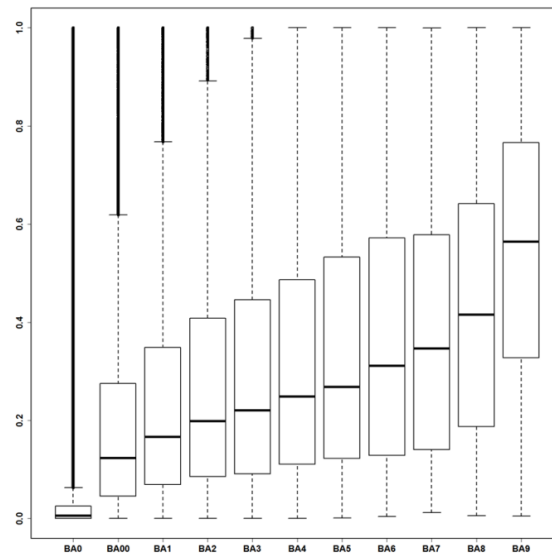


Figure 8: Box plot distributions of the LTDR BA index for all pixels with different proportions of BA for January 2008.

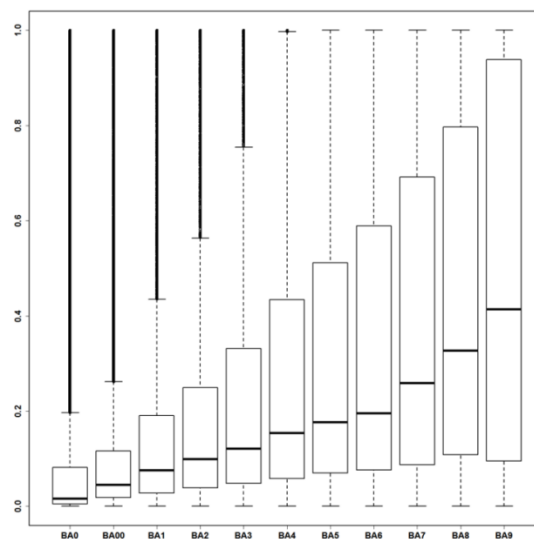


Figure 9: Same as Figure 8 but for July 2008.



Considering this wide dispersion of LTDR BA index values, we also included in the discrimination the temporal variation of this index. It is well known that spectral values similar to BA may be obtained from different phenomena (crop rotations, particularly in dark soils; seasonal floods; deforestation...). To minimize these confusions the temporal evolution of each year was taken into account as an additional discrimination variable.

Figure 10 and Figure 11 show three examples of burned and unburned pixels for two contrasting months (January and July). It is observed that burned pixels show a more divergent seasonal trend than unburned pixels that may offer medium LTDR BA index values in some months but do not offer such a contrast post-fire trend.

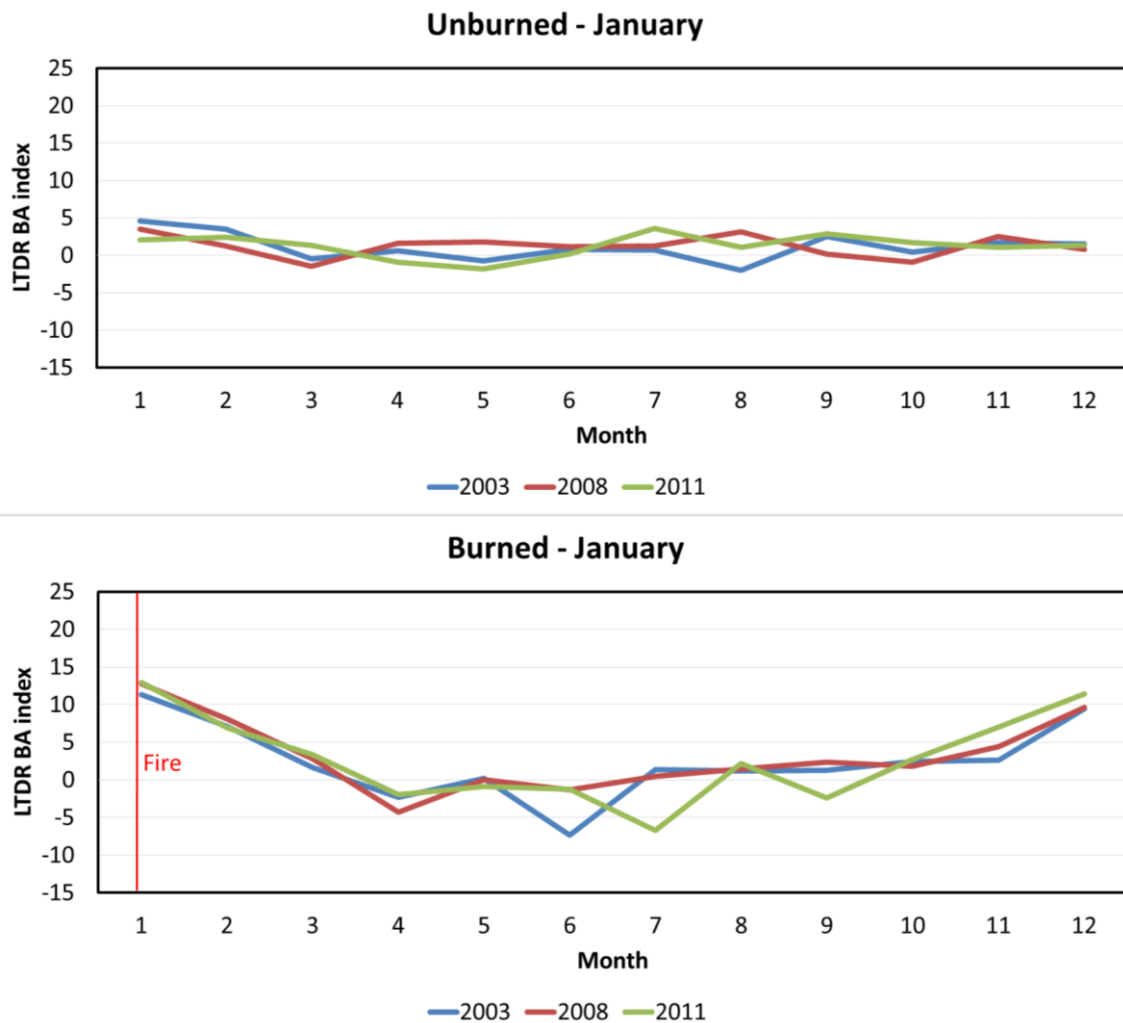


Figure 10: Monthly trajectories of the LTDR BA index for unburned and burned pixels in January in Africa

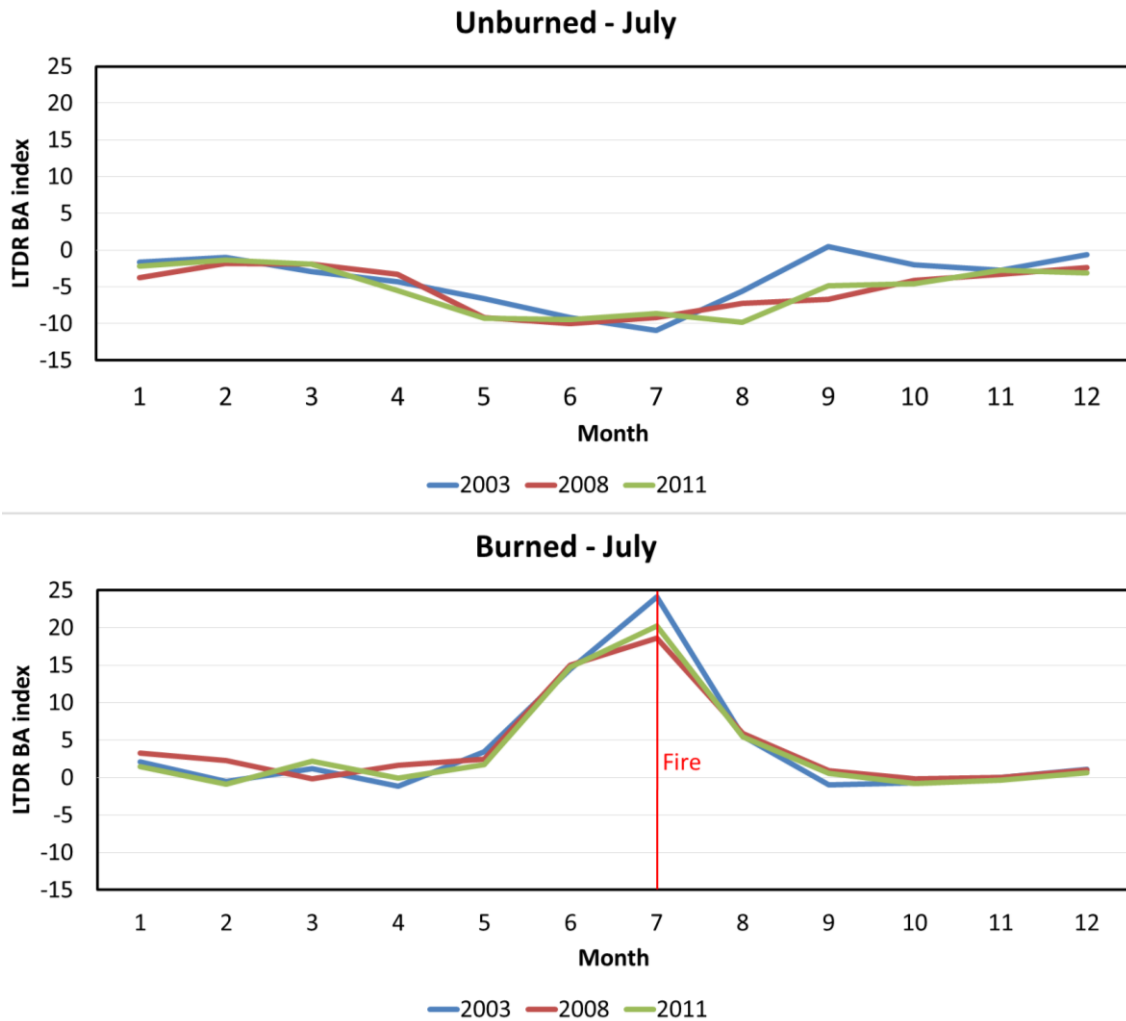


Figure 11: Monthly trajectories of the LTDR BA index for unburned and burned pixels in July in Africa

7 Classification stage

The classification models were based on the Random Forest (RF) algorithm. This approach has been widely used in the last decades for land cover mapping (Rodriguez-Galiano et al. 2012), although much less experience is available on their use in BA mapping. A recent study showed their potentials for mapping BA using MODIS 500 m images (Ramo and Chuvieco 2017). Based on this experience, we computed monthly RF models, using as inputs the LTDR BA index for several years when both LTDR and MCD64A1 BA product were available. As in the first runs we found that RF showed overtraining bias when using only a single year to calibrate the models, we used eight years instead: 2001, 2003, 2004, 2006, 2008, 2011, 2013 and 2015. We selected years with above average BA occurrence (2004 and 2011), one with lower occurrence (2013), and the rest having normal BA occurrence.

After testing different options, we finally selected a binary classification model (burned/unburned), considering a burned pixel as all that have burned area in the reference product (assuming again as reference the MCD64A1 information). Other models with several BA proportion intervals showed severe overtraining or did not

provide accurate results. Later, this assignment is corrected by computing matrix metrics with proportion of burned for those pixels detected as burned.

Following previous experiences (Ramo and Chuvieco 2017) we used 600 trees to build the models. For each tree, a 10% proportion of BA pixels were forced to be included. Output of the RF models provided a BA probability value. We compared this value with MCD64A1 c6 BA data to estimate omission and commission errors (again assuming this product as the reference BA). This was the most practical choice, as deriving a global dataset of Landsat data for the whole study period was not feasible (the one available for the project was to be used for final accuracy assessment).

The criteria to select the final probability threshold for RF classification was based on the Dice Coefficient (DC), previously used for BA assessment within the Fire_cci project (Padilla et al. 2015). The analysis was done with the whole common time series between the MCD64A1 and LTDR (2001 to 2015). The median monthly value of the RF probability providing the best DC output for the full time series was used as a threshold for the final RF classification (January 0.54 and July 0.56). The resulting files include four categories: non burnable (-2), no data (-1), unburned (0) and burned date (1-365). Figure 12 and Figure 13 show two examples of the classification results for two months in 2008. Spatial trends are reasonable, with larger fire occurrence in the Northern Hemisphere tropical regions for January and the Southern Hemisphere for July, along with some areas in central Asia and Tropical South America. Areas above 60d N are not included in the LTDR database.

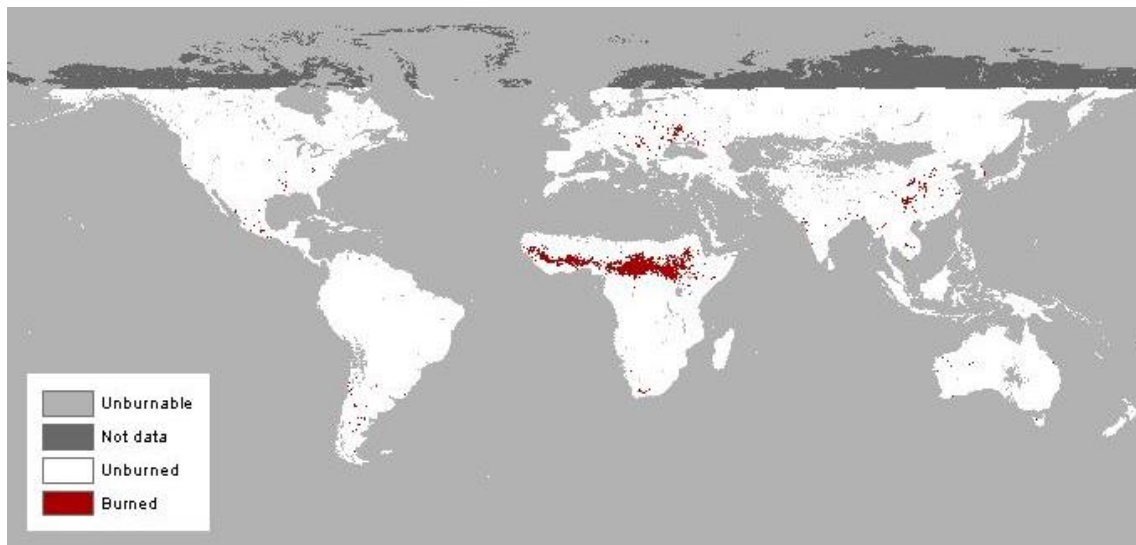


Figure 12: RF BA classification of LTDR data for January 2008

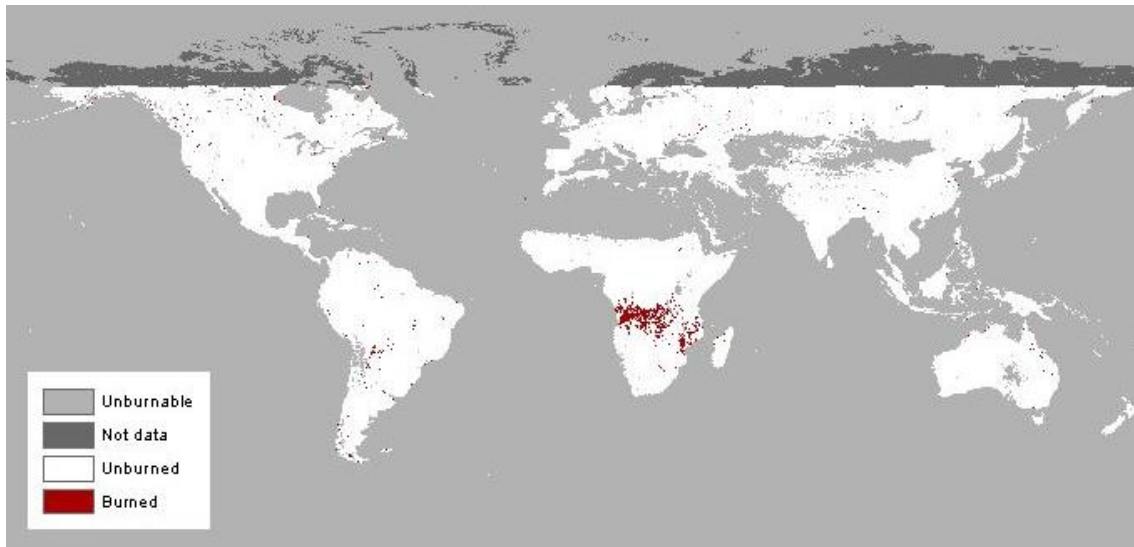


Figure 13: RF BA classification of LTDR data for July 2008

8 Calculation of proportion of burned area

Following the PSD of the LTDR option (RD-3), the final dataset will be publicly released as a grid product in NetCDF format, including the standard variables of the other Fire_cci datasets. To calculate the total BA of each cell, it would be risky to assume that the whole LTDR pixel is burned, since the training was done using cells with different proportions of burned area according to MCD64A1. For this reason, we have analysed the statistical distribution of BA proportions in each of the pixels detected as burned in different monthly RF models versus MCD64. The analysis was done separately for the different monthly time series (January, February, etc.), as we assume the BA proportions may change throughout the year.

The analysis began by grouping the burned data of the MCD64 by pixel of each specific month of the whole time series. The process has been made under the parameters of the final product where a 0.25° grid is offered. The pixels were grouped to offer that resolution, decreasing the rows and columns by 5 times. Within these pixels, the burned area was calculated. The method selected was the one able to adjust in each pixel, the LTDR burned area percentage that gives the closest value to the burned area offered by the MCD64. The data obtained from percentages of each month were combined with the results of the classification, in order to give a value to each burned pixel (Fig. 14, 15). In case that the MCD64 does not offer a burned pixel, that pixel cannot be burned in the LTDR either, but it will keep a high uncertainty. Also, another layer was created where those pixels finally burned were assigned burned date, which corresponds with the date of the pixel in the composite.

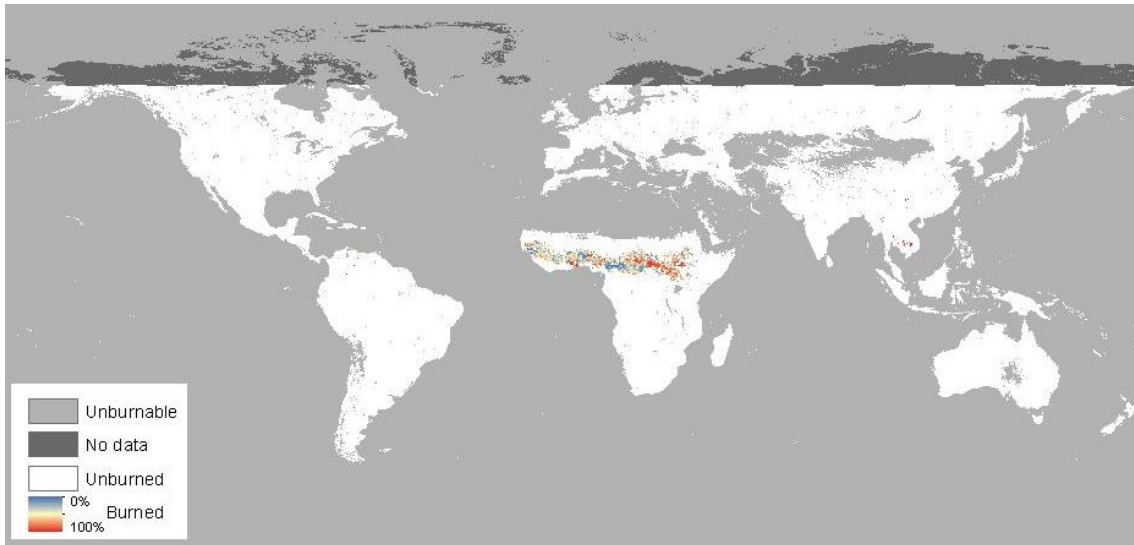


Figure 14: Percentage BA of LTDR data for January 2008

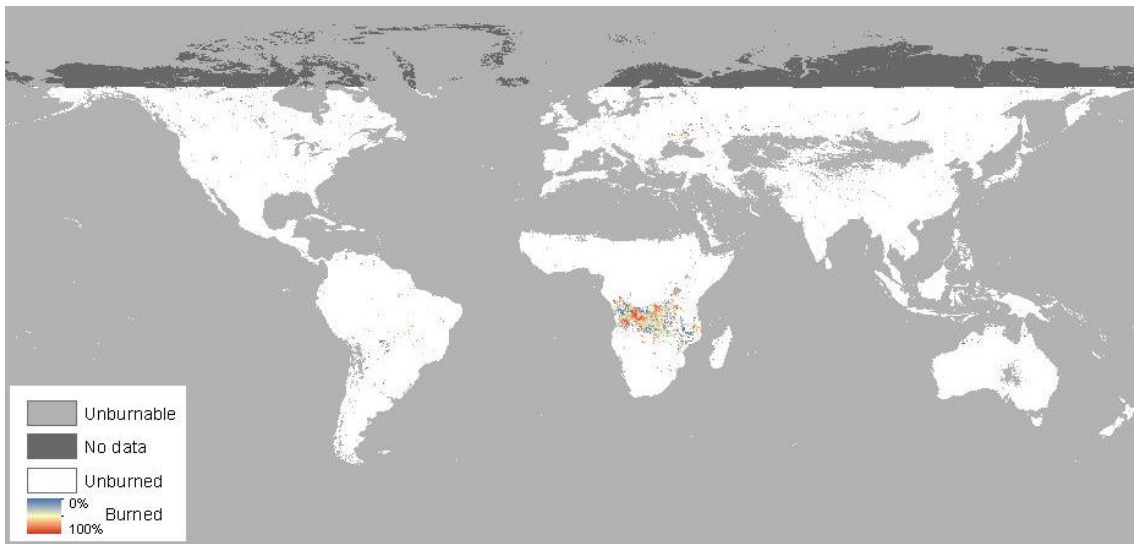


Figure 15: Percentage BA of LTDR data for July 2008

The Dates and Percentage of Burn Area is structured, so:

- Dates:
 - -2 Unburnable
 - -1 No data
 - 0 Unburned
 - 1-365 Day of burned

- Percentage of Burned Area:
 - -2 Unburnable
 - -1 No data
 - 0-1 Percentage of Burned Area

9 Uncertainty characterization

Since the binary RF model provides an estimation of the probability of each pixel being burned, we used this value as the uncertainty of the final product (Fig. 16, 17). The resulting files include two categories: probability (0-100) and No data (-1).

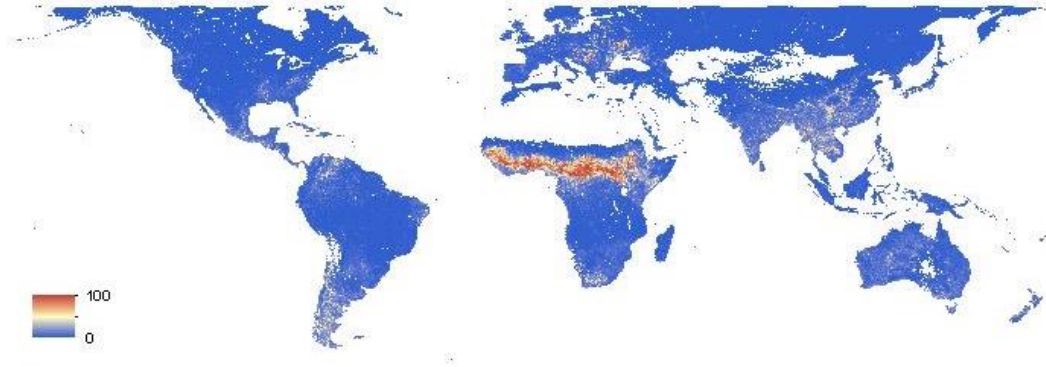


Figure 16: Uncertainty of LTDR data for January 2008

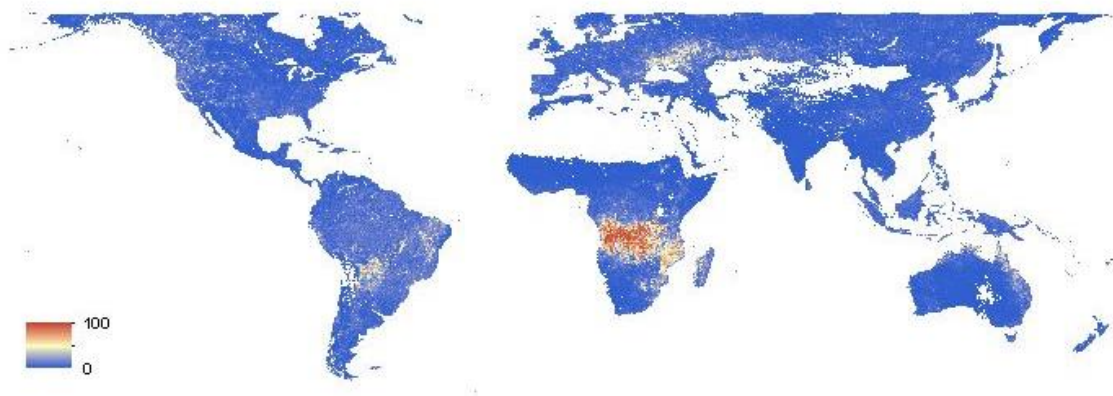


Figure 17: Uncertainty of LTDR data for July 2008

10 Validation

Preliminary validation of the classification results was done by comparing the outputs with the MCD64A1 product. Table 1 includes the cross tabulation analysis between the two products for January 2008 and Table 2 for July 2008. Overall accuracy was higher than 0.97 in both cases, but the omission and commission errors of BA were much higher, as expected. They were estimated in 0.44 and 0.47 respectively for January and 0.61 and 0.46 for July.

Table 1: Cross tabulation of LTDR results and MCD64A1 for January 2008

		MCD64A1		Total
		Burned	Not Burned	
RF	Burned	42,728	38,693	81,421
	Not Burned	34,128	4,108,291	4,142,419
Total		76,856	4,146,984	4,223,840



fire
cci

Fire_cci
Algorithm Theoretical Basis Document –
AVHRR LTDR data


Ref.:	Fire_cci_O2.D1_ATBD-AVHRR_v1.0		
Issue	1.1	Date	13/06/2018
Page			20

Commission error	0.47
Omission error	0.44
Dice Coefficient	0.54

Table 2: Cross tabulation of LTDR results and MCD64A1 for July 2008


		MCD64A1		Total
		Burned	Not Burned	
RF	Burned	39,305	33,881	73,186
	Not Burned	61,739	4,088,915	4,150,654
Total		101,044	41,22,796	4,223,840

Commission error	0.46
Omission error	0.61
Dice Coefficient	0.45


	Fire_cci	Ref.:	Fire_cci_O2.D1_ATBD-AVHRR_v1.0		
	Algorithm Theoretical Basis Document – AVHRR LTDR data	Issue	1.1	Date	13/06/2018
		Page	21		

11 References

- Adair, M., Cihlar, J., Park, B., Fedosejevs, G., Erickson, A., Keeping, R., Stanley, D., & Hurlburt, P. (2002). GeoComp-n, an advanced system for generating products from coarse and medium resolution optical satellite data. Part 1: System characterization. *Canadian Journal of Remote Sensing*, 28, 1-20
- Alcaraz-Segura, D., Chuvieco, E., Epstein, H.E., Kasischke, E.S., & Trishchenko, A. (2010). Debating the greening vs. browning of the North American boreal forest: differences between satellite datasets. *Global Change Biology*, 16, 760-770
- Barbosa, P.M., Pereira, J.M.C., & Grégoire, J.M. (1998). Compositing criteria for burned area assessment using multitemporal low resolution satellite data. *Remote Sensing of Environment*, 65, 38-49
- Cabral, A., Vasconcelos, M.J.P.d., Pereira, J.M.C., Bartholome, J.E., & Mayaux, P. (2003). Multi-temporal compositing approaches for SPOT-4 VEGETATION. *International Journal of Remote Sensing*, 24, 3343–3350
- Carmona-Moreno, C., Belward, A., Malingreau, J.P., Hartley, A., Garcia-Alegre, M., Antonovskiy, M., Buchshtaber, V., & Pivovarov, V. (2005). Characterizing interannual variations in global fire calendar using data from Earth observing satellites. *Global Change Biology*, 11, 1537-1555
- 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., Ventura, G., & Martín, M.P. (2005a). AVHRR multitemporal compositing techniques for burned land mapping. *International Journal of Remote Sensing*, 26, 1013-1018
- Chuvieco, E., Ventura, G., Martín, M.P., & Gomez, I. (2005b). Assessment of multitemporal compositing techniques of MODIS and AVHRR images for burned land mapping. *Remote Sensing of Environment*, 94, 450 – 462
- Cihlar, J., Latifovic, R., Chen, J., & Li, Z. (1999). Testing Near Real-Time Detection of Contaminated Pixels in AVHRR Composites. *Canadian Journal of Remote Sensing*, 25, 160-170
- Duchemin, B., Guyon, D., & Lagouarde, J.P. (1999). Potential and limits of NOAA-AVHRR temporal composite data for phenology and water stress monitoring of temperature forest ecosystems. *International Journal of Remote Sensing*, 20, 895-917
- Fraser, R.H., Hall, R.J., Landry, R., Lynham, T., Raymond, D., Lee, B., & Li, Z. (2004). Validation and calibration of Canada-wide coarse-resolution satellite burned-area maps. *Photogrammetric Engineering and Remote Sensing*, 70, 451-460
- Giglio, L., Loboda, T., Roy, D.P., Quayle, B., & Justice, C.O. (2009). An active-fire based burned area mapping algorithm for the MODIS sensor. *Remote Sensing of Environment*, 113, 408-420
- Giglio, L., Randerson, J.T., van der Werf, G.R., Kasibhatla, P.S., Collatz, G.J., Morton, D.C., & DeFries, R.S. (2010). Assessing variability and long-term trends in burned area by merging multiple satellite fire products. *Biogeosciences Discuss.*, 7, 1171-1186, doi:1110.5194/bg-1177-1171-2010,
- Granier, C., Bessagnet, B., Bond, T., D'Angiola, A., van Der Gon, H.D., Frost, G.J., Heil, A., Kaiser, J.W., Kinne, S., & Klimont, Z. (2011). Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period. *Climatic Change*, 109, 163-190

	Fire_cci	Ref.:	Fire_cci_O2.D1_ATBD-AVHRR_v1.0		
	Algorithm Theoretical Basis Document – AVHRR LTDR data	Issue	1.1	Date	13/06/2018
		Page	22		

- Hantson, S., Arneith, A., Harrison, S.P., Kelley, D.I., Prentice, I.C., Rabin, S.S., Archibald, S., Mouillot, F., Arnold, S.R., & Artaxo, P. (2016). The status and challenge of global fire modelling. *Biogeosciences*, *13*, 3359-3375
- Kirches, G., Krueger, O., Boettcher, M., Bontemps, S., Lamarche, C., Verheggen, A., Lembrée, C., Radoux, J., & 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-landcovercci.org/?q=documents#>*.
- Maggi, M., & Stroppiana, D. (2002). Advantages and drawbacks of NOAA–AVHRR and SPOT–VGT for burnt area mapping in a tropical savanna ecosystem. *Canadian Journal of Remote Sensing*, *28*, 231-245
- Martín, M.P., & Chuvieco, E. (1993). Mapping and evaluation of burned land from multitemporal analysis of AVHRR NDVI images. In P.J. Kennedy & M. Karteris (Eds.), *International Workshop: Satellite Technology and GIS for Mediterranean forest mapping and fire management* (pp. 71-83). Thessaloniki, Greece
- Martín, M.P., & Chuvieco, E. (1998). Cartografía de grandes incendios forestales en la Península Ibérica a partir de imágenes NOAA-AVHRR. *Serie Geográfica*, *7*, 109-128
- Martín, M.P., Díaz Delgado, R., Chuvieco, E., & Ventura, G. (2002). Burned land mapping using NOAA-AVHRR and TERRA-MODIS. In D.X. Viegas (Ed.), *IV International Conference on Forest Fire Research. 2002 Wildland Fire Safety Summit* (p. 45). Luso, Coimbra, Portugal: Millpress
- Moreno Ruiz, J.A.M., Lázaro, J.R.G., Cano, I.d.Á., & Leal, P.H. (2014). Burned area mapping in the North American boreal forest using Terra-MODIS LTDR (2001–2011): A comparison with the MCD45A1, MCD64A1 and BA GEOLAND-2 products. *Remote Sensing*, *6*, 815-840
- Mouillot, F., & Field, C.B. (2005). Fire history and the global carbon budget: a 1 degrees x 1 degrees fire history reconstruction for the 20th century. *Global Change Biology*, *11*, 398-420
- Mouillot, F., Schultz, M.G., Yue, C., Cadule, P., Tansey, K., Ciais, P., & Chuvieco, E. (2014). Ten years of global burned area products from spaceborne remote sensing—A review: Analysis of user needs and recommendations for future developments. *International Journal of Applied Earth Observation and Geoinformation*, *26*, 64-79
- Padilla, M., Stehman, S.V., Hantson, S., Oliva, P., Alonso-Canas, I., Bradley, A., Tansey, K., Mota, B., Pereira, J.M., & Chuvieco, E. (2015). Comparing the Accuracies of Remote Sensing Global Burned Area Products using Stratified Random Sampling and Estimation. *Remote Sensing of Environment* *160*, 114-121
- Pereira, J.M.C., Cadete, L., & Vasconcelos, M.J.P. (1994). An assessment of the potential of NOAA / AVHRR HRPT imagery for burned area mapping in Portugal. In, *2nd International Conference on Forest Fire Research* (pp. 665-666). Coimbra, Portugal
- Pinty, B., & Verstraete, M.M. (1992). GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetatio*, *101*, 15-20
- Power, M.J., Marlon, J.R., Bartlein, P.J., & Harrison, S.P. (2010). Fire history and the Global Charcoal Database: A new tool for hypothesis testing and data exploration. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *291*, 52-59
- 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

	Fire_cci	Ref.:	Fire_cci_O2.D1_ATBD-AVHRR_v1.0	
	Algorithm Theoretical Basis Document – AVHRR LTDR data	Issue	1.1	Date 13/06/2018
		Page	23	

- North America from NOAA/AVHRR data. *Remote Sensing of Environment*, 108, 198-208
- Ramo, R., & Chuvieco, E. (2017). Developing a Random Forest Algorithm for MODIS Global Burned Area Classification. *Remote Sensing*, 9, 1193
- Razafimpanilo, H., Frouin, R., Iacobellis, S.F., & Somerville, R.C. (1995). Methodology for estimating burned area from AVHRR reflectance data. *Remote Sensing of Environment*, 54, 273-289
- Riaño, D., Ruiz, J.A.M., Isidoro, D., Ustin, S.L., & Riaño, D. (2007). Global spatial patterns and temporal trends of burned area between 1981 and 2000 using NOAA-NASA Pathfinder. *Global Change Biology*, 13, 40-50, doi: 10.1111/j.1365-2486.2006.01268.
- Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J.P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104
- Sukhinin, A.I., French, N.H.F., Kasischke, E.S., Hewson, J.H., Soja, A.J., Csiszar, I.A., Hyer, E.J., Loboda, T., Conrad, S.G., Romasko, V.I., Pavlichenko, E.A., Miskiv, S.I., & Slinkina, O.A. (2004). AVHRR-based mapping of fires in Russia: New products for fire management and carbon cycle studies. *Remote Sensing of Environment*, 93, 546-564
- Thonicke, K., Venevsky, S., Sitch, S., & Cramer, W. (2001). The role of fire disturbance for global vegetation dynamics: coupling fire into a Dynamic Global Vegetation Model. *Global Ecology and Biogeography*, 10, 661-677
- Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., Vermote, E.F., & El Saleous, N. (2005). An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, 26, 4485-4498



Annex 1: Acronyms and abbreviations

AD	Applicable Document
ATSR	Along Track Scanning Radiometers
AVHRR	Advanced Very High Resolution Radiometer
BA	Burned Area
BAI	Burned Area Index
BRDF	Bidirectional reflectance distribution function
CCI	Climate Change Initiative
DC	Dice Coefficient
ECV	Essential Climate Variables
EPSG	European Petroleum Survey Group
ESA	European Space Agency
GAC	Global Area Coverage
GCOS	Global Climate Observing System
GCS	Geographic Coordinate System
GEMI	Global Environmental Monitoring Index
GFED	Global Fire Emissions Database
GIMMS	Global Inventory Monitoring and Modeling System
HDF	Hierarchical Data Format
HRPT	High-Resolution Picture Transmission
IFOV	Instantaneous Field of View
LC	Land Cover
LTDR	Long Term Data Record
MERIS	Medium Resolution Imaging Spectrometer
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NetCDF	NETwork Common Data Format
NIR	Near-infrared
NOAA	National Oceanic and Atmosphere Administration
PSD	Product Specification Document
RD	Reference Document
RF	Random Forest
S-2	Sentinel 2
TOA	Top of Atmosphere