



**fire**  
cci

---

## ESA Climate Change Initiative – Fire\_cci

### D2.2.3 Algorithm Theoretical Basis Document (ATBD) for the MERGED burned area product

---

<b>Project Name</b>	ECV Fire Disturbance: Fire_cci
<b>Contract Nº</b>	4000126706/19/I-NB
<b>Issue Date</b>	18/07/2024
<b>Version</b>	1.1
<b>Author</b>	Consuelo Gonzalo Martín, Ángel Mario García Pedrero and Jaime González Delgado
<b>Document Ref.</b>	Fire_cci_D2.2.3_ ATBD-Merged_v1.1
<b>Document type</b>	Public

*To be cited as: Gonzalo-Martín C., García-Pedrero A., González-Delgado, J., (2024) ESA CCI ECV Fire Disturbance: D2.2.3 Algorithm Theoretical Basis Document (ATBD) for the MERGED burned area product, version 1.1. Available at:*  
<https://climate.esa.int/en/projects/fire/key-documents/>

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.: Fire_cci_D2.2.3_ATBD-MERGED_v1.1	
		Issue 1.1	Date 18/07/2024
		Page	2

## Project Partners

Prime Contractor/ Scientific Lead & Project Management	UAH – University of Alcalá (Spain)
	UAH – University of Alcalá (Spain)
	UPM – Universidad Politécnica de Madrid (Spain)
Earth Observation Team	CNR-IREA – National Research Council of Italy - Institute for Electromagnetic Sensing of the Environment (Italy)
System Engineering	BC – Brockmann Consult GmbH (Germany)
	MPIC – Max Planck Institute for Chemistry (Germany)
Climate Research Group	CNRS – National Center for Scientific Research (France)



## Distribution

Affiliation	Name	Address	Copies
ESA	Clément Albergel (ESA)	clement.albergel@esa.int	electronic copy
Project Team	Emilio Chuvieco (UAH)	emilio.chuvieco@uah.es	electronic copy
	M. Lucrecia Pettinari (UAH)	mlucrecia.pettinari@uah.es	
	Amin Khairoun (UAH)	amin.khairoun@uah.es	
	Erika Solano (UAH)	erika.solano@uah.es	
	Consuelo Gonzalo (UPM)	consuelo.gonzalo@upm.es	
	Dionisio Rodríguez (UPM)	dionisio.rodriguez@ulpgc.es	
	Ángel García Pedrero (UPM)	angelmario.garcia@upm.es	
	Daniela Stroppiana (CNR)	stroppiana.d@irea.cnr.it	
	Thomas Storm (BC)	thomas.storm@brockmann-consult.de	
	Martin Böttcher (BC)	martin.boettcher@brockmann-cons...	
	Florent Mouillot (CNRS)	florent.mouillot@cefe.cnrs.fr	
	Philippe Ciaï (CNRS)	philippe.ciaï@lsce.ipsl.fr	
	Guido van der Werf (VU)	g.r.vander.werf@vu.nl	

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.: Fire_cci_D2.2.3_ATBD-MERGED_v1.1
		Issue 1.1      Date 18/07/2024
		Page 3

### Summary

This document presents the technical basis of the algorithms used to generate the FireCCI Burned Area Merged product version 1.0, based on the GRID version of the BA products FireCCI51, FireCCLT11 and MCD64CMQ. The document analyses the input requirements and the process to create the product, including the processor stages to get the final merged burned area product and the formatting of the data to obtain the final dataset. In addition, the validation of the product with the reference BARD database and the inter-comparison with the available BA products is also included.

	Affiliation/Function	Name	Date
<b>Prepared</b>	UPM	Consuelo Gonzalo Martín Ángel García Pedrero Jaime González Delgado	25/06/2024
<b>Reviewed</b>	UAH – Project Manager	M. Lucrecia Pettinari	18/07/2024
<b>Authorized</b>	UAH - Science Leader	Emilio Chuvieco	18/07/2024
<b>Accepted</b>	ESA - Technical Officer	Clément Albergel	20/09/2024

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

### Document Status Sheet

Issue	Date	Details
1.0	10/04/2024	First release of the document.
1.1	18/07/2024	Updated document with changes requested by ESA's Technical Officer

### Document Change Record

Issue	Date	Request	Location	Details
1.1	18/07/2024	ESA	Sections 1, 2.2.3	Small changes in the text

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page			4

## Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>7</b>
<b>2</b>	<b>BA Algorithm Description .....</b>	<b>8</b>
2.1	General Scheme .....	8
2.2	Input Data.....	8
2.2.1	FireCCI Burned Area Products.....	9
2.2.2	MCD64CMQ Climate Modelling Grid Burned Area Product .....	10
2.2.3	Burned Area Reference Dataset (BARD).....	11
2.3	Pre-processing of the input data.....	12
2.4	Methodology for Comparison Between BARD and BA Products .....	14
2.5	Phase 1 (2000-2020) .....	15
2.6	Evaluation of the BA estimation model.....	21
2.6.1	Evaluation of the BA estimation model for the year 2008 .....	24
2.6.2	Evaluation of the BA estimation model for years 2003 and 2017 .....	27
2.7	Phase 2 (1982-2018) .....	32
2.7.1	Generation of Single Long-Term BA Series .....	35
2.7.2	Confidence level estimation .....	38
2.7.3	Conclusion.....	39
2.8	FireCCIM10 Product.....	39
2.8.1	FireCCIM10 assessment.....	40
<b>3</b>	<b>Conclusions.....</b>	<b>44</b>
	<b>References.....</b>	<b>45</b>
	<b>Annex 1: General acronyms and abbreviations.....</b>	<b>47</b>
	<b>Annex 2: Climatic variable abbreviations .....</b>	<b>47</b>

## List of Figures

Figure 1: First phase for the generation of single long-term merged series of BA (2001-2018).....	8
Figure 2: Second phase for the generation of single long-term merged series of BA (1982-1993, 1995-2018). .....	8
Figure 3: An example of the intersection of BARD with grid cells. Some parts show that there are some cells that do not cover all the grid area.....	12
Figure 4: Number of cells that intersect with BARD for each product and year. a) All cells intercepted (35318). b) Number of cells after applying the selection criteria (2564). c) Number of cells after applying the land mask (3484). d) Number of cells after eliminating the small “non data” areas (4044). .....	13
Figure 5: Distribution of cells by ecoregion and year. ....	14

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page			5

Figure 6: Histogram of the distribution of cells that correspond to the burned area category (886 cells). .....14

Figure 7: Final version of the manual rules-based BA estimation model .....16

Figure 8: Scatterplot shows the difference between BARD BA values and selecting max, min or middle BA depending on the  $\Delta NDVI$ . .....18

Figure 9: Catplots comparing BARD BA values against the BA selection when  $\Delta NDVI < -0.1$ ,  $-0.1 < \Delta NDVI < 0.1$  and  $\Delta NDVI > 0.1$ . .....18

Figure 10: Accumulate BA for the period 2001-2018. ....21

Figure 11: Average annual BA for each ecoregion .....23

Figure 12: Average monthly BA for the period 2000-2018 .....23

Figure 13: Monthly BA for BA products (2008). ....25

Figure 14: Monthly number of cells selected from each BA product (2008). ....25

Figure 15: Comparison among BA products of the monthly BA for the different ecoregions (2008). .....27

Figure 16: a) Distribution of the difference in the number of BA cells provided by the BARD database and estimated by the new product for year 2008. b) Scatterplot between BA provided by the BARD database against BA estimated by the new product. ....27

Figure 17: Monthly accumulated BA for BA products for the years 2003 (a) and 2017 (b). .....28

Figure 18: Accumulated monthly BA for the year 2003 (a), and for the year 2017 (b)..31

Figure 19: Distribution of the difference in the number of BA cells provided by the BARD database and estimated by the new product for years 2003 (a) and 2017 (b). Scatterplot between BA provided by the BARD database against BA estimated by the new product for years 2003 (c) and 2017 (d). .....31

Figure 20: Long-term BA series estimated by the local regression model for two particular grid cells, for two different ecoregions (Mediterranean and Tropical Forest) and for the months of July (a) and January (b) .....33

Figure 21: Results of the analysis of the importance of the characteristics for different biomes and months. a) Boreal Forest north in January. b) Temperature savanna in May. c) Tropical forest north in January. d) Tropical forest south in August. (see Annex 2 for a variable abbreviation list). Numbers 1, 2 and 3 correspond to the previous, current, and following months. ....34

Figure 22: First row: BA prediction model based on the Random Forest Regressor. a) Training scheme. b) Generation of the product FireCCIM10 and validation (m: month, e: ecoregion, h: hemisphere). Second row: BA prediction model based on local polynomial regression. c) Training scheme. d) Generation of the FireCCIM10 product and validation. ....36

Figure 23: Annual BA estimation – SLT2. ....36

Figure 24: Monthly BA estimation – SLT2. ....36

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.: Fire_cci_D2.2.3_ATBD-MERGED_v1.1
		Issue 1.1      Date 18/07/2024
		Page 6

Figure 25: Monthly BA estimation pre-MODIS era – SLT2. ....	37
Figure 26: Monthly BA estimation MODIS era – SLT2.....	37
Figure 27: Distribution of BA differences between FireCCIM10 and BARD (km <sup>2</sup> ). a) Training phase, b) testing phase, and c) pre-MODIS era. ....	40
Figure 28: Average monthly BA time series for the FireCCILT11, FireCCI51, MCD64CMQ, FireCCIM10, GFED4 and V1.3.3 products. ....	40
Figure 29: BA time series for the FireCCILT11, FireCCI51, MCD64CMQ, FireCCIM10, GFED4 and V1.3.3 products, average all-time series, for different biomes. ....	42
Figure 30: FireCCIM10 product for February 2017. Total burned area at 0.25 spatial resolution (top) and Confidence Interval (bottom).....	43
Figure 31: FireCCIM10 product for December 2013. Total burned area at 0.25 spatial resolution (top) and Confidence Interval (bottom).....	44

### **List of Tables**

Table 1: Spatial resolution of BA products .....	9
Table 2: Comparison results for rule 1 .....	16
Table 3: Comparison results for rule 2 .....	17
Table 4: Results of the assessment of the training phase for the different versions of the estimation model .....	20
Table 5: Results of the assessment of the testing phase for the different versions of the estimation model .....	20
Table 6: Results of the assessment of the training phase for the input BA products and the estimation model. ....	20
Table 7: Results of the assessment of the testing phase for the input BA products and the estimation model. ....	21
Table 8: Average BA for the period 2000-2018.....	24
Table 9: Accuracy assessment of the estimated product against BARD for year 2008 ..24	24
Table 10: BA over the years 2008.....	25
Table 11: Accuracy assessment of the estimated product against BARD for years 2003 and 2017 .....	28
Table 12: Year accumulated BA for the years 2003 and 2017.....	28
Table 13: Results of the comparative assessment in the training phase for the BA source products and estimation models in the MODIS era.....	37
Table 14: Comparative assessment results in the testing phase for the source BA products and estimation models in the MODIS era. ....	38
Table 15: Comparative assessment results for CONUS_BA product in pre-MODIS era (years 1988, 1993 and 1998) .....	38

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page			7

## 1 Introduction

Currently there are a variety of burned area (BA) products obtained from satellite data, among others: FireCCILT11 (Otón et al. 2021), FireCCI51 (Lizundia-Loiola et al. 2020), MCD641 and MCD45A1 (Giglio et al., 2018), GFED4 (Giglio et al., 2014). Most of the global BA products currently available have a relatively short time series (2000 to present). In spite of that, the FireCCILT11 dataset spans from 1981 to 2018, although this product has been obtained from AVHRR imagery which suffers of low radiometric and geometric quality (MacGregor and Gorman 1994; Weber and Wunderle 2019).

Apart from differences in the temporal extent of these products, there are discrepancies in the estimates they provide for the same area and time period. This uncertainty is due, among other factors, to the coarse spatial resolution of the global BA products, but also to the lack of long-term BA time series (Chuvieco et al., 2019). In addition, a temporal extension of 22 years is not long enough for some end users, as for example atmospheric and carbon modellers, but many times also for the climate community in general. This is the reason why the Global Climate Observing System (GCOS) recommended extending the available BA datasets, based on sensors launched around 2000, backward to the 1980s (GCOS 2016).

The availability of diverse products poses challenges in terms of data integration, cross-comparison, and temporal consistency. Achieving homogeneity across these products is essential for creating a seamless and accurate historical record of burned areas, which is invaluable for scientific research. Extending the homogenization process backward in time is equally important for researchers seeking to build long-term datasets for trend analysis and climate change studies. However, this task may be challenging due to the lack of consistent and reliable historical satellite data.

It is possible to find some work in the literature that address this issue. Perhaps the most representative is (Giglio et al., 2010). In this work, a global, monthly BA product aggregated to 0.5° spatial resolution for the time period July 1996 through mid-2009 was developed. That BA series was compared to some other global burned area products (GFED2, L3JRC, GLOBCARBON, and MODIS MCD45A1) and found substantial differences in many regions.

The main objective of the FireCCI project is to generate a global single long-term time series of BA called FireCCIM10 spanning from 1982 to 2018 and with an accuracy at least equal to the current available products. At this moment, the two main grid global BA products developed by FireCCI are: FireCCILT11 (Otón et al. 2021), which was developed from AVHRR-LTDR data, spanning from 1982 to 2018, and FireCCI51 (Lizundia-Loiola et al. 2020), which is based on the Moderate Resolution Imaging Spectroradiometer (MODIS) and covers the period from 2001 to 2021. Both have a version with a resolution of 0.25 degrees. Moreover, the NASA grid BA product MCD64C6 is also available at the same resolution and time frame. The FireCCIS311 (Lizundia-Loiola et al. 2022) BA product, although producing better results than FireCCI51, only starts its time series on 2019, outside the scope of this analysis.

The generation of the FireCCIM10 product was split into two phases: (1) the integration of the aforementioned three products for the period 2001-2018 (MODIS period), and (2) the backward estimation for the period 2000 down to 1982 (pre-MODIS) and 2001-2018 (MODIS era). To drive the generation process, a database with a higher accuracy, used

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page			8

as a validation dataset for different BA products, was used to train the algorithms (BARD, (Franquesa et al., 2022) was used.

As other FireCCI BA products, the resulting FireCCIM10 product includes maps of the global burned area developed and tailored for use by climate, vegetation, and atmospheric modellers, as well as fire researchers or fire managers interested in historical burned patterns. This dataset covers the period 1982-2018, complementing and extending the temporal range of the previous BA products developed by the FireCCI project. This product is a GRID burned area product at a resolution of 0.25 degrees.

The methodologies proposed and implemented to generate this product in both phases are described in this document.

## 2 BA Algorithm Description

The objective of creating the FireCCIM10 algorithm has been to generate a model capable of estimating long BA time series (1982-2018) with higher precision than the BA products available for the two periods considered. For that, different strategies have been investigated for each period.

### 2.1 General Scheme

Figure 1 and Figure 2 illustrate the two phases considered in our methodology. Different strategies were needed for each period or phase.

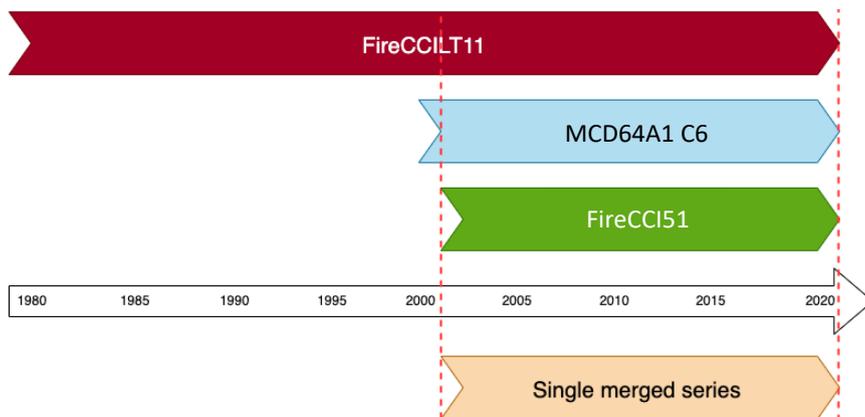


Figure 1: First phase for the generation of single long-term merged series of BA (2001-2018).

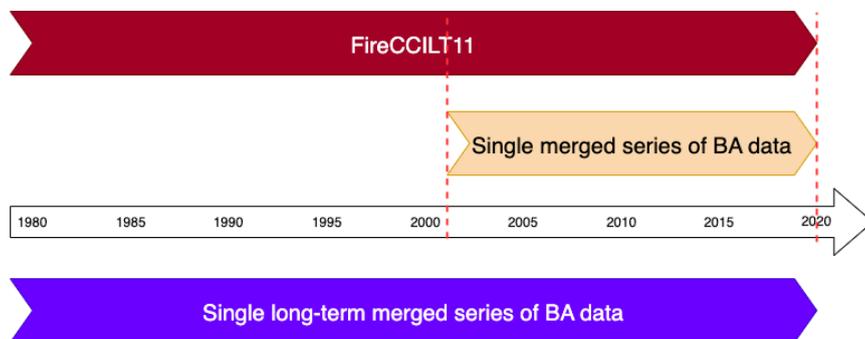


Figure 2: Second phase for the generation of single long-term merged series of BA (1982-1993, 1995-2018).

### 2.2 Input Data

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page			9

The three main inputs for the process to generate this merged BA dataset in the first phase have been FireCCILT11 (Otón et al. 2021) developed from AVHRR-LTDR data spanning from 1982 to 2018; FireCCI51 (Lizundia-Loiola et al. 2020), which is based on MODIS, available from 2001 to 2020; and MCD64CMQ (Giglio et al., 2020) also developed from MODIS Collection 6 and with a temporal extension of 20 years (2000-2020). The last product is distributed as a grid product (derived from the MCD64A1 pixel product), and the two first BA products are distributed at two different spatial resolutions (pixel and grid). The spatial resolution of the pixel versions is different for the two products, as shown in Table 1. Only the grid products have the same spatial resolution of 0.25 degrees.

*Table 1: Spatial resolution of BA products*

Spatial Resolution version	Burned Area Products		
	FireCCILT11	FireCCI51	MCD64CMQ
<b>Pixel</b>	0.05°	0.0002246° ( $\approx$ 250 m)	--
<b>Grid</b>	0.25°	0.25°	0.25°

The grid products are global products that aggregate for each cell the BA values of the pixels that belong to such cell. Details of how aggregation is done for each of the input BA products are included in Lizundia-Loiola et al. (2018) and Otón and Chuvieco (2021). According to the description of the FireCCI51 and FireCCILT11 products, there is a good geometric overlap between them.

## 2.2.1 FireCCI Burned Area Products

### 2.2.1.1 FireCCI51 grid burned area product

MODIS FireCCI51 (Lizundia-Loiola et al. 2020) is a BA product based on MODIS data and developed within the Fire CCI project (<https://climate.esa.int/en/projects/fire/>, last accessed in November 2023). This product covers the period 2001-2020 and complements and extends the temporal range of the previous BA products developed by the Fire CCI project: MERIS Fire\_cci v4.1 (FireCCI41, Chuvieco et al. 2016), which comprised the 2005-2011 period, and MODIS Fire\_cci v5.0 (FireCCI50, Chuvieco et al. 2018), which comprised the 2001-2016 period. FireCCI51 is provided as a global and monthly dataset at two different spatial resolutions: 250 m at pixel resolution and 0.25° at grid resolution. For generating FireCCIM10, only the grid resolution product has been used. Although the grid version of FireCCI51 includes different layers, only the total burned area associated with each cell has been used to generate FireCCIM10.

A complete description of the FireCCI51 product, as well as its characteristics, is included in Lizundia-Loiola et al. (2020) and Pettinari et al. (2021).

### 2.2.1.2 FireCCILT11 grid burned area product

FireCCILT11 is a BA dataset based on LTDR-AVHRR data (Otón et al. 2021). Its temporal extension, from 1982 to 2018, doubles the time series of the previous BA products developed within CCI. FireCCILT11 also adds an additional year of coverage (2018) to the previous Beta version FireCCILT10 (Otón et al. 2019), using an improved algorithm that addresses many of the issues found in that version. Two global monthly products at different spatial resolution are available (<https://climate.esa.int/en/>

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page			10

[projects/fire/](#), last accessed in November 2023): pixel files at 0.05° and grid files at 0.25°. For generating FireCCIM10, only the latter product has been used.

Despite the long temporal extension of this product, given the quality limitations of the images from which it has been generated (low radiometric and geometric quality and low spatial resolution), there is a need to generate other products with the same temporal extension that avoid or at least minimize these limitations (MacGregor and Gorman 1994; Weber and Wunderle 2019).

### 2.2.1.3 *General characteristics of the FireCCI grid BA input products*

The two source datasets generated by the FireCCI project, FireCCILT11 and FireCCI51, are in geographical coordinates, as is the merged product, being the unit of analysis of all these products the standard tile.

FireCCI BA products include different layers: the sum of BA, the standard error of the burned area, the fraction of burnable area, the fraction of observed area, and the sum of burned area of each land cover class. A brief description of these layers is included next (Chuvieco et al, 2017; Pettinari et al., 2021):

- The sum of burned area is computed by identifying for each target grid cell the burned pixels from the pixel product located inside the grid cell and adding their areas. The unit of this attribute is m<sup>2</sup>.
- Standard error is also computed as an aggregation of the confidence level of each source pixel. It is also given in units of m<sup>2</sup>.
- The fraction of burnable area (FBA) takes values between 0 and 1. A cell with a value equal to 1 indicates that the entire area covered by the cell consists of burnable pixels while a value of 0 indicates that no source pixels of the area covered by the cell are burnable. This variable is unitless.
- The fraction of observed area (FOA) 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.
- The sum of BA in each land cover burned class allows discriminating in each grid cell the different land covers affected by the fire.

Chuvieco et al. (2017) provide a comprehensive description of the process to generate each of these layers.

### 2.2.2 **MCD64CMQ Climate Modelling Grid Burned Area Product**

The CMQ burned area product is a monthly gridded summary intended for use in regional and global modelling. The MCD64CMQ files follow the standard MODIS product naming convention. Each data layer has 720 rows and 1440 columns that comprise a global 0.25° grid. The centre of the upper left grid cell is located at longitude 179.875°W, 89.875°N. One data layer has an additional dimension (i.e., it is three-dimensional) as described below.

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page	11		

Each product file nominally contains four data layers: BurnedArea, QA, UnmappedFraction, and LandCoverDist, each stored as a separate HDF4 Scientific Data Set (SDS)<sup>1</sup> (last accessed on December 2023):

- **BurnedArea:** Monthly area burned. The scaled value stored in this SDS must be multiplied by a factor of 0.01 to obtain burned area in hectares.
- **QA:** 8-bit quality assurance bitfield. 0=water, 1=unprocessed land, 2=processed land.
- **Unmapped Fraction:** Combined spatial and temporal fraction of the unmapped land area within each grid cell. Specifically, this quantity is the average number of days that could be mapped during the one-month product period over all 500-m land grid cells falling within each 0.25° bin, divided by the number of days in the product calendar month. The resulting proportion is expressed as a percentage. Unprocessed land grid cells (see QA layer above) will have an unmapped fraction of 100%.
- **LandCoverDist:** A breakdown (to the nearest percent) of the area burned in each grid cell by the 16 different UMD land cover classes available in the Collection 6 MCD12Q1 land cover product. These classes are shown below. The number in each row is the index into the specific plane of the three-dimensional LandCoverDist array that contains the percentage of the area burned for the corresponding land cover type.

0	water	8	woody savannas
1	evergreen needleleaf forests	9	savannas
2	evergreen broadleaf forests	10	grasslands
3	deciduous needleleaf forests	11	Permanent wetlands
4	deciduous broadleaf forests	12	croplands
5	mixed forest	13	urban and built-up
6	closed shrublands	15	barren
7	open shrublands	16	unclassified

The NASA MCD64CMQ product is in sinusoidal projection. That means that the NASA data must be reprojected to geographical coordinates.

### 2.2.3 Burned Area Reference Dataset (BARD)

BARD is a public database created from current products by different international burned area projects. It contains more than 2600 files from Landsat and Sentinel-2 images. This product arises from the need to validate new products generated in the field of remote sensing of burned areas. It can also be used as reference data to train and test the performance of new burned area algorithms.

<sup>1</sup> [https://lpdaac.usgs.gov/documents/875/MCD64\\_User\\_Guide\\_V6.pdf](https://lpdaac.usgs.gov/documents/875/MCD64_User_Guide_V6.pdf)

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page			12

BARD includes different datasets from different products (indicated with the name of the project that financed its development, their geographical extent, and the years for which the reference datasets were created): FireCCI global (2018), FireCCI global (2003-2014), FireCCI Africa (2016), CONUS Landsat Burned Area (1988-2013), NOFFi Greece (2016-2018) and C3S global (2017-2019).

The description of this database is provided in Franquesa et al., (2022), and the data is available at <https://edatos.consorciomadrono.es/dataverse/BARD> (last accessed on March 2024).

Other input data used in the generation of the estimation models have been the NDVI provided by the monthly MOD13C2 product<sup>2</sup> (last accessed on December 2023) with a resolution of 0.05°, as well as the biomes scheme proposed in Olson et al. (2001).

### 2.3 Pre-processing of the input data

Prior to the use of the BARD as the ground truth of our study, a preprocessing of this database was needed. Due to the difference in spatial resolution of the BARD products and the BA products (see Figure 3), a spatial adjustment was required to be able to use the database as validation data and, moreover, to better approach the comparison to reality.

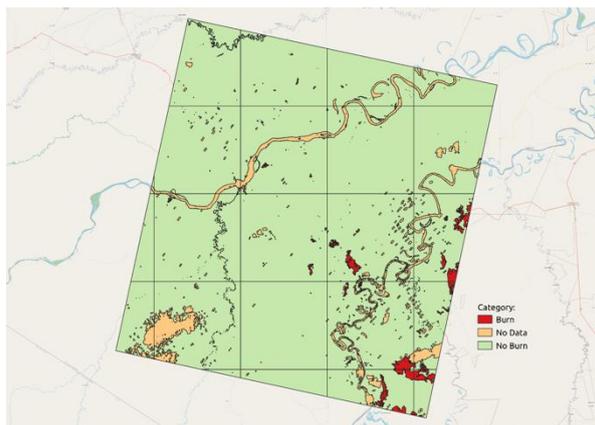


Figure 3: An example of the intersection of BARD with grid cells. Some parts show that there are some cells that do not cover all the grid area.

To accomplish this, each grid cell was labelled with a number to be able to identify them. Then, the difference BARD products got intersected with the BA area grid cells, and the BA calculated for each cell of the grid. Figure 4 a), shows the number of grid cells that intersect with BARD for each product and year. But not all these cells will be useful for training and validation of models. It is necessary to find those that completely cover a grid cell and whose category is not "no data" (no observable zone or, in some products, water zones). Only cells that meet both criteria were used to generate the rules. With this process, only 2564 cells remained (Figure 4 b) for further processing. Some of the areas without data are unburnable zones, so a mask was applied to categorise those areas as areas without data or with no burnable areas. This new category helped to avoid losing more information, as it was then possible to use the cells where the no-data information is categorised as not burnable, which increases the number of total cells that could be used

<sup>2</sup> <https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD13C2#overview>

to 3484 cells (Figure 4 c). Trying to find more cells that could be used, an additional strategy was used to eliminate areas classified as 'no data'. This strategy looks for small areas that cannot be registered by the sensor, and these areas will get the same category as the neighbouring polygons, increasing the number of cells to 4044 (Figure 4 d).

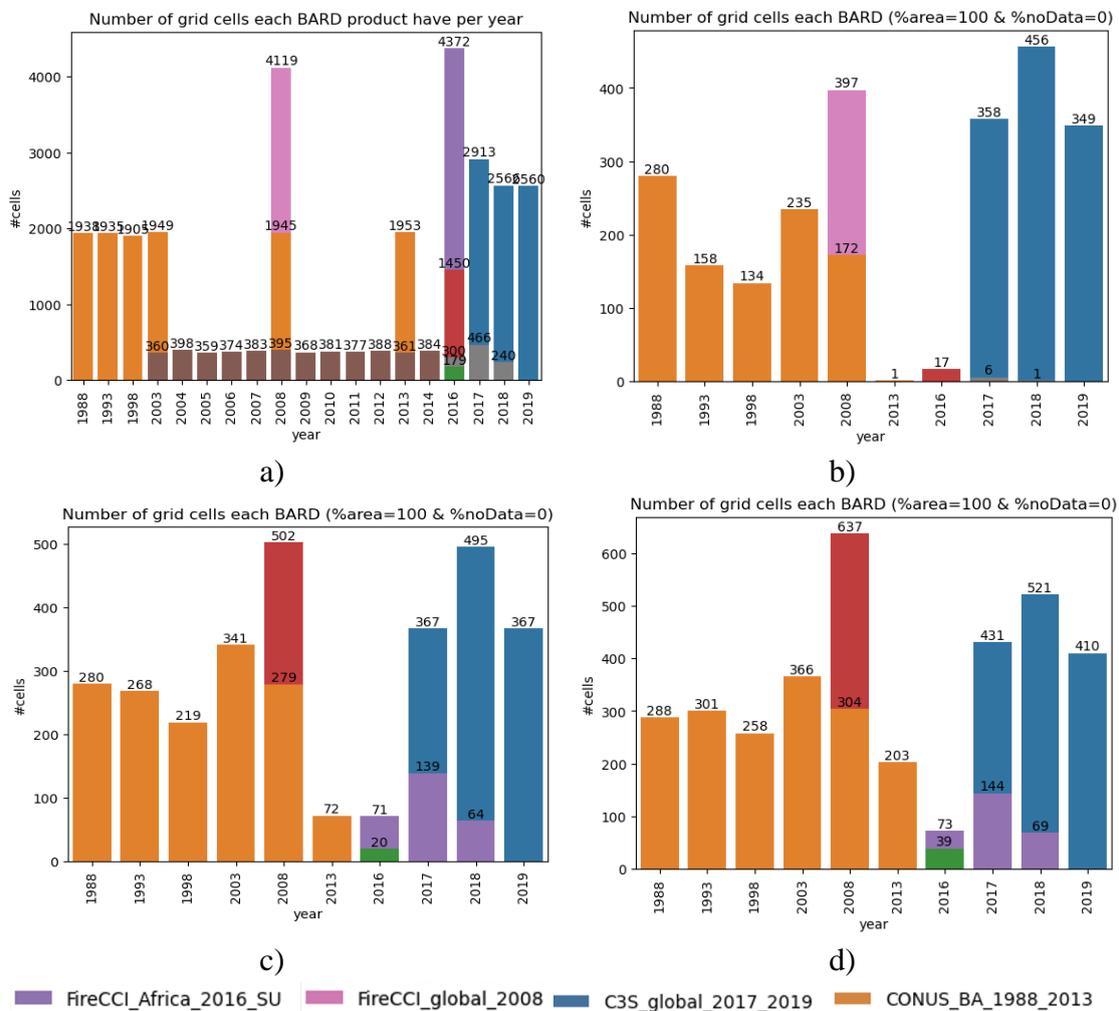


Figure 4: Number of cells that intersect with BARD for each product and year. a) All cells intercepted (35318). b) Number of cells after applying the selection criteria (2564). c) Number of cells after applying the land mask (3484). d) Number of cells after eliminating the small "non data" areas (4044).

Finally, and considering the ecoregions defined in Olson et al. (2001), only those cells that correspond with only one ecoregion have been selected. Leading to 2648 cells that can be used in the generation and validation of the models (Figure 5).

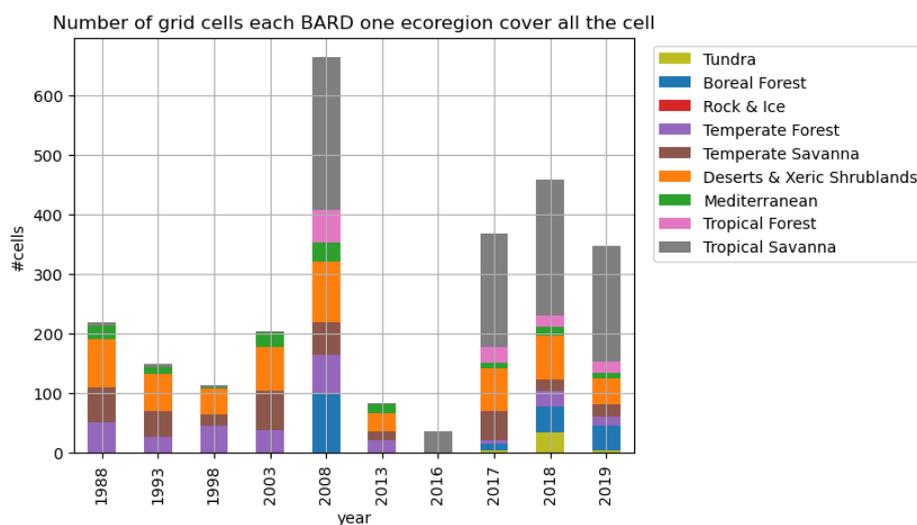


Figure 5: Distribution of cells by ecoregion and year.

After the pre-processing and selection process described before, a total of 1816 cells are available for the period 2001-2018. These cells have been used to generate and evaluate the burned area estimation models. Figure 6 represents the distribution of the number of cells against the BA extension. Most of them have a burned area smaller than 122 km<sup>2</sup>. These BARD data have been split into training and testing sets. The training set consists of 1123 cells for the years 2008 and 2018. These years have been selected for training because of the distribution of ecoregions. The testing set includes 693 cells for years 2003, 2013, 2016 and 2017.

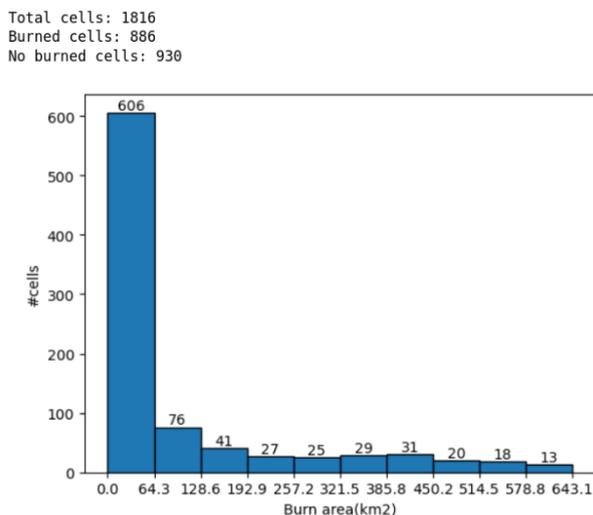


Figure 6: Histogram of the distribution of cells that correspond to the burned area category (886 cells).

## 2.4 Methodology for Comparison Between BARD and BA Products

To evaluate the performance of the models generated for estimating BA values, a methodology for comparison of BA estimated by BARD and the models was designed and implemented.

The comparison between the BARD and BA product was carried out cell by cell. Therefore, the same cells selected in the BARD are considered in the BA products.

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page	15		

Special attention should be paid to the temporal aspect since the BA products have a monthly temporal resolution, while the BARD products could span from some days of observation to a year. To avoid this problem, the evaluation was carried out with the accumulated BA of the products during the period determined by the BARD. Spatial distribution differences amongst BARD and each BA product were evaluated.

## 2.5 Phase 1 (2000-2020)

Among the strategies investigated in the first phase were the training of different machine learning models, such as decision trees and random forests, using as target selected cells from the BARD. In both cases, the available information was not sufficient to generate robust models. Finally, the best results were obtained when manual rules were established. These manual rules were obtained from a detailed comparison between the BA source products and the ground truth (BARD).

An incremental and iterative process was carried out for the design of the final version of the burned area estimation model based on manual rules. These manual rules were obtained from a deep comparative analysis between the BA values of the selected cells provided by the BA input products and the BARD dataset. Figure 7 includes the flowchart of the final model, where each colour corresponds to a different version of the model, as shown in the legend of the figure. A tolerance of 2% was considered in the analysis.

The rules determine for the baseline model (white colour in Figure 7) from the comparison between the different sources of burned areas where:

- Baseline rule 1:** If for the N-burned area products the BA values is 0, then the BA value assigned to this cell is 0. For the available data, the agreement between this estimate and the values provided by the BARD data was 99.33 % in the training data set and 98.71 % in the testing data set. The disagreement is due mainly to the BA regions included in the BARD dataset with an extension always lower than 50 km<sup>2</sup> and mostly less than 12.5 km<sup>2</sup>. For the analysis, it is suspected that some of the error could be produced for the temporal issues commented above, that is, the month selected in the dataset does not temporally match the date in the BARD dataset. To minimise this type of errors, an analysis of the burned area was also done for all the months included in the BARD observation. For that, the accumulated value of BA during the pre-date and the post-date is calculated and we apply the rule directly to the accumulated value. Under this consideration, the agreement rates increase to 99.81% and 100% for training and testing sets, respectively. Consequently, the rest of the analysis was carried out under this condition. Table 2 summarizes the results corresponding to the analysis discussed above for this first rule. All the errors found by this rule are by underestimation, where the BARD predict that a small area gets burned and it is not detected in the BA products. By this analysis it is concluded that it is better to use the accumulated burn area for compare with the BARD. The cells used to determine this rule were removed for the rest of the analysis at this stage.

Table 2: Comparison results for rule 1

Dataset	# of correct cells	# of non-correct cells	Total number of cells	% agreement
Training (accumulated)	514	1	515	99.81%
Testing (accumulated)	244	0	244	100%

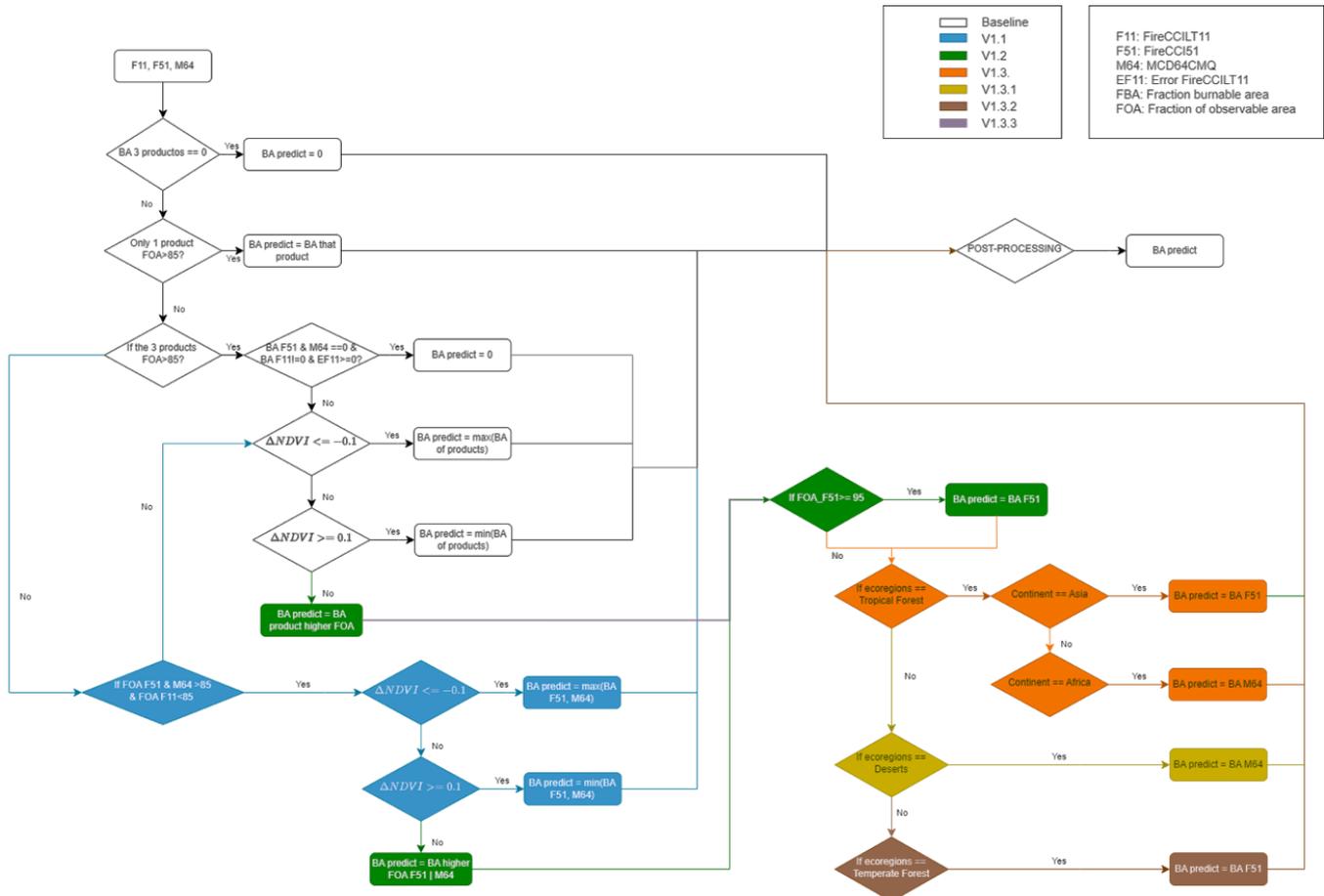


Figure 7: Final version of the manual rules-based BA estimation model

- Baseline rule 2:** The estimated BA value will be the BA value provided by the BA product with a higher FOA. For this rule, the analysis will only be made with the accumulated value during the pre-burn date and the postdate from the compared cells. The number of cells available to perform this analysis was 247 cells for training and 153 for testing. Table 3 illustrates a situation where most of the errors are because an overestimation of the rule respect to the BARD. Some of these errors can be explained by the temporal resolution of the products and the BARD, where the products have a monthly resolution, while the BARD depending, on which cell, can be for some days to a couple of months. On the other hand, it can be suspected that this is the reason why the real underestimation errors are lower.

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>		Ref.: Fire_cci_D2.2.3_ATBD-MERGED_v1.1
			Issue 1.1      Date 18/07/2024
			Page 17

Table 3: Comparison results for rule 2

Dataset	# of correct cells	# of non-correct cells	% agreement	% overestim.	% underestim.
Training	181	66	73.28%	72.73%	27.27%
Testing	137	16	89.54%	75%	25%

- Baseline rule 3:** If the BA value of one product is greater than 85% of the total grid area, then the predicted BA value will be this BA value. The number of cells available that met this condition was very low. In fact, only 6 cells for the training dataset did. For 5 of them, good predictions were obtained, but there is one where the error was very high. This error was analysed, and it was proved that the error appears because of a temporal problem between the BARD and the BA products. According to the BA pixel product, the date of appearance of all burned areas was before the date in the BARD dataset.
- Baseline rule 4:** Selection of the maximum or minimum value of the BA product depending on the variation of NDVI. An analysis of how the difference in the NDVI affects the selection process was conducted. This analysis was done for the training BARD reference cell (2008, 2018) where not all products predict 0 and where more than a product have a FOA higher than 85 or none of them have a FOA higher than 85. The results of this analysis are represented in Figure 8. This figure shows a different behaviour from selecting the max, min, and middle BA against BARD with the biggest difference made when  $\Delta NDVI$  is higher than -0.1 and lower than 0.1. Therefore, to better understand this behaviour, catplot representations were created to analyse which selection is better for each case when the  $\Delta NDVI$  is higher than 0.1, is lower than -0.1 or in between -0.1 and 0.1 (Figure 9). Considering that a low variability is desirable, from Figure 9, it can be concluded that when  $\Delta NDVI$  is lower than -0.1 is the maximum value of BA which must be selected; while, in the case that  $-0.1 < \Delta NDVI < 0.1$ , even though none of the values (max, min, middle) provided a good adjustment with the BARD BA values, since there are 248 cells from the selection of middle BA that predict better against the 180 cells that work better selecting max, it is decided to use the selection of the middle value.

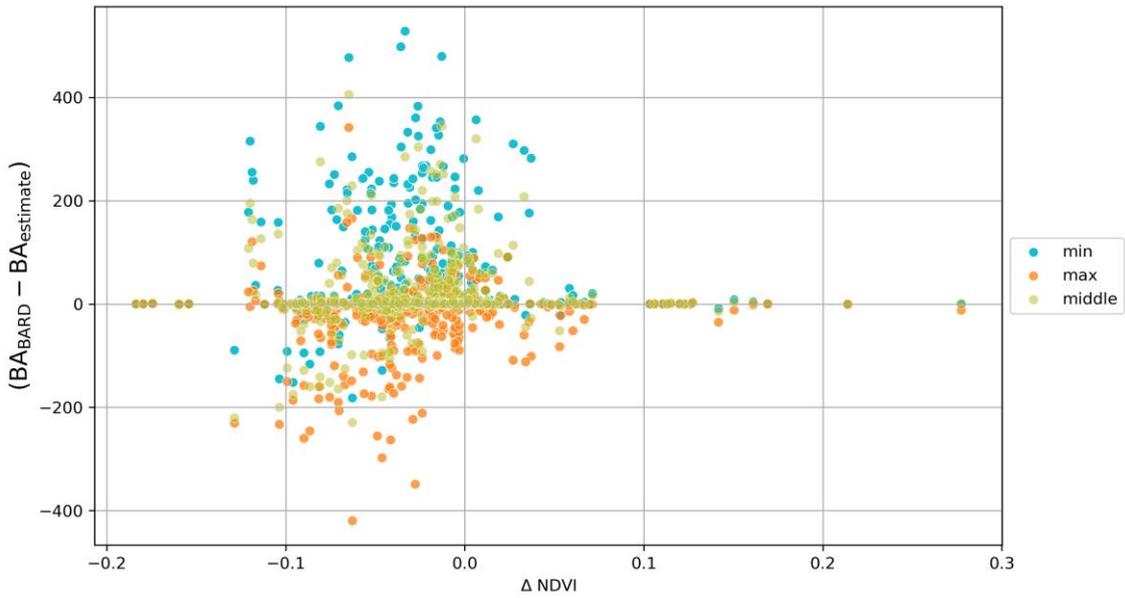


Figure 8: Scatterplot shows the difference between BARD BA values and selecting max, min or middle BA depending on the  $\Delta NDVI$ .

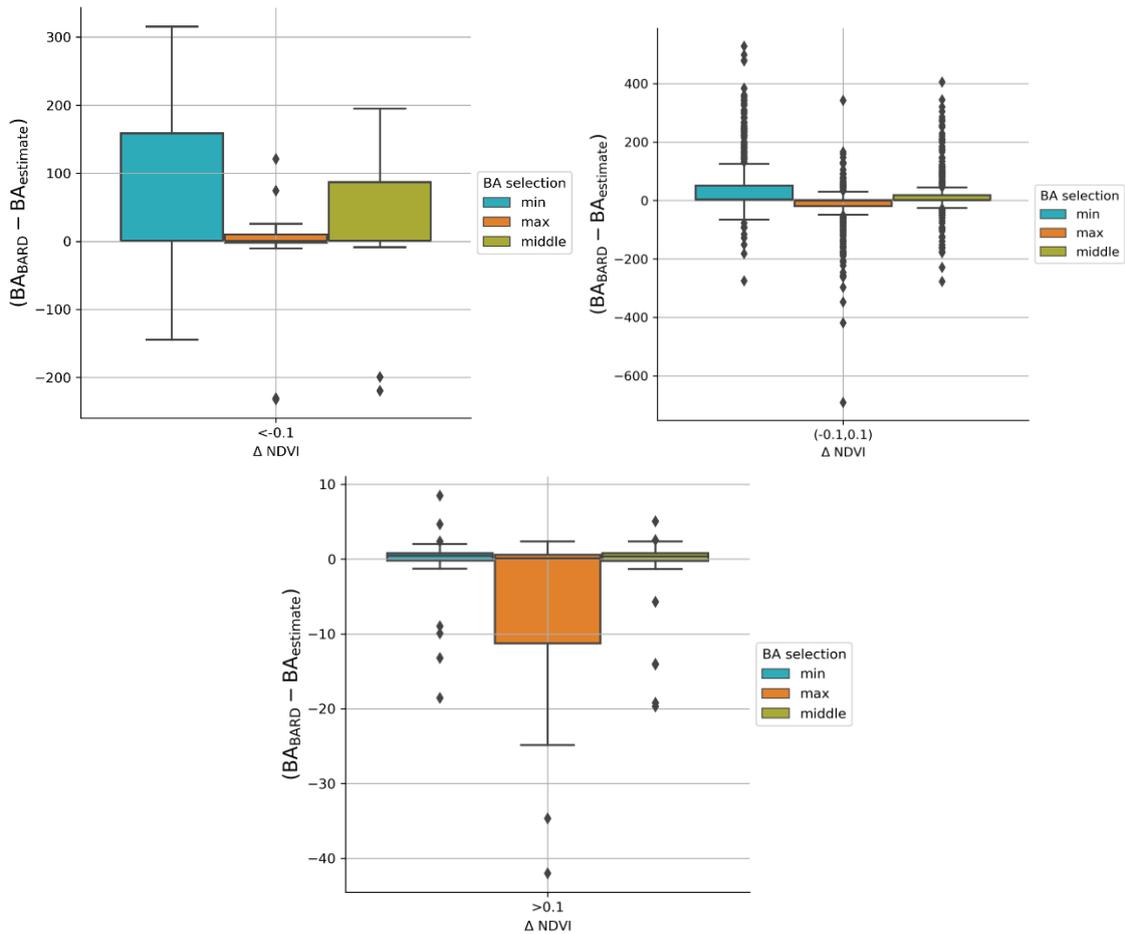


Figure 9: Catplots comparing BARD BA values against the BA selection when  $\Delta NDVI < -0.1$ ,  $-0.1 < \Delta NDVI < 0.1$  and  $\Delta NDVI > 0.1$ .

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.: Fire_cci_D2.2.3_ATBD-MERGED_v1.1
		Issue 1.1      Date 18/07/2024
		Page 19

For the incorporation of new rules to upgrade the different versions, similar analyses were carried out. Only the descriptions of the new rules included in each version are summarised below.

- **V1.1:** Add new rules when the FOA of FireCCI51 and MCD64CMQ are greater than 85%, and the FOA of FireCCILT11 is lower than 85%. If  $\Delta NDVI < -0.1$ , use  $\max(\text{BA}_{F51}, \text{BA}_{M64})$ ; if  $\Delta NDVI > 0.1$ , use  $\min(\text{BA}_{F51}, \text{BA}_{M64})$ ; otherwise, use the product with the higher FOA.
- **V1.2:** Modify the rule when  $-0.1 < \Delta NDVI < 0.1$ . If all products have a FOA higher than 85%, the product with the higher FOA is selected, unless  $\text{FOA}_{F51} > 95\%$ , in which case the selected product was FireCCI51.
- **V1.3:** For the selections made by the  $\Delta NDVI$  between -0.1 and 0.1, when the FOA of all products is greater than 85%, or when  $\text{FOA}_{F51}$  and  $\text{FOA}_{M64}$  are greater than 85%, if the cell ecoregion is Tropical Forest and the cell is in Asia, the product FireCCI51 is selected. If the cell is in Africa or in Tropical Forest, the MCD64CMQ should be used.
- **V1.3.1:** For the selections made by the  $\Delta NDVI$  between -0.1 and 0.1, when the FOA of all products are greater than 85%, or when  $\text{FOA}_{F51}$  and  $\text{FOA}_{M64}$  are greater than 85%, if the cell ecoregion is Desert & Xeric Shrublands, MCD64CMQ is selected.
- **V1.3.2:** For the selections made by the  $\Delta NDVI$  between -0.1 and 0.1, when the FOAs of all products are higher than 85%, or when  $\text{FOA}_{F51}$  and  $\text{FOA}_{M64}$  are higher than 85%, if the cell ecoregion is the Temperate Forest, FireCCI51 must be selected.
- **V1.3.3:** Add ecoregion rules also when the FOAs of all products are lower than 85%.
- **V1.4:** Perform non-linear regression between Hotspot and the BA predicted in v1.3.3.

The performance of the models obtained in all versions was compared with the BA values provided by the BARD for the training and testing datasets, considering a tolerance of 2%. The total number of training and testing cells used in the evaluation was 1123 and 693, respectively. The evaluation results are included in Table 6 and Table 7. In these tables, the number of correct cells has been included for the BA of the BARD data and the BA estimated by the models, having a difference lower than 2% of the cell. For the cells in which this difference is higher of 2% (2<sup>nd</sup> column), the cells that provide higher (overestimation) and lower (underestimation) BA values than BARD have been calculated separately (3<sup>rd</sup> and 4<sup>th</sup> columns). The fifth column includes the percentage of agreement and in the last two columns, the metric MAE and  $R^2$  have been included. In general, an improvement in performance was obtained from one version of the model to the next, except for version v1.4, since in this case the performance is worse than that of version 1.3.3, which was considered the final version.

Table 4: Results of the assessment of the training phase for the different versions of the estimation model

Version	Correct cells	Overest. error	Underst. error	% agreem.	MAE	R <sup>2</sup>
<b>Baseline</b>	926	60	137	82.46	16.227	0.828
<b>V1.1</b>	932	66	125	83.00	14.610	0.862
<b>V1.2</b>	927	71	125	82.55	14.775	0.859
<b>V1.3</b>	927	71	125	82.55	14.710	0.859
<b>V1.3.1</b>	931	72	120	82.90	13.574	0.857
<b>V1.3.2</b>	931	72	120	82.90	13.574	0.857
<b>V1.3.3</b>	931	72	120	82.90	13.574	0.857
<b>V1.4</b>	924	79	120	82.28	14.120	0.850

Table 5: Results of the assessment of the testing phase for the different versions of the estimation model

Version	Correct cells	Overest. error	Underst. error	% agreem.	MAE	R <sup>2</sup>
<b>Baseline</b>	550	54	89	79.36	21.879	0.684
<b>V1.1</b>	554	62	77	79.94	20.010	0.739
<b>V1.2</b>	559	60	74	80.66	19.078	0.750
<b>V1.3</b>	559	60	74	80.66	19.210	0.745
<b>V1.3.1</b>	558	57	78	80.52	18.901	0.751
<b>V1.3.2</b>	558	57	78	80.52	18.901	0.751
<b>V1.3.3</b>	558	57	78	80.52	18.901	0.751
<b>V1.4</b>	545	55	93	78.64	22.083	0.700

The estimated BA values provided by this version were also compared with the BA values provided for the different source BA products for the training and testing cells. The comparison results have been included in Table 6 and Table 7. The results included in Table 6 show that the agreement between the BA estimated by the model corresponding to V1.3.3 and BARD is the highest, providing the best values of the metrics considered in comparison to the rest of the BA products during the training phase. During the testing phase, the best agreement is provided by the FireCCI51 product, with a small difference (0.72) with respect to the model corresponding to V1.3.3.

Table 6: Results of the assessment of the training phase for the input BA products and the estimation model.

Version	Correct cells	Overest. error	Underst. error	% agreem.	MAE	R <sup>2</sup>
<b>FireCCILT11</b>	871	111	141	77.56	22.058	0.728
<b>FireCC51</b>	928	62	133	82.64	14.476	0.866
<b>MCD64CMQ</b>	928	62	133	82.64	16.957	0.775
<b>Baseline</b>	926	60	137	82.46	16.227	0.828
<b>V1.3.3</b>	931	72	120	82.90	13.574	0.857

Table 7: Results of the assessment of the testing phase for the input BA products and the estimation model.

Version	Correct cells	Overest. error	Underst. error	% agreement	MAE	R <sup>2</sup>
<b>FireCCILT11</b>	480	119	94	69.26	27.778	0.684
<b>FireCC51</b>	563	50	80	81.24	18.781	0.866
<b>MCD64CMQ</b>	551	42	100	79.51	25.281	0.619
<b>Baseline</b>	550	54	89	79.36	21.879	0.684
<b>V1.3.3</b>	558	57	78	80.52	18.901	0.751

## 2.6 Evaluation of the BA estimation model

Once the rules corresponding to version V1.3.3 were established as the final ones, the global monthly BA product was estimated for the whole period 2001-2018. Figure 10 shows the evolution of BA for the period 2001-2018 for the three source BA products, as well as the BA provided by the baseline estimation model and the final version of the BA estimation model (V1.3.3).

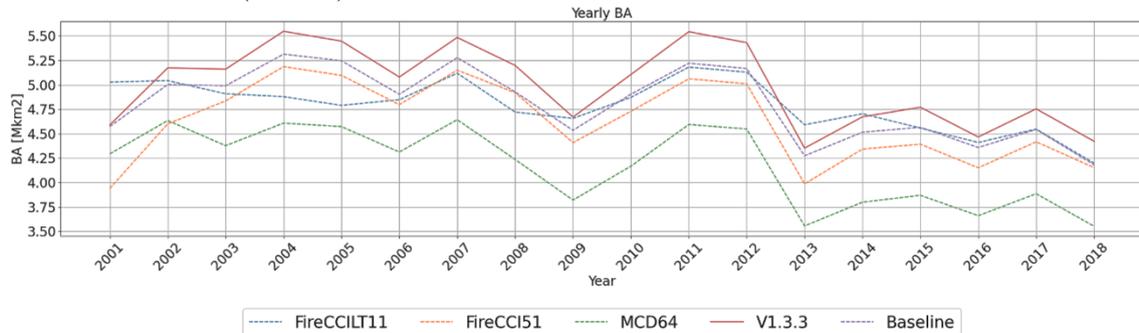
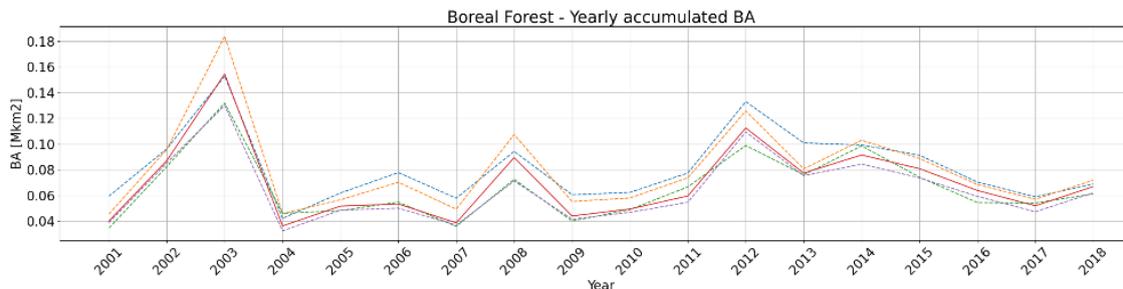


Figure 10: Accumulate BA for the period 2001-2018.

The MCD64CMQ product provides the lowest BA values for the entire period, while the new BA product mostly estimates the highest values. Differences between the baseline model, more similar to the Fire BA source products, and the version 1.3.3 can be appreciated.

The same representations for each of the ecoregions are shown in Figure 11, where a similar behaviour to the global analysis can be observed for most of the ecoregions, with some exceptions as Boreal Forest and Temperate Savana, where input Fire BA products provide highest values of BA than the new estimates. It can be also pointed that the time series corresponding to the product FireCCILT11 overestimate the rest of the products in several ecoregions, as Mediterranean, Temperate Forests and Temperate Savanna.





fire  
cci

**Fire\_cci**  
**Algorithm Theoretical Basis**  
**Document - MERGED**

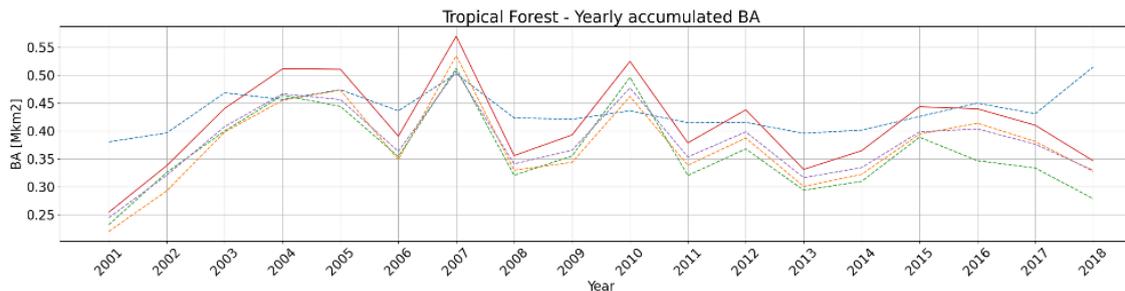
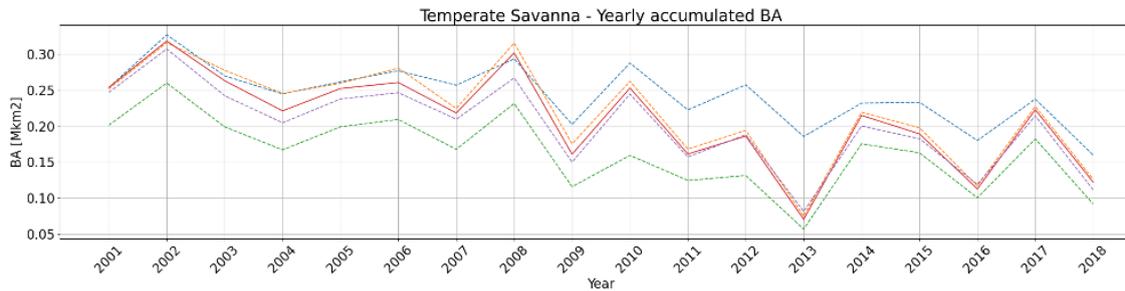
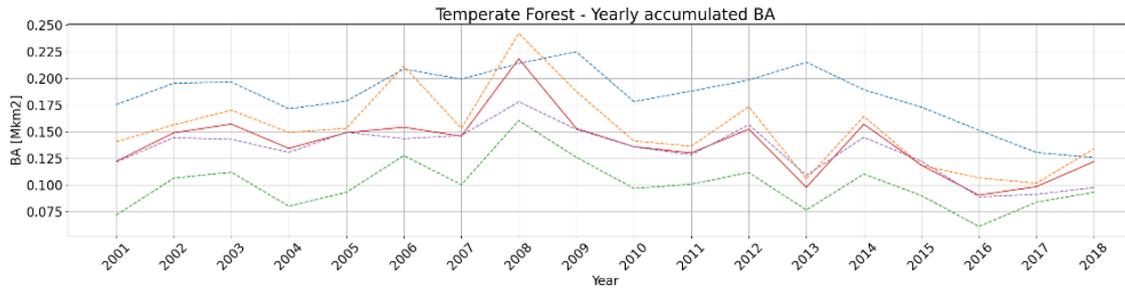
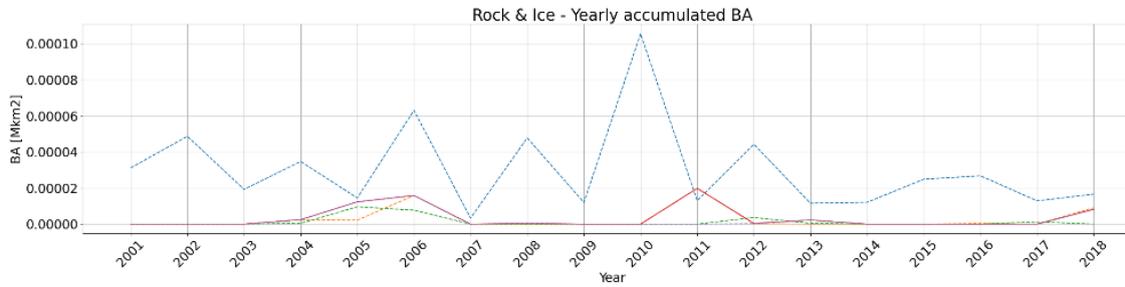
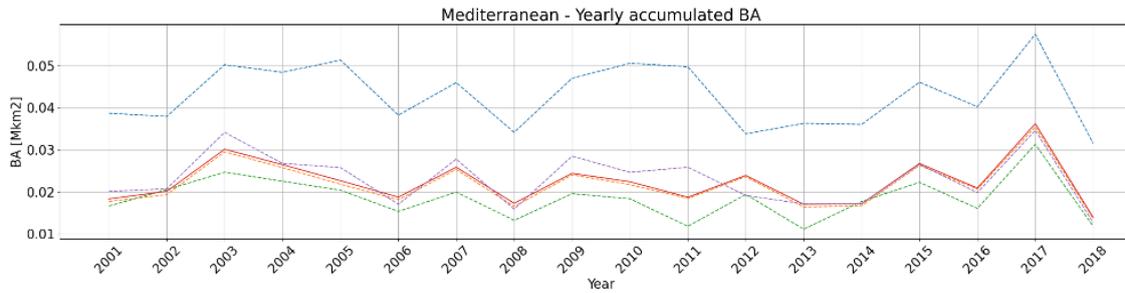
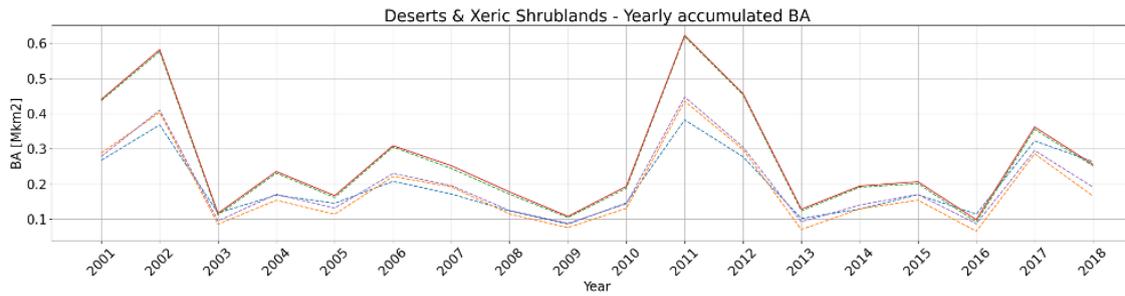
Ref.: Fire\_cci\_D2.2.3\_ATBD-MERGED\_v1.1

Issue 1.1

Date 18/07/2024

Page

22



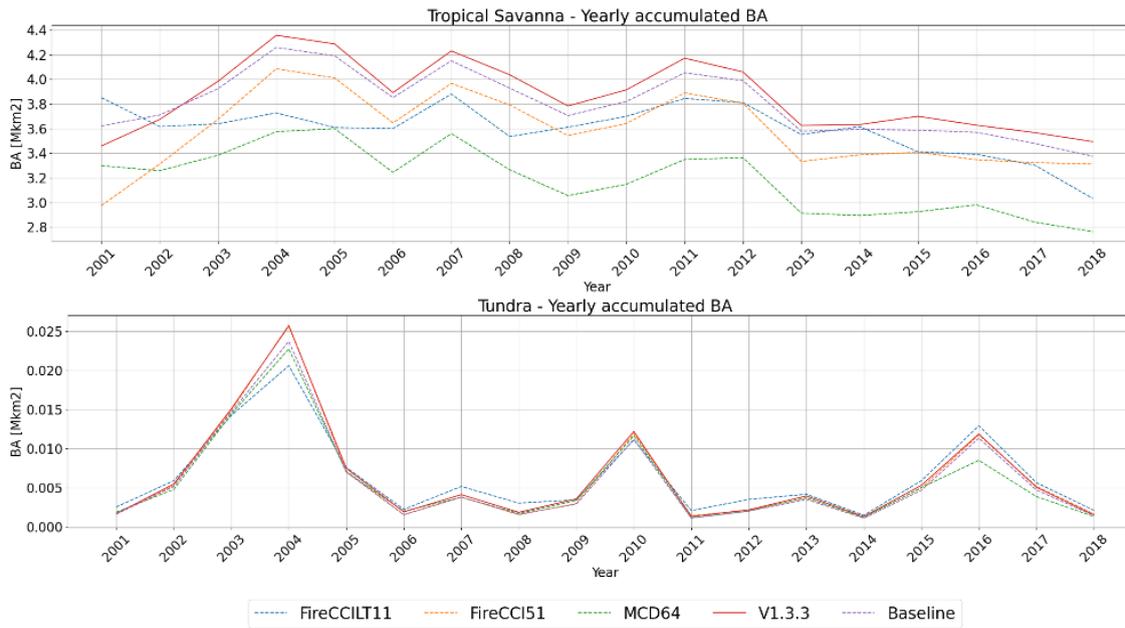


Figure 11: Average annual BA for each ecoregion

The average monthly BA series for the entire period 2001-2018 for all products and ecoregions considered by BA is shown in Figure 12: Average monthly BA for the period 2000-2018. Again, the MCD64CMQ product presents the largest differences against the rest of the products. The average BA for the whole period is included in Table 8. Furthermore, these values reflect that the new product overestimates the BA values given by the rest of the products, which cannot be considered to be an error, since different studies have shown that, in general, available BA products underestimate the real total burned area.

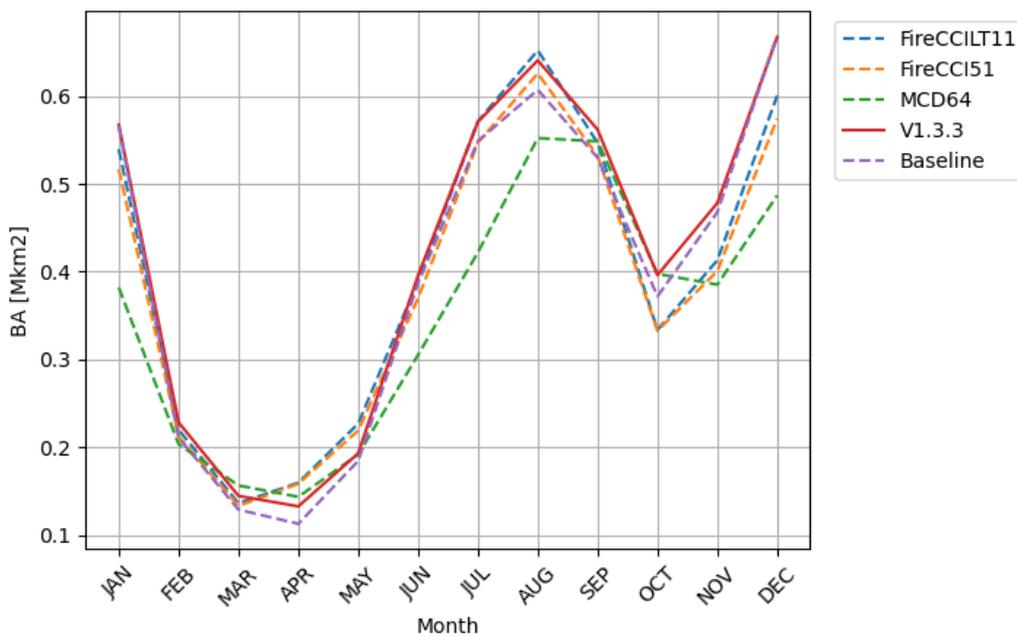


Figure 12: Average monthly BA for the period 2000-2018

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page	24		

Table 8: Average BA for the period 2000-2018

BA product	BA (M km2)
FireCCILT11	4.788
FireCCI51	4.621
MCD64CMQ	4.173
Baseline	4.783
V1.3.3	4.975

After the global evaluation of the new BA product, a more specific analysis for some years has been carried out. In particular, the training year 2008 and the testing years 2003 and 2017 were selected.

### 2.6.1 Evaluation of the BA estimation model for the year 2008

First, the 664 BARD cells susceptible to be compared with the estimated product were analysed. Table 9 summarises the results of the assessment. It should be pointed to the high agreement, almost 90%, and the higher overestimation error versus the underestimation one.

Table 9: Accuracy assessment of the estimated product against BARD for year 2008

Correct cells	Overestimation error	Underestimation error	% agreement
598	53 (80.30%)	13 (19.70%)	90.06

Since the rules have been established based on the training set of paired cells BARD-BA products, the model behaviour has been globally analysed for the training year 2008. The total burned area for each month in 2008 and for the three BA products used in the generation of the model and the new product generated was plotted (Figure 13: [Monthly BA for BA products \(2008\)](#)). The trend for the new product for this year is similar to the average for the entire period. It has a similar distribution to the other products, with the biggest difference in December, where the model predicts a higher BA for that month than the other products. Furthermore, the accumulated BA for the whole year 2008 was calculated and the values are included in Table 10. It can be observed that the largest difference between the new BA product and the source BA products corresponds to MCD64CMQ, while there is a better adjustment with the Fire BA products, the best with FireCCI51.

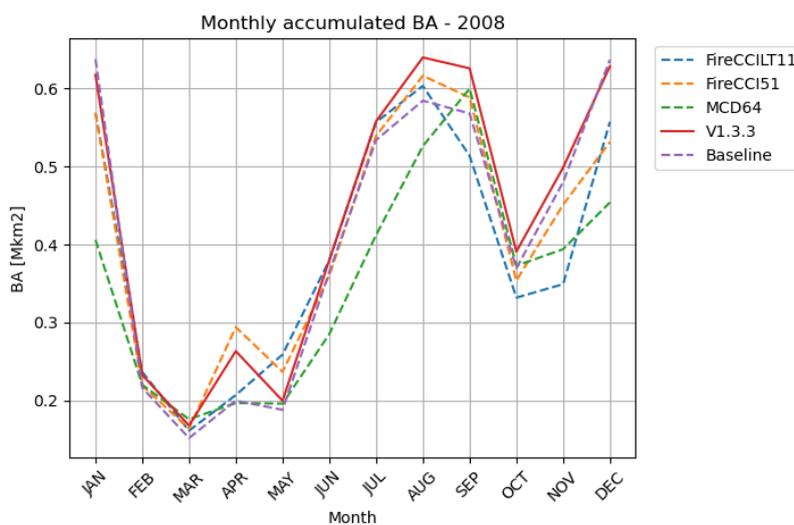


Figure 13: Monthly BA for BA products (2008)

Table 10: BA over the years 2008

BA product	BA (M km <sup>2</sup> )
FireCCILT11	4.721
FireCCI51	4.919
MCD64CMQ	4.236
Baseline	4.927
V1.3.3	5.199

Moreover, the distribution of the selected cells for each product was analysed, and the results are shown in Figure 14. This graph justifies the best adjustment of the new BA product with the FireCCI51 product. And also, that FireCCI51 product has the best adjustment with the BARD database of the BA source products used to generate the estimation model.

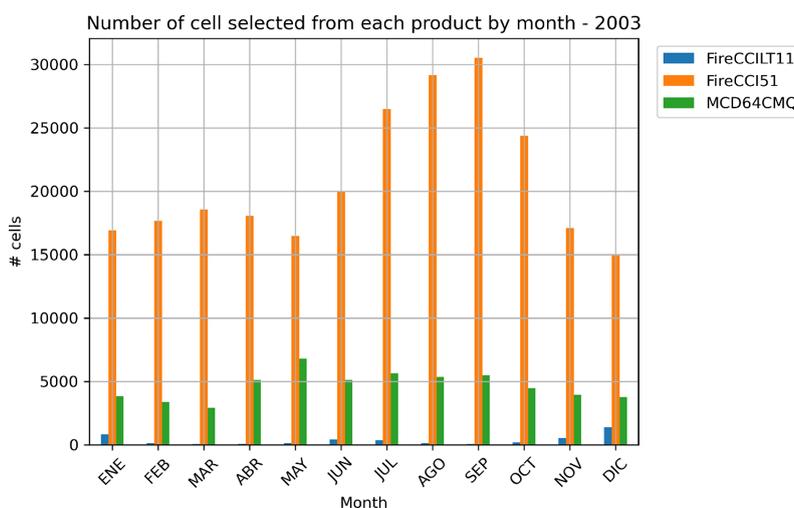


Figure 14: Monthly number of cells selected from each BA product (2008).

The previous analysis was also performed for each ecoregion. Figure 15 shows the comparison of the monthly-accumulated BA for each of the ecoregions.



fire  
cci

Fire\_cci  
Algorithm Theoretical Basis  
Document - MERGED

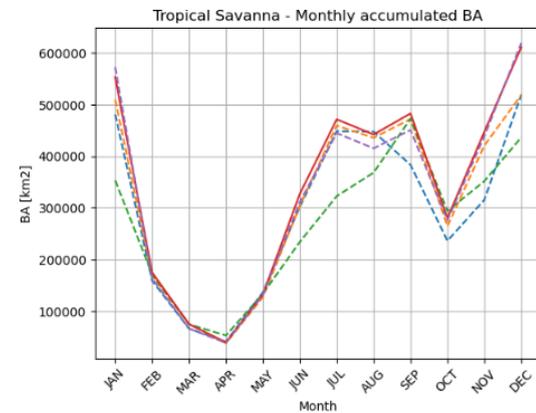
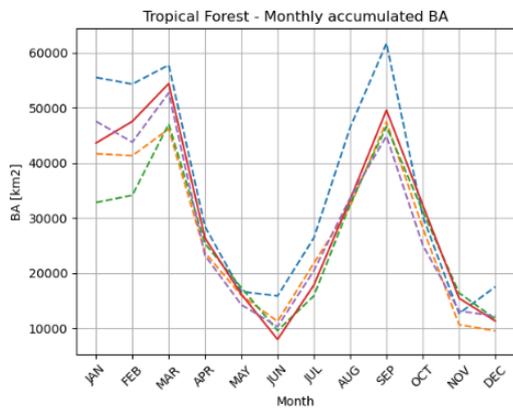
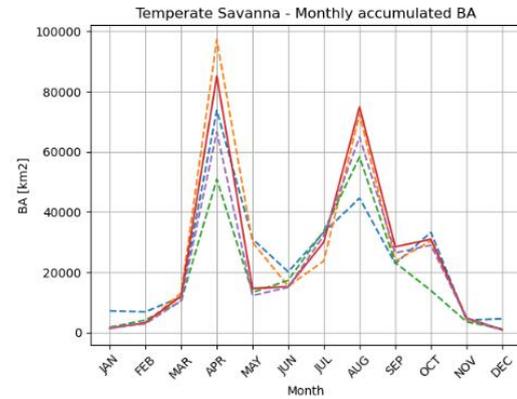
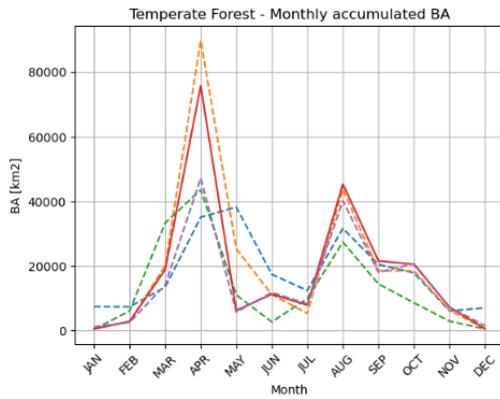
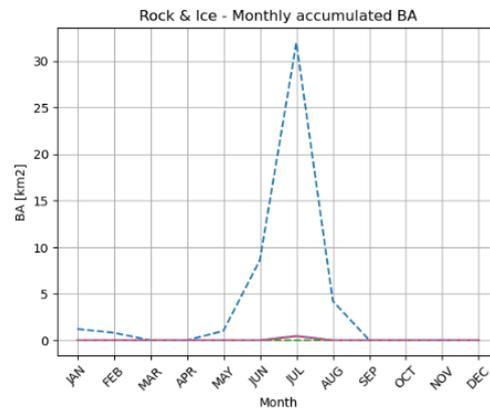
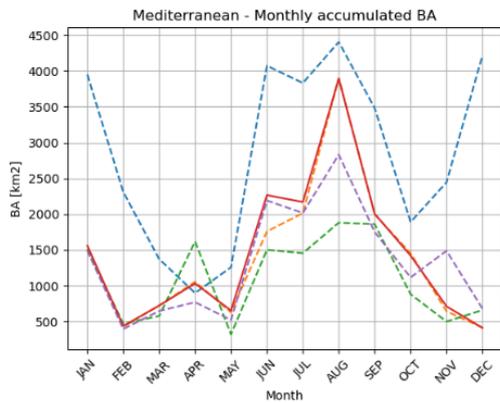
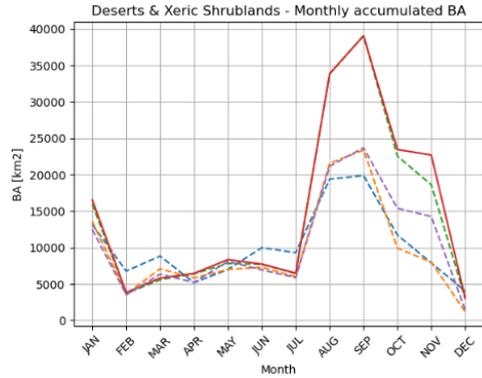
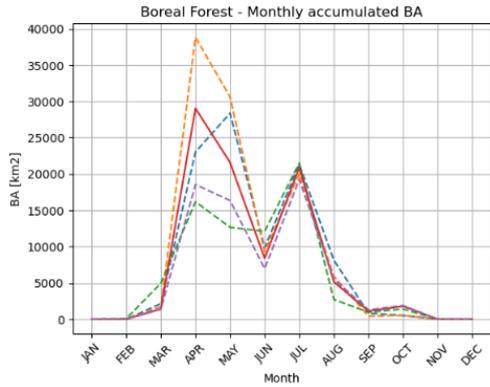
Ref.: Fire\_cci\_D2.2.3\_ATBD-MERGED\_v1.1

Issue 1.1

Date 18/07/2024

Page

26



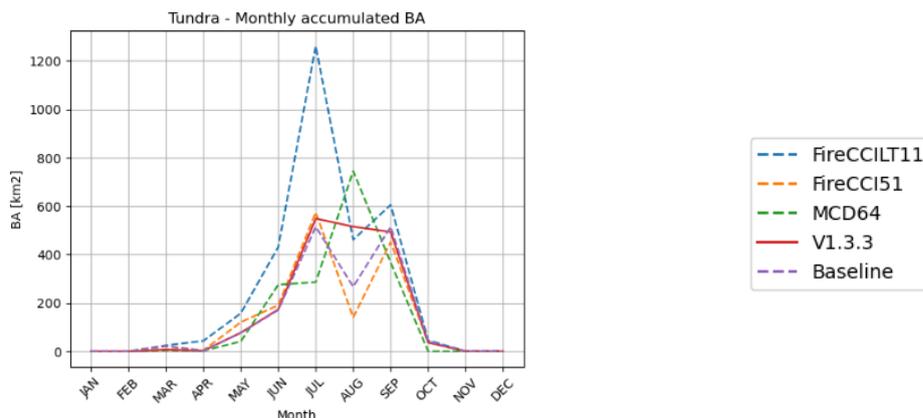


Figure 15: Comparison among BA products of the monthly BA for the different ecoregions (2008).

The graphs in Figure 15 show that the trend of the new BA product is similar to that of the source BA products, with some exceptions. As already mentioned, it is more similar to the FireCCI51 product, showing the biggest differences with the FireCCILT11 product, which shows a very different behaviour to the set of products in particular for some ecoregions (Mediterranean, Rock&Ice and Tundra) and months.

The results of the comparative analysis between BARD data and the BA values estimated by v1.3.3 are included in Figure 16. Figure 16 a) shows that for most cells a small deviation of less than 25 km<sup>2</sup> occur. As mentioned above, only cells with a higher difference than 2% of the total area are considered as error in the prediction. Then, 57 cells are considered failure in the prediction from a total of 628 cells (9.07%) and 571 are considered hits (90.93%). For the failures cells 46 cases are due to overestimation of the model, where in 11 of them the BARD covers more than one month and in 35 cases the BARD cover less than a month. There are also 11 cases of failure due to underestimation.

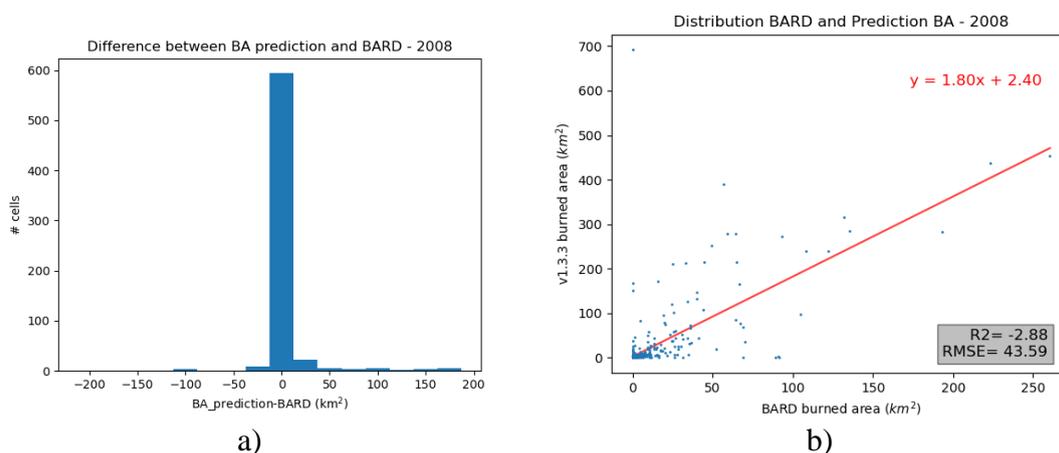


Figure 16: a) Distribution of the difference in the number of BA cells provided by the BARD database and estimated by the new product for year 2008. b) Scatterplot between BA provided by the BARD database against BA estimated by the new product.

## 2.6.2 Evaluation of the BA estimation model for years 2003 and 2017

The year 2003 is one of the test years of the BARD with more cells (204) susceptible to be used, so it is one of the best years to analyse the product. However, it should be pointed that it only contains data for the United States of America so it will not give us a worldwide information of the new BA product. Regarding the year 2017 (369 cells), it is

the only year for the BARD database that has global information, and it could be used as a test data. Perhaps this may be the reason why the percentage of agreement is much lower than for the year 2008 and 2003.

Table 11: Accuracy assessment of the estimated product against BARD for years 2003 and 2017

Year	Correct cells	Overestimation error	Underestimation error	% agreement
2003	195	7 (77.78%)	2 (22.22%)	95.59
2017	253	41 (35.34%)	75 (64.66%)	68.56

Figure 17 and Table 12, include the monthly accumulated BA and the yearly accumulated BA, respectively, for the years 2003 and 2017. A similar trend than the one obtained for the year 2008 can be observed.

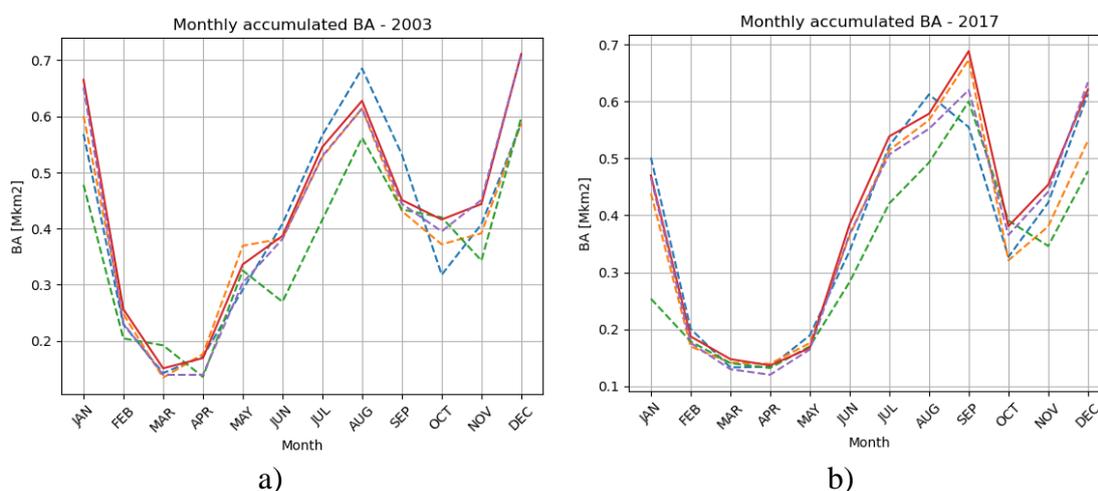


Figure 17: Monthly accumulated BA for BA products for the years 2003 (a) and 2017 (b).

Table 12: Year accumulated BA for the years 2003 and 2017

BA product	BA (M km <sup>2</sup> ) 2003	BA (M km <sup>2</sup> ) 2017
FireCCILT11	4.908	4.547
FireCCI51	4.834	4.416
MCD64CMQ	4.377	3.885
Baseline	4.990	4.543
New product	5.160	4.754

Furthermore, the monthly accumulated BA for each ecoregion for the two testing years has been plotted for the source BA products and for the baseline and version 1.3.3. The plots are included in Figure 18.



fire  
cci

Fire\_cci  
Algorithm Theoretical Basis  
Document - MERGED

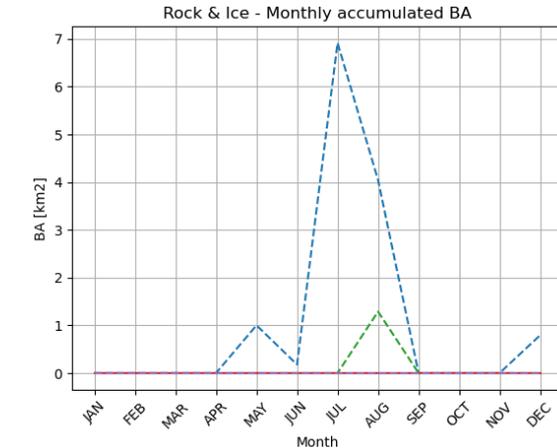
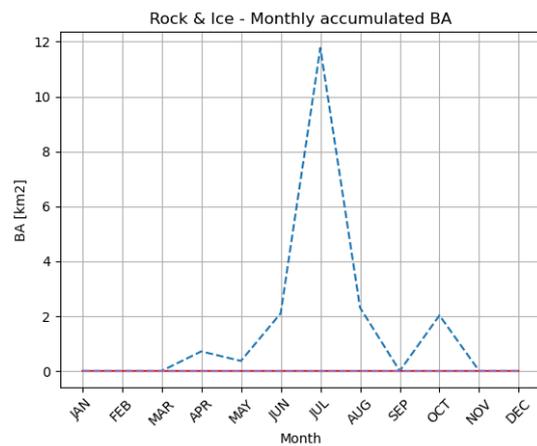
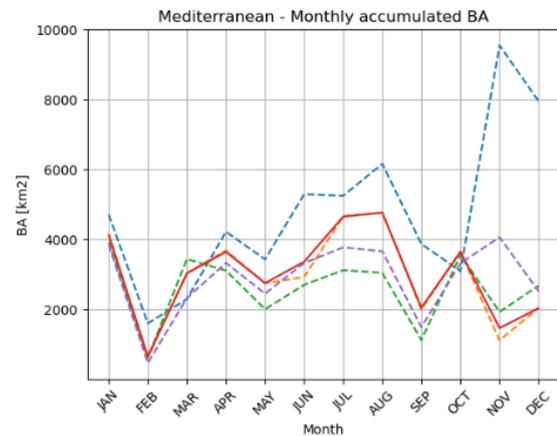
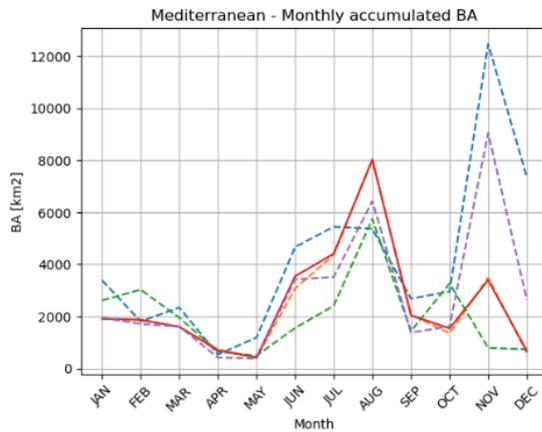
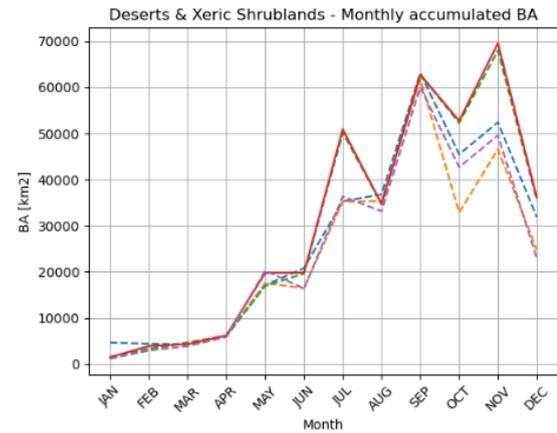
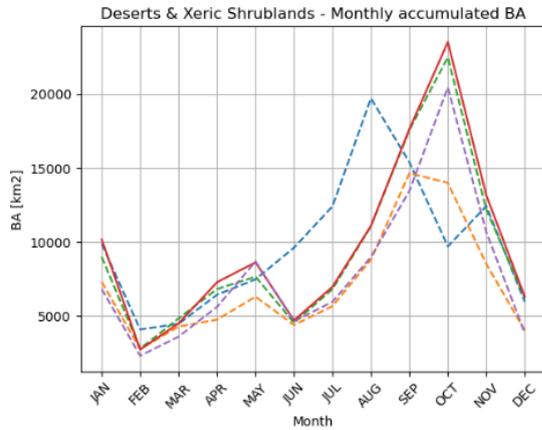
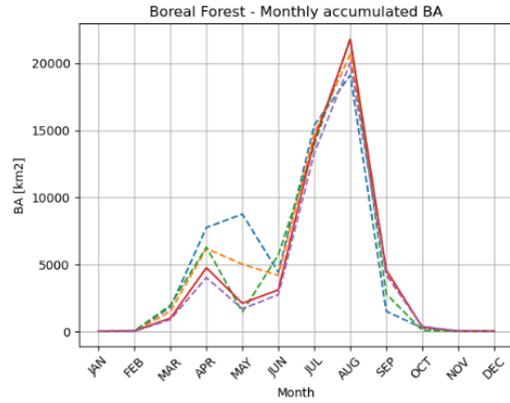
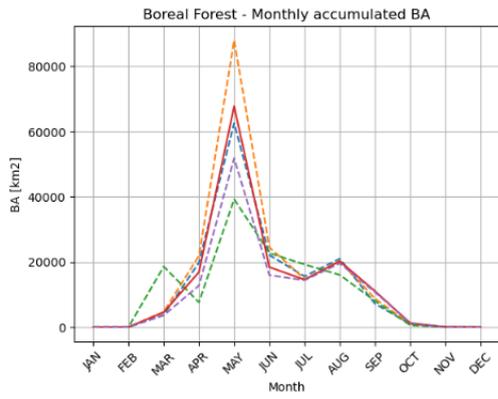
Ref.: Fire\_cci\_D2.2.3\_ATBD-MERGED\_v1.1

Issue 1.1

Date 18/07/2024

Page

29





fire  
cci

Fire\_cci  
Algorithm Theoretical Basis  
Document - MERGED

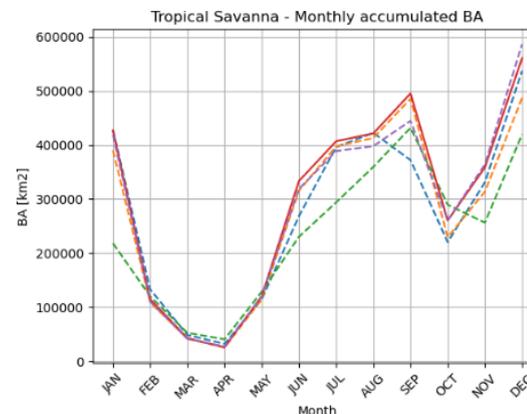
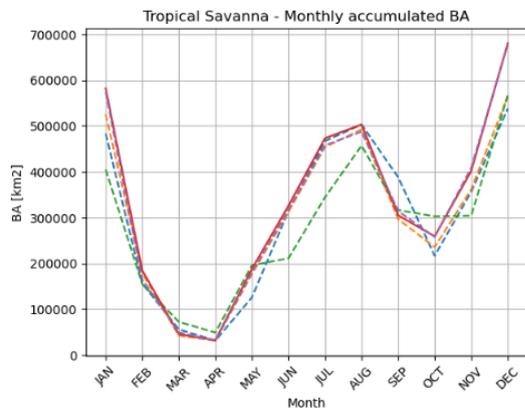
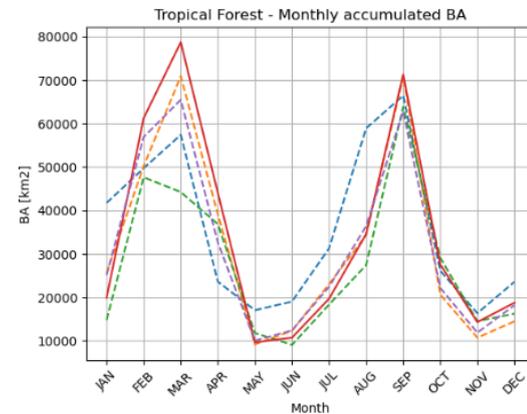
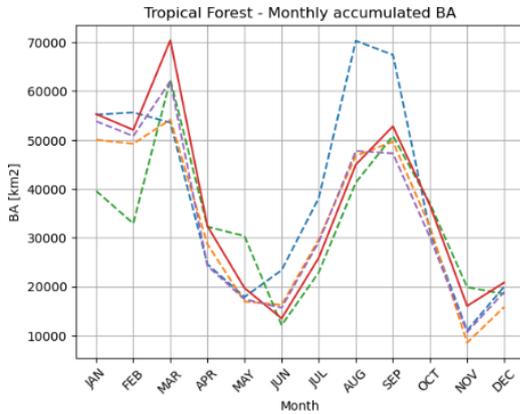
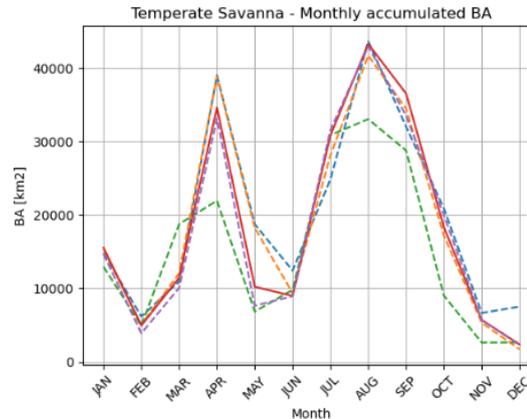
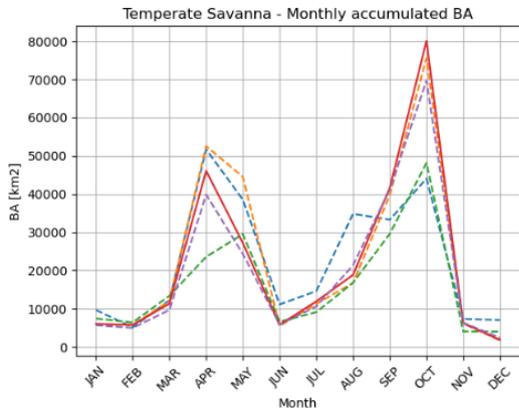
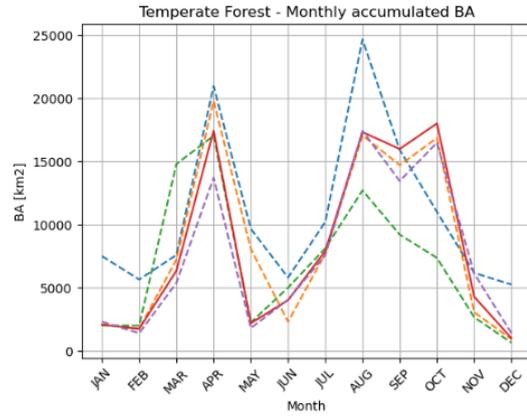
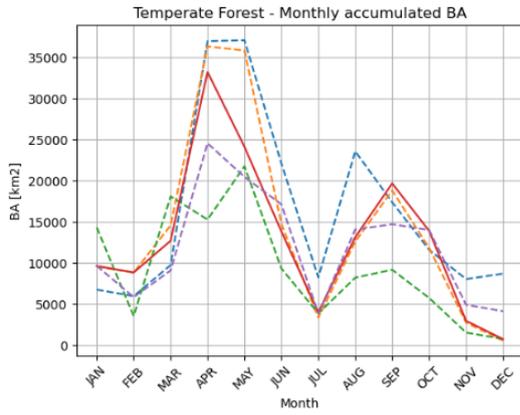
Ref.: Fire\_cci\_D2.2.3\_ATBD-MERGED\_v1.1

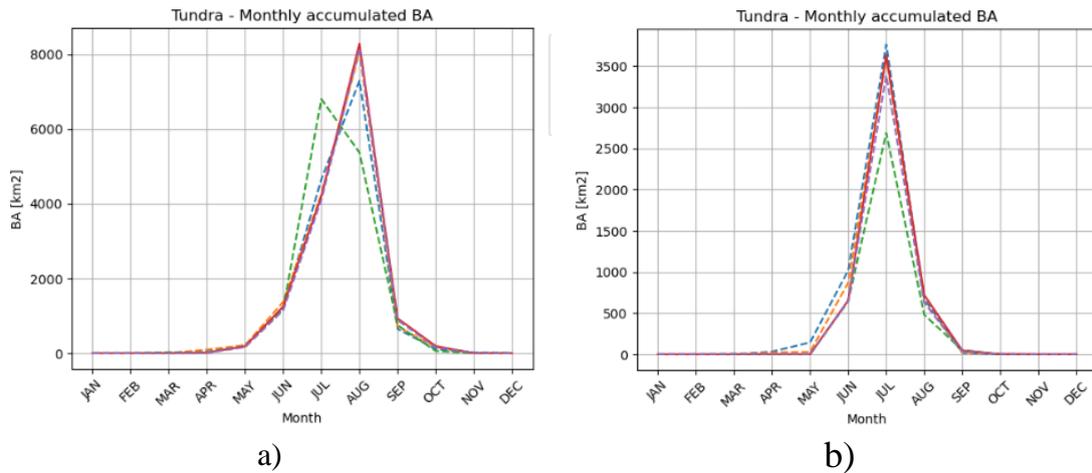
Issue 1.1

Date 18/07/2024

Page

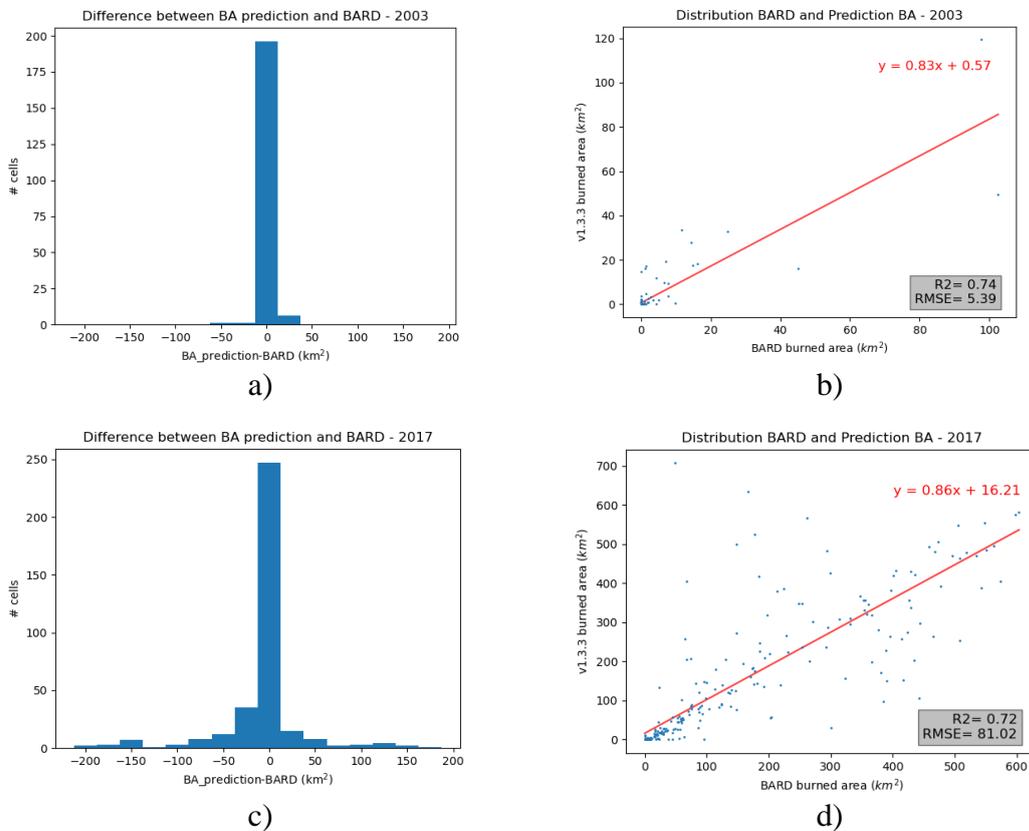
30





a) b)  
*Figure 18: Accumulated monthly BA for the year 2003 (a), and for the year 2017 (b).*

In Figure 18, a consistent behaviour for years 2003 and 2017 compared to that of the year 2008 can be observed. The results of the comparative analysis between BARD data and the BA values estimated by v1.3.3 for years 2003 and 2017 are included in Figure 19. The results obtained for 2003 are very similar to the ones obtained for the year 2008. Figure 19 a) shows that for most cells a small deviation of less than 25 km<sup>2</sup> occurs. However, the results showed for year 2017 (Figure 19 c and d) are slightly different, appearing a large number of cell with higher differences between the BARD and the BA estimate by V1.3.3.



a) b)  
c) d)  
*Figure 19: Distribution of the difference in the number of BA cells provided by the BARD database and estimated by the new product for years 2003 (a) and 2017 (b). Scatterplot between BA provided by the BARD database against BA estimated by the new product for years 2003 (c) and 2017 (d).*

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page			32

## 2.7 Phase 2 (1982-2018)

As illustrated in Figure 2, the aim of the second phase is to generate a single long-term series of BA with a temporal extension from 1982 to 2018. Therefore, after developing the short-merged series v1.3.3 for the years 2001-2018, it is necessary to extend it back. Now the main goal is to develop a methodology that allows us to generate the final BA product, FireCCIM10. Inspired by the product GFED (Giglio et al., 2010), we decided to use as input data the only BA product available for the period (1982-1999), FireCCILT11, as well as other datasets with long time series of climatic information such as the ERA5-Land monthly dataset<sup>3</sup> and Fire Danger Index<sup>4</sup>. The last two data sets have proved in the first phase to be useful for the prediction of the BA value.

To generate this final, extend series from 1982 to 2018, it was first proposed to use regression machine learning models that can establish the BA value for each cell using as input the datasets mentioned before and FireCCILT11. Since this information was not enough for training the regression models for all cells, more complex models such as the random forest regression (RFR) were also considered.

Firstly, a local second-degree polynomial regression was performed using the BA values from FireCCILT11 and from the Fire Danger Index, using the variables KDBI min, FDI max, DC min, SC mean (see Annex 2). These four variables proved to be the variables with better performance applying the next nonlinear regression.

$$BA(i, m) = \alpha(i)VAR(i, m)^{\beta(i)} \quad \text{Equation 1}$$

Where,  $BA(i, m)$  represents the BA value of  $i^{th}$ -cell at month  $m$ ; and  $\alpha(i)$  and  $\beta(i)$  represent the model parameters for each variable. For the training of the regression, the BA estimation from v1.3.3 was used as a target, but only the BA data when the FOA is higher than 80% were considered for training. In addition, it should be noted that the minimum number of values to perform the training of the regression for each cell and for each month is eight. Once the regression model is generated, the estimated BA for the cell in the pre-MODIS era for each month is obtained, as well as the associated confidence interval.

To explore the results obtained after applying this method, particular grid cells were considered to analyse the corresponding BA long time series. Figure 20 shows the long-term BA series estimated by the local regression model for two particular grid cells, for two different ecoregions (Mediterranean and Tropical Forest) and for the months of January and July, respectively. Moreover, in this figure the BA values provided by the FireCCILT11 and V1.3.3 products, for comparison, as well as the confidence level (CL) have been included. The geographical coordinates of the cell are included in the figure. It is observed in Figure 20, that better CL is obtained in the MODIS era than in the pre-MODIS era. It can be observed also that there is a not negligible number of estimations that are lower than zero, which does not have physical meaning. Moreover, for some cells BA values higher than maximum possible values have been estimated by the regression. In summary, the main problems we have found are: (i) it is not possible to apply the regression model to all cells, so there is no estimated BA for a large set of cells; (ii) for

<sup>3</sup> <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means?tab=overview>

<sup>4</sup> <https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-fire-historical-v1?tab=overview>

some cells a negative BA prediction has been obtained; and (iii) for some cells the estimated BA is higher than the maximum burnable area for a cell.

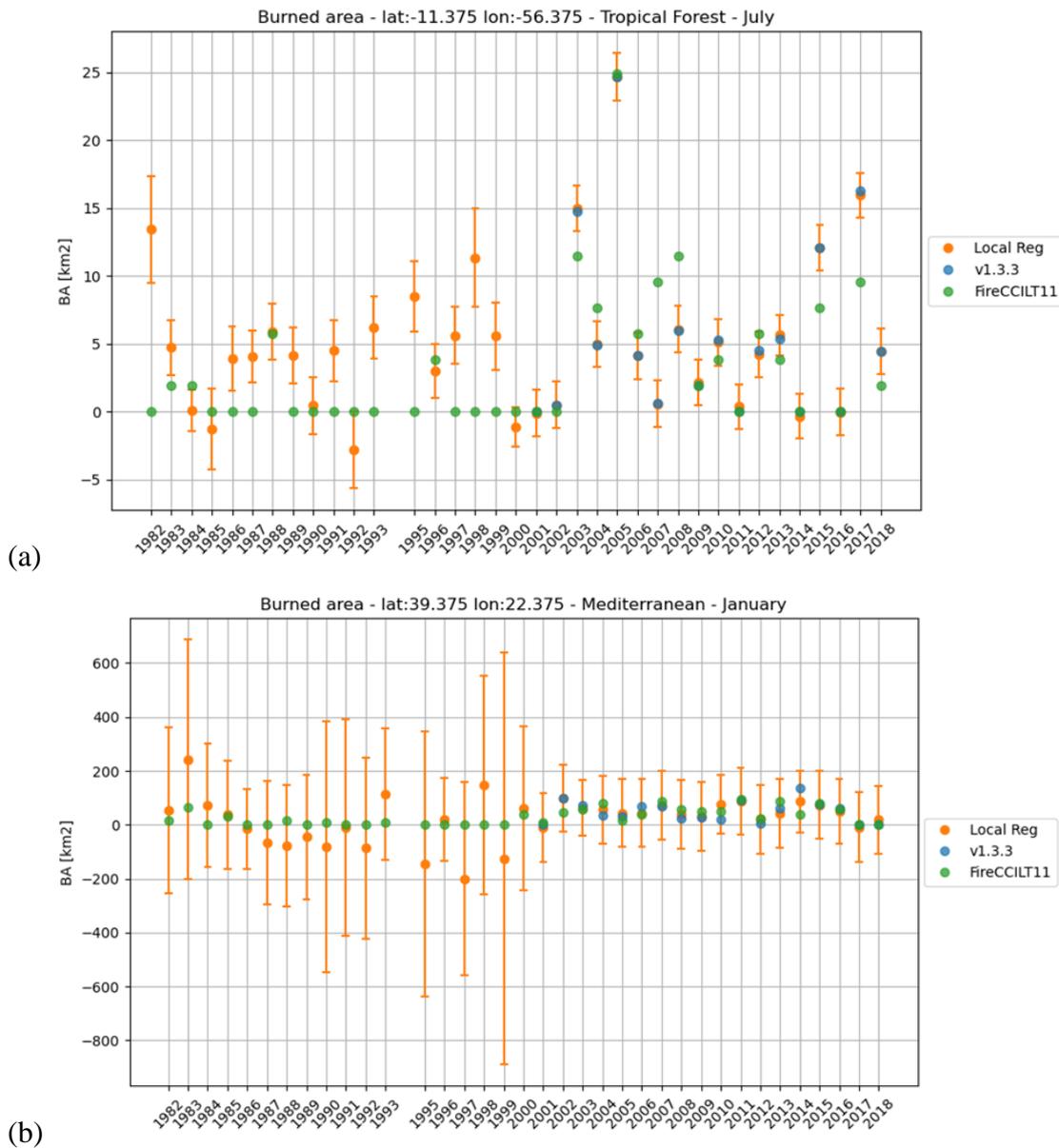


Figure 20: Long-term BA series estimated by the local regression model for two particular grid cells, for two different ecoregions (Mediterranean and Tropical Forest) and for the months of July (a) and January (b)

With the aim of addressing the first problem, it was decided to train a random forest regression for estimating the BA value for the cells that the local regression does not calculate. Since fire does not have the same behaviour throughout the world, there are places where it is more likely to have fires in December than in July and vice versa; weather seasons in the northern and southern hemisphere do not follow the same pattern etc. For this reason, different models have been built for each Olson ecoregion, hemisphere, and month. Considering the eight burnable ecoregions established by Olson, (Rock & Ice ecoregion does not get burned, so it does not need to have a model), and that the Boreal Forest ecoregion is only present in the north hemisphere, there is a total of 15 different ecoregion-hemispheres. As is mentioned before, for each month we will have a

different Random Forest Regressor, so in the end we will have to train 180 different Random Forest Regressors. During the training phase, the same input as for the local regression model (FireCCILT11 (BA, FOA, FBA), different meteorological variables provided in the database ERA5-Land and Fire Danger Indices) were used. But with the aim of considering the temporal context, the values of these variables have been considered for three months (previous, current, and following). The associated output for each input pattern has been the estimated BA (V1.3.3) for this cell and month. All these RFR models were trained for the MODIS era, using as testing the data for the years 2003, 2013, 2016 and 2017. They were trained using GridSearchCV to find the best features for each case. In most models, the feature of the greatest importance, by far, is normally the Burned Area value from FireCCILT11 of the month to be predicted (ba\_f11\_2), as shown in Figure 21, which shows some examples of the 180 RFR. The main problem with these models is that they always estimate a small value of BA instead of 0. That is why we could usually find patches of areas that have a small estimation of BA.

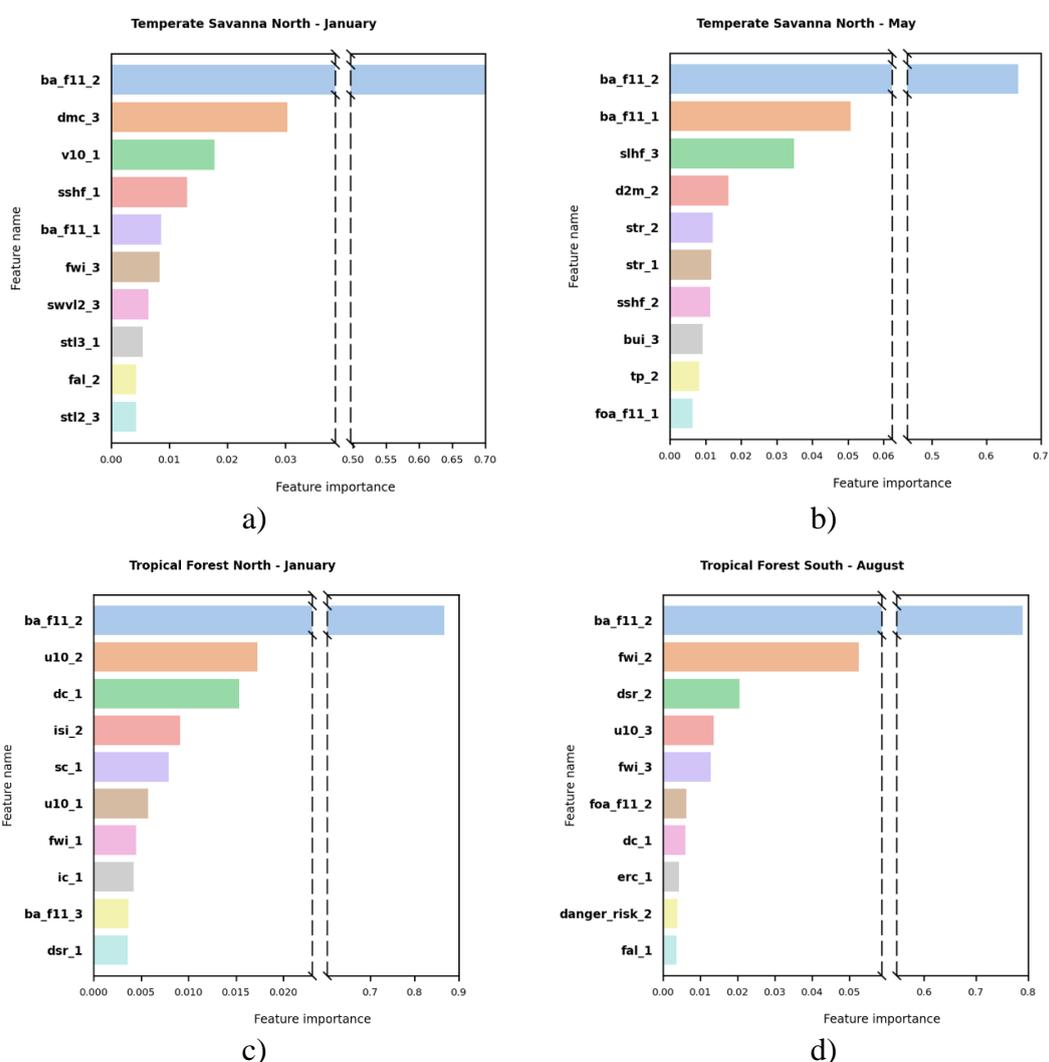


Figure 21: Results of the analysis of the importance of the characteristics for different biomes and months. a) Boreal Forest north in January. b) Temperature savanna in May. c) Tropical forest north in January. d) Tropical forest south in August. (see Annex 2 for a variable abbreviation list). Numbers 1, 2 and 3 correspond to the previous, current, and following months.

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page			35

### 2.7.1 Generation of Single Long-Term BA Series

Two different approaches have been taken to generate the Single Long-Term BA Series, which will end with the generation of the final BA product, FireCCIM10. These approaches differ in which of the two previously explained methodologies will be applied first. In the first approach (SLT-1) the main way to estimate the burned area is using the local polynomial regression (LPR). In the cells where this regression cannot be applied because there is not enough data for the training, the BA value the estimation was obtained from the Random Forest Regressor (RFR). It was found that for the MODIS period, both annual and monthly SLT1 estimates are very similar to the other products. Meanwhile, in the pre-MODIS the BA estimation from SLT1 is much higher, as the annual estimations in average are close to 12Mkm<sup>2</sup>.

Attempting to improve the estimation, a second approach, SLT-2, was implemented. In this second approach, RFR predictions were used to estimate the bulk of cells to generate the time series. In those cells where it cannot be applied, we have used de polynomial regression. In both approaches, there are some cells, mainly the ones on the coast, where we will still have a null value. This issue is something that will be treated when we decide which approach has a better performance.

A scheme of training the 180 RFR models and the estimation of the BA and CL is illustrated in Figure 22. The same set of features input that were determined by the feature selection process is used in both processes. During training, the target is BA estimated by V1.3.3 for the current month, and the corresponding ecoregion and hemisphere. After training, the model is used to estimate the BA and CL of the FireCCIM10 product. Since these models cannot estimate the BA values for all cells at any time, the LPR is applied to them. Figure 22 illustrates this process for training and estimating BA for particular cells.

As in the previous approach, post-processing will be done to avoid errors in the BA values, such as negative estimations or predictions higher than the fraction of burnable area. But the main issue that has been found using this methodology is the small deviation that the RFR predictions have when the cell does not get burned. For most of the RFR models we will not get a prediction of 0, so it is needed to filter them. For this first study of this approach, the filter was in 0.9 km<sup>2</sup>, so all cells where the BA value is less than a squared kilometre will be adjusted to 0.

Figure 23 shows a smaller disparity amongst the BA values estimated by the new product and the previous BA products than for the approach SLT1. Also, it can be appreciated that the new BA product estimates BA values slightly higher than the other BA products in the whole time series, except for the case of product V1.3.3 in the MODIS era. This behaviour is also appreciated when the monthly BA values have been represented for the entire period (Figure 24) and separating the pre-MODIS era (Figure 25) and the MODIS era (Figure 26).

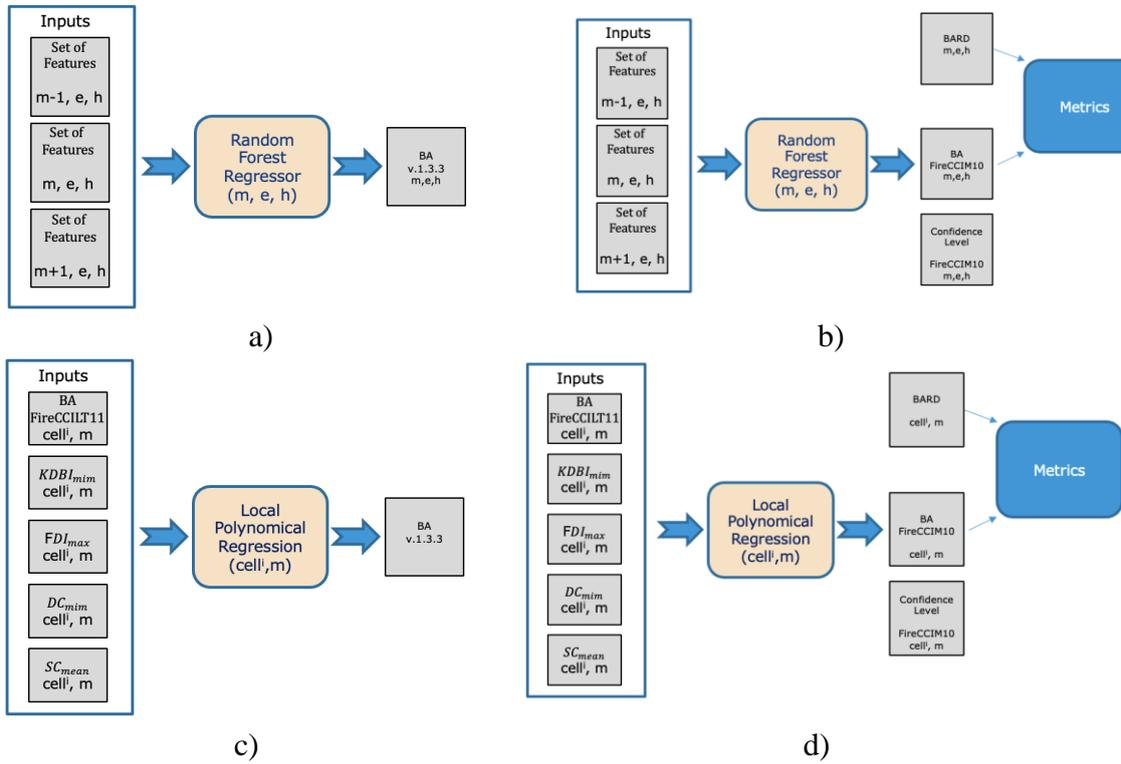


Figure 22: First row: BA prediction model based on the Random Forest Regressor. a) Training scheme. b) Generation of the product FireCCIM10 and validation ( $m$ : month,  $e$ : ecoregion,  $h$ : hemisphere). Second row: BA prediction model based on local polynomial regression. c) Training scheme. d) Generation of the FireCCIM10 product and validation.

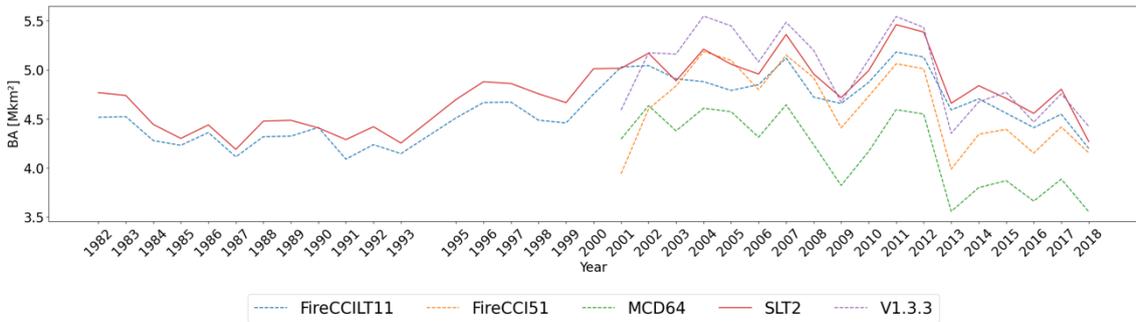


Figure 23: Annual BA estimation – SLT2.

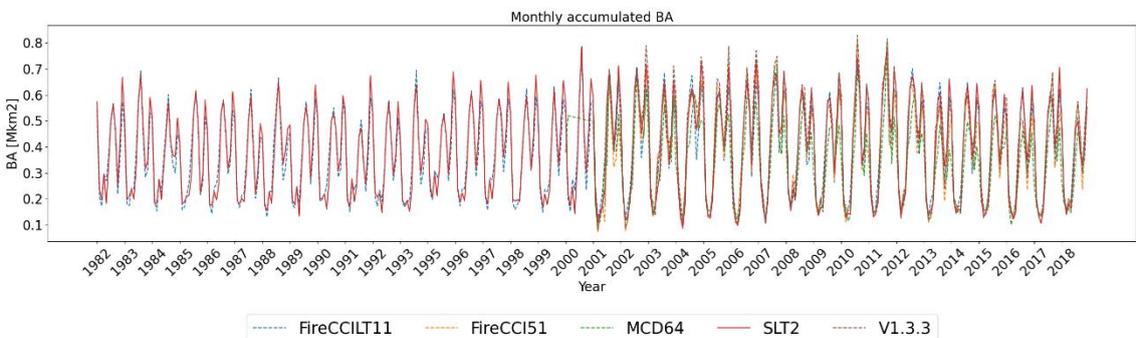


Figure 24: Monthly BA estimation – SLT2.

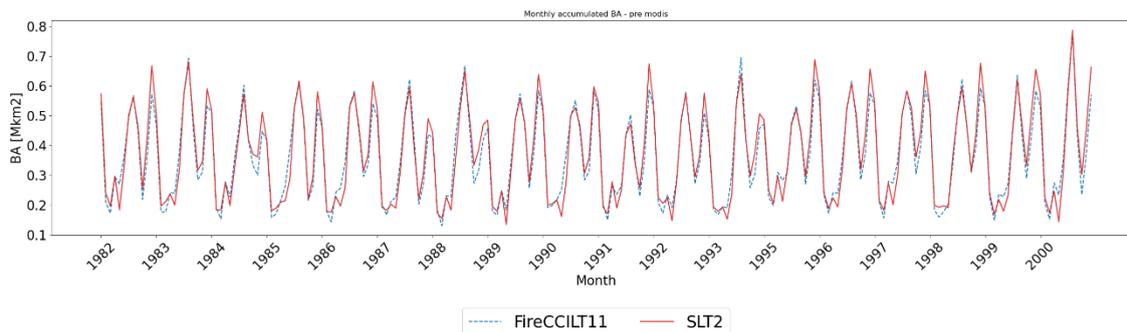


Figure 25: Monthly BA estimation pre-MODIS era – SLT2.

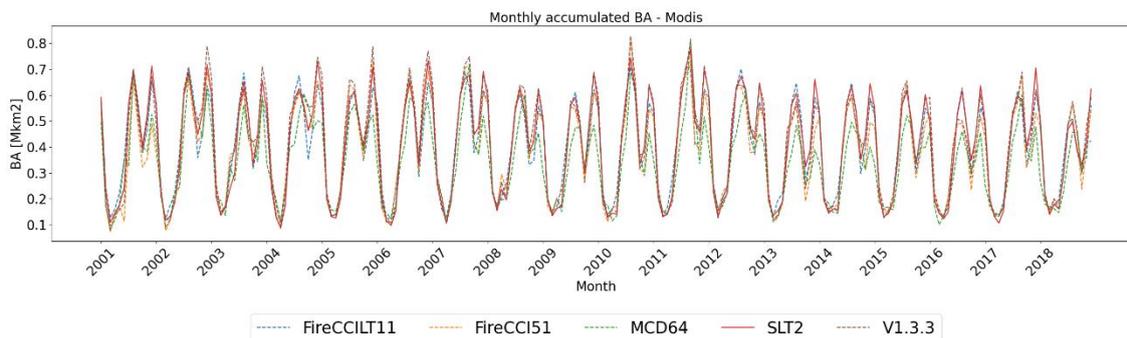


Figure 26: Monthly BA estimation MODIS era – SLT2.

By the same evaluation methodology described in Section 2.6, comparing the BA provided by the reference database BARD and the source and estimated BA products, the new time series estimated by approaching SLT-1 and SLT-2 has been evaluated. Table 13 and Table 14 summarise the results of the comparative assessment for the MODIS era.

Table 13: Results of the comparative assessment in the training phase for the BA source products and estimation models in the MODIS era.

Version	Training						
	Correct cells	Overest. error	Underst. error	% agreem.	MAE	R <sup>2</sup>	MSE
<b>FireCCILT11</b>	871	111	141	77.56	22.056	0.7279	3726.3345
<b>FireCC51</b>	928	62	133	82.64	14.476	0.8665	1827.7796
<b>MCD64CMQ</b>	928	62	133	82.64	16.957	0.7748	3083.6470
<b>Baseline</b>	926	60	137	82.46	16.227	0.8282	2352.9444
<b>V1.3.3</b>	931	72	120	82.90	13.574	0.8575	1951.8096
<b>SLT-1</b>	923	79	121	82.19	14.695	0.8465	2102.7625
<b>SLT-2</b>	896	91	136	79.79	19.856	0.7750	3082.1291

From the results, it can be observed that during the training phase and the testing phase SLT-1 presents better performance than SLT-2. When comparing the new estimated long time series (SLT-1 and SLT-2) with the source BA products, it can be observed that both outperform FireCCILT11 and provide assessment values very close to FireCCI51 and MCD64CMQ, with an average agreement rate of 80%. Meanwhile, the SLT-2 approach for the training dataset (years: 2008 and 2018) had a better behaviour than the FireCCILT11, in terms of percentage of agreement and MAE, but for the testing dataset

	<b>Fire_cci</b>		Ref.: Fire_cci_D2.2.3_ATBD-MERGED_v1.1
	<b>Algorithm Theoretical Basis</b>		Issue 1.1      Date 18/07/2024
	<b>Document - MERGED</b>		Page 38

(years: 2003, 2013, 2016 and 2017) the product FireCCILT11 had a better MAE but a worse agreement than SLT-2.

Table 14: Comparative assessment results in the testing phase for the source BA products and estimation models in the MODIS era.

Version	Testing						
	Correct cells	Overest. error	Underst. error	% agreem.	MAE	R <sup>2</sup>	MSE
<b>FireCCILT11</b>	480	119	94	69.26	27.778	0.6842	4583.7895
<b>FireCC51</b>	563	50	80	81.24	18.781	0.8665	3472.8879
<b>MCD64CMQ</b>	551	42	100	79.36	25.281	0.6187	5533.6586
<b>Baseline</b>	550	54	89	79.36	21.879	0.6842	4583.3198
<b>V1.3.3</b>	558	57	78	80.52	18.901	0.7514	3608.5446
<b>SLT-1</b>	535	85	73	77.20	21.058	0.7366	3822.9606
<b>SLT-2</b>	501	99	93	72.01	29.059	0.6282	5395.2580

Also, an analysis of the performance of these approaches have been done for the Pre-MODIS era, where the BARD only provide data from the United States of America for the years 1988, 1993 and 1998 (Table 15). As expected from the observation from the SLT-1 time series, the performance of this approach is the one with the worst results. On the other hand, the results obtained for the SLT-2 approach show a good performance of it during this period with even better results than the FireCCILT11 product.

Table 15: Comparative assessment results for CONUS\_BA product in pre-MODIS era (years 1988, 1993 and 1998)

Version	Pre-MODIS era: CONUS_BA (USA): 1988, 1993 and 1998						
	Correct cells	Overest. error	Underst. error	% agreem.	MAE	R <sup>2</sup>	MSE
<b>FireCCILT11</b>	439	41	4	90.70	3.4544	-3.3920	94.6764
<b>SLT-1</b>	414	66	4	85.54	24.3541	-0.0469	11861.46
<b>SLT-2</b>	466	15	3	96.28	1.9247	0.2891	70.8775

### 2.7.2 Confidence level estimation

The confidence level (CL) layer included in the product FireCCIM10 has been computed by the method proposed in (Wagner et al., 2014), using the Python implementation provided by the library Forestci <https://contrib.scikit-learn.org/forest-confidence-interval/index.html>. In this method, the jackknife-after-bootstrap and infinitesimal jackknife (IJ) methods (Efron, 1992, 2013) for estimating the variance of bagged predictors are analysed, demonstrating that both estimators suffer from considerable Monte Carlo bias, and proposing a bias- corrected versions of the methods. Thus, it proposes to estimate the variance (CL) by the expression:

$$CL = \sum_{i=1}^n (C_i^{*,2} - \overline{C_i^{*,2}})^2$$

Where  $n$  the number of training samples, and  $C_i^* = Cov_*[N_{bi}^*, t_b^*(x)]$  is the covariance between  $t_b^*(x)$  and the number of times  $N_{bi}^*$  the  $i$ th training example appears in a bootstrap sample.

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page			39

### 2.7.3 Conclusion

The results obtained show that for the SLT-1 approach we are experiencing an overfitting with a remarkable performance during the MODIS era, but these results are not consistent for the pre-MODIS period. If this inconsistency can be solved, this local regression approach might be a good method to be able to extend the BA series. But due to the results and the consistency obtained in the BA series from the SLT-2 approach, this is the approach that will be used for the generation of the Single Long-Term BA Series, FireCCIM10. As can be seen in the following table, the SLT-2 approach had a better behaviour according to BARD for both the MODIS and pre-MODIS periods than the FireCCILT11.

### 2.8 FireCCIM10 Product

As explained above, 180 different RFR have been trained to obtain the estimation of the BA. This training was done in the MODIS era, using as features the data provided by FireCCILT11 (BA, FOA, FBA), Era5 Land data set and Fire Danger Indices, and as reference the BA estimate from the product v1.3.3. This RFR uses the data from the previous, the following and the current month to predict BA. Using the Python library Forestci we were able to obtain the confidence interval of the predictions done using this RFR. The years 2003, 2013, 2016 and 2017 will be used as testing of the models.

As mentioned before, there will be cells where some of the features of the RFR do not have data, so in those cells the RFR models can be applied. For these cases, the local polynomial regression will be applied to estimate the BA, and the confidence interval will be obtained too.

Despite of using the RFR and the local regressions, there will be some cells located normally in the coast where neither of both can be applied. This situation is produced due to the lack of data on the features of the models. To solve this problem, filling the missing data of the cells with the mean value of the neighbours, that is, the cells surrounding them. To be able to generate the BA estimation after filling the missing data, the RFR will be applied, also getting the confidence interval.

Cells that are entirely covered by water bodies are not processed either the RFR or local polynomial regressor, as their BA values will be 0 as there are not burnable areas. This also happens in the cells located in the Rock & Ice ecoregion. For these cells, the confidence interval will be 0.

In this way, all cells can be processed and will have BA and confidence interval (CI) values. Over these predictions a post-processing had been done to assure that the obtained BA estimations are not negatives or higher than the maximum burnable area. Also, a filter was needed to avoid the introduction of a small error in the predictions made by the RFR. These models instead of predicting 0, always estimate a small value. Therefore, to avoid this problem, the filter was established at 0.9 km<sup>2</sup>.

NetCDF (NC) files have been generated for each month from 1982 to 2018, except for 1994 (no data from FireCCILT11). These NC files have 3 different layers:

- Burned\_area (BA): the value of BA estimated for the cell.
- Confidence\_interval (CI): confidence interval of the BA estimation.

- Filter\_apply (FA): this layer only includes the values 1 and 0. 1 when the filter is applied, 0 for the cells where it is not.

Two types of NC files have been generated: in the first one the confidence layer was masked in the areas where the BA is 0, while in the second one the confidence layer was not.

### 2.8.1 FireCCIM10 assessment

The same methodology described above has been applied to compare the BA values estimated by the FireCCIM10 product with the BA values provided by the BARD reference database. Table 13 to Table 15 provide the results of the FireCCIM10 accuracy assessment, considering that it has been generated from the SLT2 approach. As is reflected in them, the accuracy of FireCCIM10 is better than the product FireCCILT11 for the whole series. Figure 27 includes the cell distribution of the differences between the BA estimated by FireCCIM10 and the BARD. It can be observed that for most cells the estimation is almost the same as the reference data; only few cells present a difference larger than 50 km<sup>2</sup>, which is a value lower than the maximum BA of a grid cell.

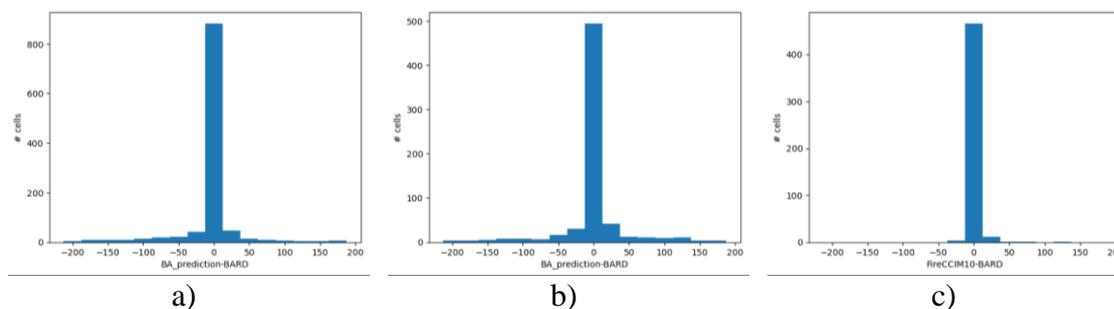


Figure 27: Distribution of BA differences between FireCCIM10 and BARD (km<sup>2</sup>). a) Training phase, b) testing phase, and c) pre-MODIS era.

To analyse the BA values estimated by FireCCIM10 with the rest of the products, the monthly average BA for the entire time series (1982-2018) was estimated and represented in Figure 28. The largest differences between the estimated values by FireCCIM10 and other BA products correspond to FireCCIM10 and MCD64, changing the differences through the different months. And the lowest difference corresponds to the FireCCI51 product.

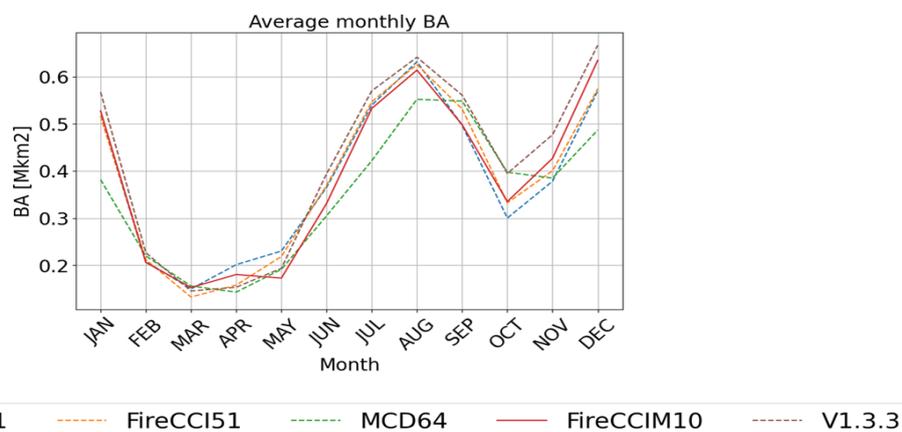
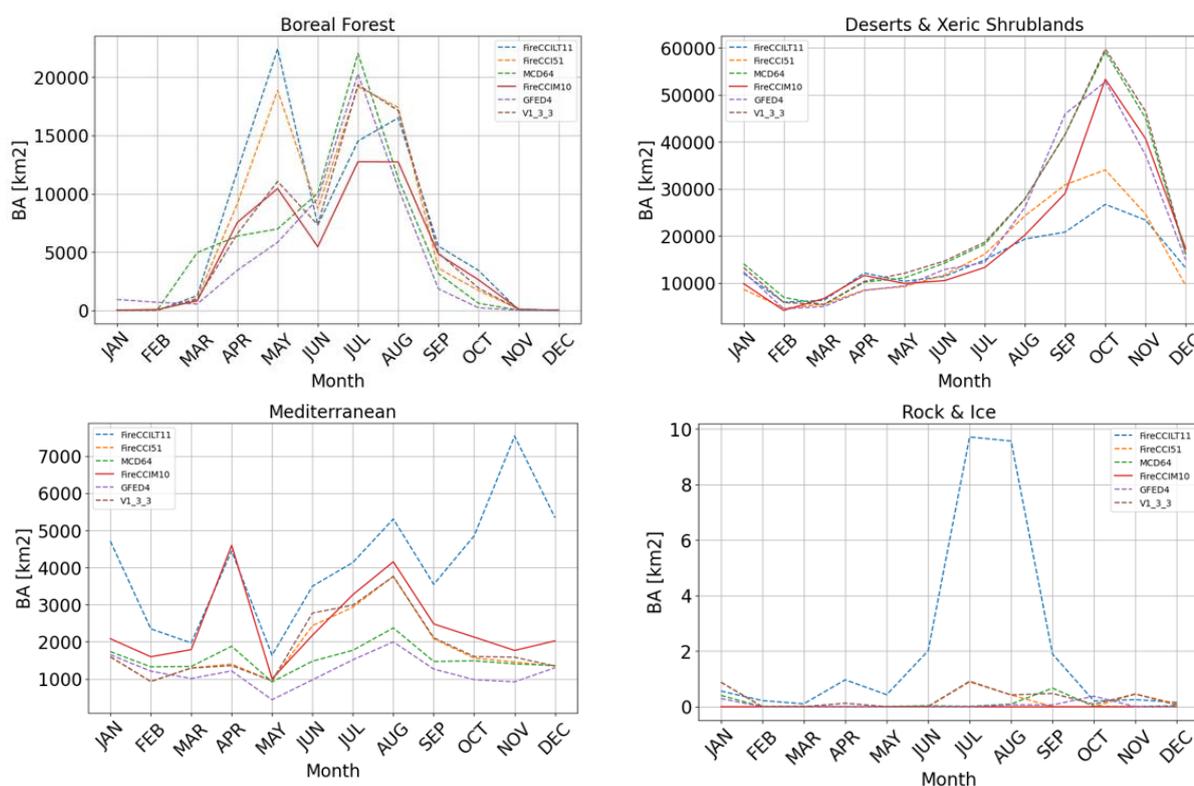
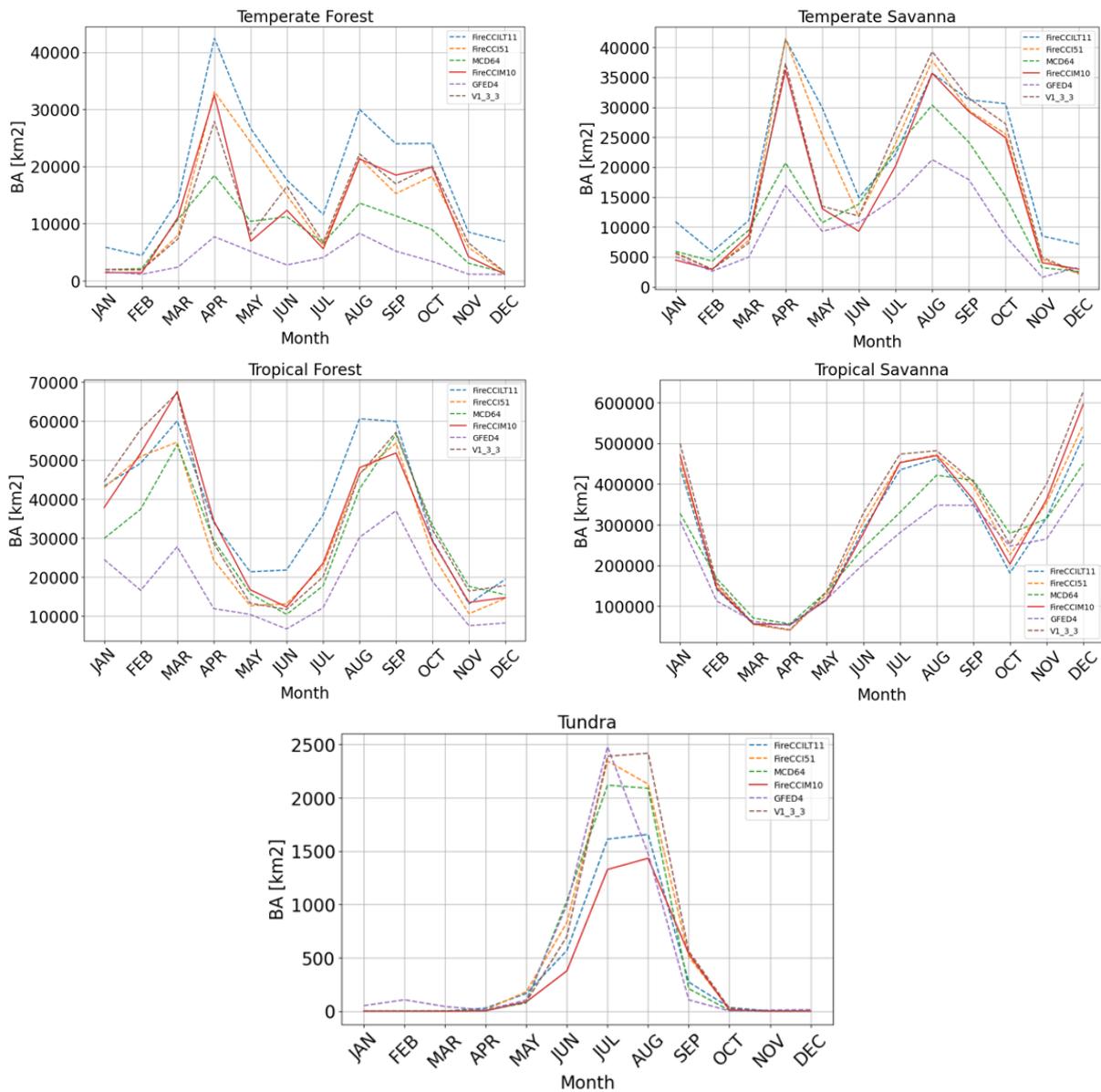


Figure 28: Average monthly BA time series for the FireCCILT11, FireCCI51, MCD64CMQ, FireCCIM10, GFED4 and V1.3.3 products.

In addition, this analysis was also performed for each biome. The resulting graphs are included in Figure 29. When a biome to biome analysis is performed, some aspects should be highlighted with the aim of considering them in future versions of this new BA product.

- Boreal forest: FireCCIM10 provides an underestimation for most months, mainly from May to August and for the FireCCI products.
- Deserts&Xeric Shrublands: the trend of FireCCIM10 matches the other products, except October and November when BA FireCCI products underestimate BA values respect the rest of the products.
- Mediterranean: FireCCI BA products, including FireCCIM10, provide highest values than GFED4 and MCD64.
- Rock&Ice: the BA values estimated by FireCCIM10 are almost negligible compared with the values provided by FireCCILT11.
- Temperate forest: the best match is between FireCCI51 and FireCCIM10.
- Temperate savanna: the best match is between FireCCI51 and FireCCIM10.
- Tropical forest: From January to August FireCCIM10 overestimate other products, except FireCCILT11, which provides larger BA values than FireCCIM10 from May to December. From August to December FireCCIM10 presents a similar trend to the rest of the resto of the products.
- Tropical Savanna: A good agreement between all products is obtained for most of the months; from May to August this agreement is better amongst the FireCCI products.
- Tundra: FireCCIM10 provides an underestimation for most months, mainly from May to August for all products.



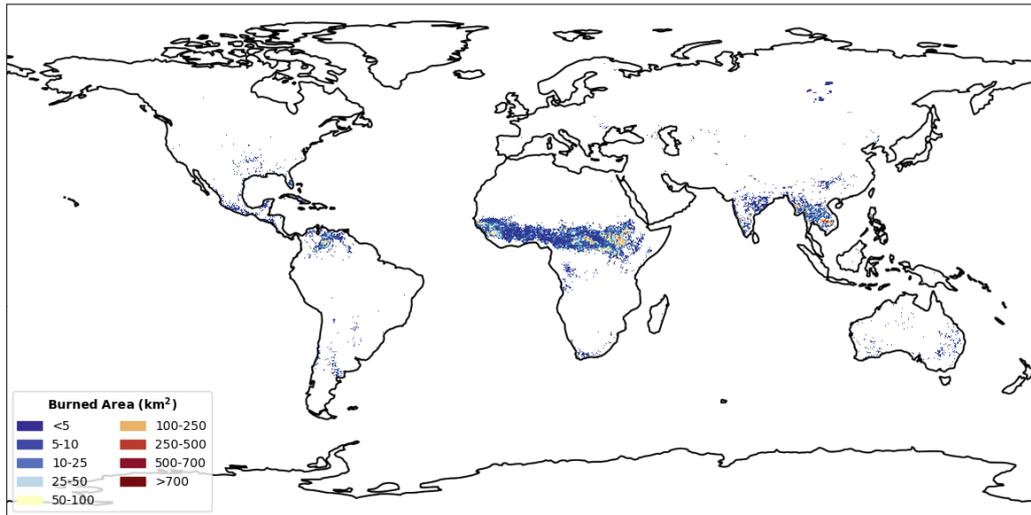


--- FireCCILT11   
--- FireCCI51   
--- MCD64   
--- FireCCIM10   
--- GFED4   
--- V1.3.3

*Figure 29: BA time series for the FireCCILT11, FireCCI51, MCD64CMQ, FireCCIM10, GFED4 and V1.3.3 products, average all-time series, for different biomes.*

Finally, as an example, the layers of BA and CL for February 2007 and December 2013 have been included in Figure 30 and Figure 31.

Total BA FireCCIM10 - 2017/02



CI FireCCIM10 - 2017/02

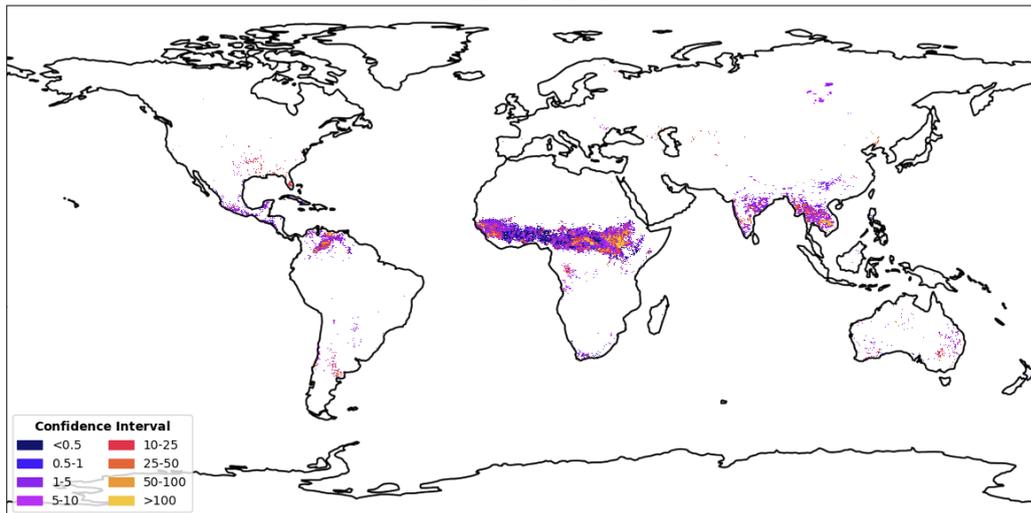
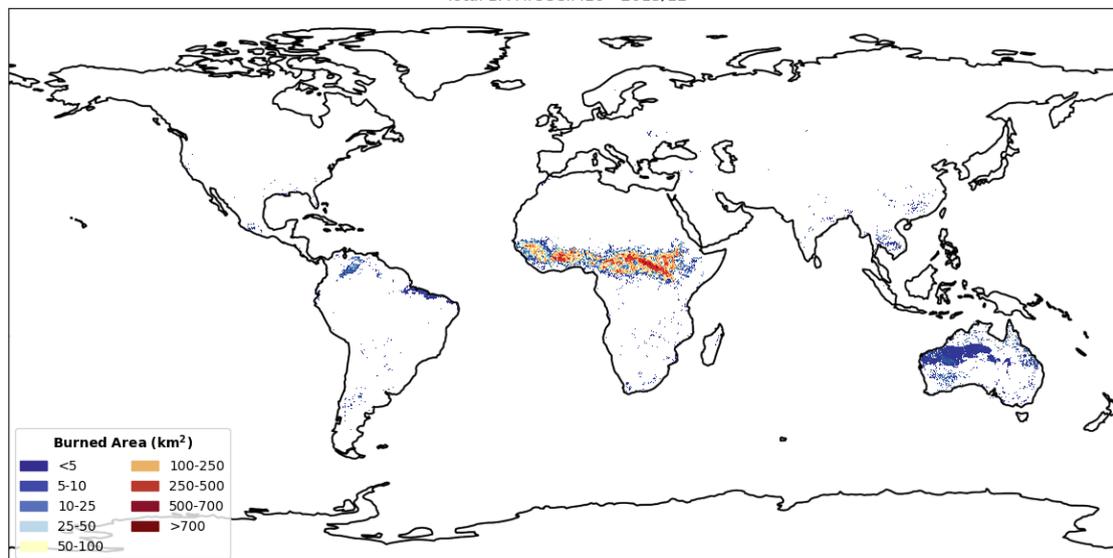


Figure 30: FireCCIM10 product for February 2017. Total burned area at 0.25 spatial resolution (top) and Confidence Interval (bottom).

Total BA FireCCIM10 - 2013/12



CI FireCCIM10 - 2013/12

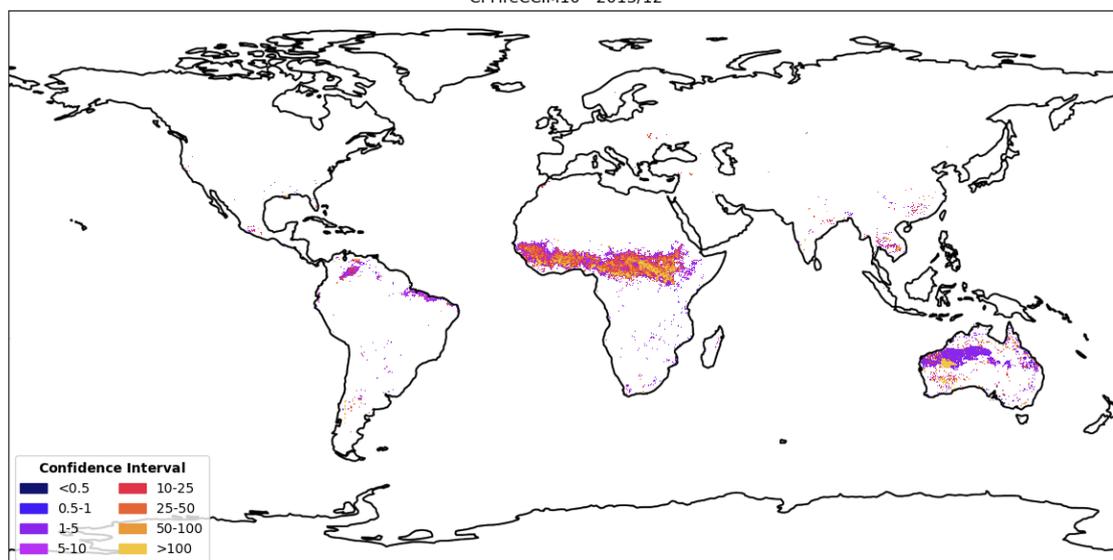


Figure 31: FireCCIM10 product for December 2013. Total burned area at 0.25 spatial resolution (top) and Confidence Interval (bottom).

### 3 Conclusions

A new time series BA product has been generated for the period 2001 to 2018, with more than 80% agreement according to BARD. This allowed generating a new BA product with more than 70% agreement according to BARD between 1982 and 2018. There is a need for more reference data to validate the results obtained, allowing one to improve the models too.

Please notice that the dataset is not publicly available, but a copy can be obtained upon request to the FireCCI team (<https://climate.esa.int/en/projects/fire/contacts>).

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page			45

## References

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>.

Chuvieco, E., Mouillot, F., van der Werf, G. R., San Miguel, J., Tanase, M., Koutsias, N., García, Yebra, M., Padilla, M., Gitas, I., Heil, A., Hawbaker, T. J., Giglio, L. (2019). Historical background and current developments for mapping burned area from satellite Earth observation, *Remote Sensing of Environment*, 225, 45-64, <https://doi.org/10.1016/j.rse.2019.02.013>

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: <https://climate.esa.int/en/projects/fire/key-documents/>

Chuvieco, E., Yue, C., Heil, A., Mouillot, F., Alonso-Canas, I., Padilla, M., Tansey, K. (2016). A new global burned area product for climate assessment of fire impacts. *Global Ecology and Biogeography*, 25(5), 619-629.

Efron, B. Jackknife-after-bootstrap standard errors and influence functions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 83–127, 1650, 1992.

Efron, B. Estimation and accuracy after model selection. *Journal of the American Statistical Association*, 109 (507), 991-1007, 10.1080/01621459.2013.823775, 2013.

Franquesa, M., Lizundia-Loiola, J., Stehman, S. V. and Chuvieco, E. (2022) Using long temporal reference units to assess the spatial accuracy of global satellite-derived burned area. *Remote Sensing of Environment*, 269, 112823. <https://doi.org/10.1016/j.rse.2021.112823>. GCOS, 2016. The Global Observing System for Climate: Implementation Needs. GCOS-200. World Meteorological Organization, Geneva, Switzerland.

Giglio, L., Randerson, J. T., van der Werf, G. R., Kasibhatla, P. S., Collatz, G. J., Morton, D. C., and DeFries, R. S.: Assessing variability and long-term trends in burned area by merging multiple satellite fire products, *Biogeosciences*, 7, 1171–1186, <https://doi.org/10.5194/bg-7-1171-2010>, 2010.

Giglio et al. (2020) Collection 6 MODIS Burned Area Product User's Guide Version 1.3. Available at: [https://lpdaac.usgs.gov/documents/875/MCD64\\_User\\_Guide\\_V6.pdf](https://lpdaac.usgs.gov/documents/875/MCD64_User_Guide_V6.pdf)

Giglio et al, The Collection 6 MODIS burned area mapping algorithm and product, *Remote Sensing of Environment*, Volume 217, 72-85, 2018.

Giglio, L., J. T. Randerson, and G. R. van der Werf. Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4) *J. Geophys. Res. Biogeosci.*, 118, 317–328, 2013.

Lizundia-Loiola, J., Franquesa, M., Khairoun, A., Chuvieco, E. (2022) Global burned area mapping from Sentinel-3 Synergy and VIIRS active fires. *Remote Sensing of*

	<b>Fire_cci</b> <b>Algorithm Theoretical Basis</b> <b>Document - MERGED</b>	Ref.:	Fire_cci_D2.2.3_ATBD-MERGED_v1.1		
		Issue	1.1	Date	18/07/2024
		Page	46		

Environment 282, 113298. <https://doi.org/10.1016/j.rse.2022.113298>

Lizundia-Loiola, J., Otón, G., Ramo, R., Chuvieco, E. (2020) A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data. *Remote Sensing of Environment* 236, 111493, <https://doi.org/10.1016/j.rse.2019.111493>

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 2.0. Available at: <https://climate.esa.int/en/projects/fire/key-documents/>

MacGregor, J., Gorman, A. J. (1994). Some considerations for using AVHRR data in climatological studies: orbital characteristics of NOAA satellites. *Int. J. Remote Sens.* 15, 537-548.

Olson, D. M., et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity, *BioScience*, Volume 51, Issue 11, November 2001, Pages 933–938, [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)

Otón, G., Chuvieco, E. (2021) ESA CCI ECV Fire Disturbance: D2.1.3 Algorithm Theoretical Basis Document (ATBD) for the AVHRR-LTDR burned area product, version 1.0. Available at: <https://climate.esa.int/en/projects/fire/key-documents/>

Otón, G., Lizundia-Loiola, J., Pettinari, M. L., Chuvieco, E. (2021) Development of a consistent global long-term burned area product (1982–2018) based on AVHRR-LTDR data, *International Journal of Applied Earth Observation and Geoinformation*, Volume 103, <https://doi.org/10.1016/j.jag.2021.102473>

Otón, G., Ramo, R., Lizundia-Loiola, J., Chuvieco, E. (2019) Global detection of long-term (1982-2017) burned area with AVHRR-LTDR data. *Remote Sensing* 11 (18), 2079, <https://doi.org/10.3390/rs11182079>

Pettinari, M. L., Lizundia-Loiola, J., Chuvieco, E. (2021). ESA CCI ECV Fire Disturbance: D4.2.1 Product User Guide - MODIS, version 1.1. Available at: <https://climate.esa.int/en/projects/fire/key-documents/>

Weber, H., Wunderle, S. (2019). Drifting effects of NOAA satellites on Long Term Active Fire Records of Europe. *Remote Sens.* 11(4), 467. <https://doi.org/10.3390/rs11040467>

Wagner, S., Hastie, T. and Efron, B., Confidence Interval for Random Forests: The Jackknife and the Infinitesimal Jackknife, *Journal of Machine Learning Research*, 15, 1625-1651, 2014.

## Annex 1: General acronyms and abbreviations

AVHRR	Advanced Very High Resolution Radiometer
BA	Burned Area
BARD	Burned Area Reference Dataset
CCI	Climate Change Initiative
CI	Confidence Interval
CL	Confidence Level
ECV	Essential Climate Variables
ESA	European Space Agency
FBA	Fraction of Burnable Area
FOA	Fraction of Observed Area
GCOS	Global Climate Observing System
GFED	Global Fire Emissions Database
IJ	Infinitesimal Jackknife
LPR	Local Polynomial Regression
LTDR	Long Term Data Record
MODIS	Moderate Resolution Imaging Spectroradiometer
NA	Not applicable
NASA	National Aeronautics and Space Administration
NC	NetCDF
RFR	Random Forest Regressor

## Annex 2: Climatic variable abbreviations

Variable name	Long name
<b>Fire Danger Indices</b>	<a href="https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-fire-historical-v1?tab=overview">https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-fire-historical-v1?tab=overview</a>
bui	Build-up Index
danger_risk	Danger rating
dc	Drought code
dmc	Duff moisture code
erc	Energy release component
ffmc	Fine fuel moisture code
dsr	Fire daily severity index
fdi	Fire danger Index
fwi	Fire weather Index
ic	Ignition component

isi	Initial spread Index
kbdi	Keetch-Bryam drought Index
sc	Spread component
<b>ERA5_Land</b>	<a href="https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means?tab=overview">https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means?tab=overview</a>
d2m	2 metre dewpoint temperature
fal	Forecast albedo
lai_vh	Leaf area index, high vegetation
lai_lv	Leaf area index, low vegetation
skt	Skin temperature
slhf	Surface latent heat flux
sp	Surface pressure
src	Skin reservoir content
sshf	Surface sensible heat flux
ssr	Surface net solar radiation
ssrd	Surface solar radiation downwards
stl1	Soil temperature level 1
stl2	Soil temperature level 2
stl3	Soil temperature level 3
stl4	Soil temperature level 4
str	Surface net thermal radiation
strd	Surface thermal radiation downwards
swvl1	Volumetric soil water layer 1
swvl2	Volumetric soil water layer 2
swvl3	Volumetric soil water layer 3
swvl4	Volumetric soil water layer 4
t2m	2 metre temperature
tp	Total precipitation
u10	10 metre U wind component
v10	10 metre V wind component