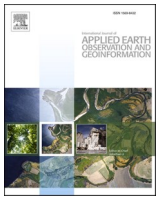
International Journal of Applied Earth Observation and Geoinformation 120 (2023) 103350

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Refining historical burned area data from satellite observations

Víctor Fernández-García [a,b,*], Christian A. Kull [a]

[a] Institute of Geography and Sustainability, Faculty of Geosciences and Environment, Université de Lausanne, Géopolis, CH-1015 Lausanne, Switzerland
[b] Ecology, Department of Biodiversity and Environmental Management, Faculty of Biological and Environmental Sciences, Universidad de León, 24071 León, Spain

ARTICLE INFO

Keywords:
Sentinel-2
MODIS
Random forest
Africa
Madagascar

ABSTRACT

Sentinel-2 imagery has revealed a substantial underestimation of burned area (BA) compared with earlier satellite products with coarser spatial resolution. In this context, we investigate the predictability of biases between the reference Sentinel-2 BA product developed for Sub-Saharan Africa (FireCCISFD) in 2019 and commonly used global coarse resolution BA products (MCD64, Fire CCI and C3S), providing tools to refine historical annual BA data before the Sentinel-2 era. To do so, we built a comprehensive dataset of environmental predictors of BA biases, with variables or proxies of (I) the annual BA estimated from the coarse-resolution product, (II) BA sizes, (III) the persistence and strength of BA signals, (IV) the maximum potential BA, and (V) the obstruction of land surface observation from satellites. Full and parsimonious random forest models were performed and validated through out-of-bag (OOB) estimations, and reconstructed BAs were validated with external data over space, and over time. The explained variance in BA biases was ≥78.58% (OOB) for all full and parsimonious models. The reconstructed BA data showed a high correspondence with the reference BA in the validation sites over space (≥91.15% var. explained) and time (≥90.37% var. explained), notably reducing biases of coarse resolution products. As an example of the model applicability, the spatial patterns of Madagascar's BA were reconstructed for 2005, 2010, 2015 and 2020, revealing a burned extent between two and four times higher than previous estimations. The proposed models are operational solutions to obtain regional and global virtually unbiased BA estimates since 2000.

1. Introduction

Burned area (BA) detection and quantification is a research topic of general interest to the scientific community because of the multidisciplinary importance of landscape fires (Andela et al., 2017). Among others, pyrogeographers are interested in a better characterization of fire patterns (Chuvieco et al. 2018; Roteta et al., 2019); fire ecologists often require BA data to properly identify fire relationships with organisms and ecosystem functioning (Fernández-García et al., 2020); geoscientists and climatologists need accurate BA data to model C emissions (Lasslop et al., 2019) and Earth energy fluxes (Liu et al., 2019); environmental epidemiologists have use BA information to analyze health problems (Cascio, 2018); and social scientists focus on exploring the fire-human relationships (Kull, 2004) and fire impacts on ecosystem services (Roces-Díaz et al., 2022).

Given the importance of global BA characterization, time series of several global remote sensing BA products have been developed. Notable among these are the MCD64 products from the Moderate Resolution Imaging Spectroradiometer (MODIS) on board the Terra and Aqua satellites launched by the U.S. National Aeronautics and Space Administration (NASA) in 1999 and 2002, respectively. The MCD64 products (e.g. Giglio et al., 2018) are obtained from the MODIS 500 m surface reflectance imagery coupled with 1 km MODIS active fire observations, taking advantage of the good revisit period of MODIS sensors (almost daily combining Terra and Aqua). MCD64 products are available since 2000 (Giglio et al., 2018) and widely used in estimating the global BA, revealing around 4.78 M km^2 $year^{-1}$ for 2003–2019 (Alonso-González and Fernández-García, 2021); carbon emissions, with estimates around 1.8 Gt C $year^{-1}$ for 2000–2019 (Zheng et al., 2021); and fire severity, revealing that severe burning concentrates in high-latitude biomes (Alonso-González and Fernández-García, 2021).

A second time series is the FireCCI product. With the aim of improving spatial resolution and consequently the detection of small fires, Chuvieco et al. (2018) this product was developed at a resolution of 250 m within the framework of the European Space Agency's (ESA) Climate Change Initiative (CCI). This product is also based on MODIS

* Corresponding author at: Institute of Geography and Sustainability, Faculty of Geosciences and Environment, Université de Lausanne, Géopolis, CH-1015 Lausanne, Switzerland.
E-mail addresses: victor.fernandezgarcia@unil.ch, vferg@unileon.es (V. Fernández-García).

https://doi.org/10.1016/j.jag.2023.103350
Received 31 January 2023; Received in revised form 20 April 2023; Accepted 8 May 2023
Available online 17 May 2023

reflectance and thermal anomaly data but includes the MODIS reflective bands at 250 m (Chuvieco et al., 2018). Since 2017, the Copernicus Climate Change Service (C3S) has generated a third global BA product (C3SBA10; Lizundia-Loiola et al., 2021) by adapting the FireCCI algorithm to the reflectance data from 300 m Sentinel-3 OLCI.

Images from the Landsat satellites, with higher spatial resolution (30 m), have also been used to systematically generate BA products at the regional level (conterminous United States; Hawbaker et al., 2020), and can be applied to build historical series of BA in some regions (e.g. Fernández-García et al., 2018a; Belhadj-Khedher et al., 2018). However, due to the long revisit period (16 days), Landsat satellites have major limitations in areas where there is a high persistence of cloud cover, such as equatorial regions and high latitudes (Ju and Roy, 2008), and in vegetation types where fire scars fade quickly, such as savannas (Melchiorre and Boschetti, 2018). Both limitations are common in sub-Saharan Africa, where 70% of the area burned worldwide is concentrated (Andela et al., 2017; Ramo et al., 2021).

In this context, the launch of the Sentinel-2A (in 2015) and 2B (in 2017) satellites has opened new possibilities for more accurate BA detection worldwide, since the combined action of both satellites provides multispectral images at 10 to 60 m resolution with a revisit period of 5 days. Sentinel-2 capabilities for BA mapping have been exploited regionally to map large fires (e.g. Llorens et al., 2021), as well as to map all large and small fires (e.g. van Dijk et al., 2021; Deshpande et al., 2022) revealing better results than reference data at higher spatial resolutions. Probably the most extent BA database based on Sentinel-2 was compiled by Roteta et al. (2019) and Chuvieco et al. (2022) for Sub-Saharan Africa at 20 m (FireCCISFD) for 2016 (FireCCISFD11) and 2019 (FireCCISFD20) respectively, showing that BA products from Sentinel-2 exceed others almost everywhere. Likewise, these authors have observed that the total BA in Sub-Saharan Africa is between 80% and 120% greater than estimated using MODIS-derived products (Chuvieco et al., 2022), and suggest this underestimation to be spatially heterogeneous, the highest errors being where the BA patches are smaller (Ramo et al., 2021) as detected for other regions (Deshpande et al., 2022). Then, a detailed assessment of the sources of error impacting global BA products by Franquesa et al. (2022a) also found as determining factors the spatial resolution of the BA product, the active fire densities, size and shape of burned patches and land cover types.

The error metrics of all these products have been calculated using independent reference BA data, usually from 20 to 30 m resolution imagery (i.e., Landsat or Sentinel imagery), visually interpreted or compared by pairs, which is a highly costly and time-consuming process (Franquesa et al., 2022a, 2022b). The first stage-3 validations of MCD64A1 product reported 40.2% commission error (CE) and 72.6% omission error (OE) with a relative bias of −54% (Boschetti et al., 2019), and the first error estimates of the FireCCI product revealed 54% CE and 67% OE with underestimations around −28% (Padilla et al., 2018). However, Franquesa et al. (2022b) used long-time series of reference data to avoid the interference of temporal resolution on error metrics, resulting in biases decreased to −38.7% for MCD64 and −29.4% for FireCCI and −38.7% for C3S, still far from the 15% error threshold recommended by the Global Climate Observing System (GCOS) (GCOS, 2016). In contrast, the validation of the Sentinel-2A product (FireCCISFD20) following the last approach resulted in 8.5% OE, 15.0% CE and relative bias of 8.4% (Chuvieco et al., 2022), meeting the GCOS guidelines. In view of its high accuracy, fine resolution and independence acquisition, the FireCCISFD20 product meets the criteria to be considered as reference when compared to coarse resolution products, and has the advantage over other existing reference BA data of being available for a larger spatial extent (Sub-Saharan Africa) for two complete years. Thus, one can use the FireCCISFD20 product as a reference to identify biases of coarse resolution BA products and for the development of robust statistical models able to refine historical fire data.

Based on these premises, we investigate the predictability of biases between annual BA from the Sentinel-2 2019 BA product developed for Sub-Saharan Africa (FireCCISFD20) and commonly used global BA products (MCD64, Fire CCI and C3S), and thus provide tools to refine historical annual BA data. Specifically, we (I) build comprehensive models to identify the predictability of biases between the annual BA of Sentinel-2 and coarser-resolution products, and to analyze the relationship of each predictor with the response variable (II) develop efficient predictive models (parsimonious) to compensate biases in annual BA of coarse-resolution products, and (III) show the applicability of the developed parsimonious model to reconstruct historical BA patterns in a target study site.

2. Methods

2.1. Study area

We have selected Sub-Saharan Africa because the majority of the Earth's BA concentrates there (Andela et al., 2016; Ramo et al., 2021), and because of the availability of accurate regional burned area products (FireCCISFD). The analysis was restricted to the area shown in Fig. 1, South of the 20° N parallel.

2.2. Data

The response variables in this work were the biases between the Sentinel-2 BA product developed for Sub-Saharan Africa in 2019 (FireCCISFD20) and three global BA products (MCD64, Fire CCI and C3S). The four mentioned products were acquired in a monthly temporality for 2019. The FireCCISFD20 was acquired from the ESA CCI Open data portal (https://climate.esa.int/en/odp/) at 0.05° and then aggregated at 0.25°, the MCD64 collection 6 BA data was downloaded from Fire Information for Resource Management System US/Canada (https://firms.modaps.eosdis.nasa.gov/usfs/download/) at 0.25° (MCD64CMQ) and converted to km2, and the FireCCI version 5.1.1 cds and C3S version 1.1 products were downloaded from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/) at 0.25°. The total BA in km^2 for each grid was obtained by calculating the sum of all months, and biases were calculated in km^2 for each grid (Fig. 2) by subtracting the BA detected by FireCCISFD20 from the BA detected in the global coarse-

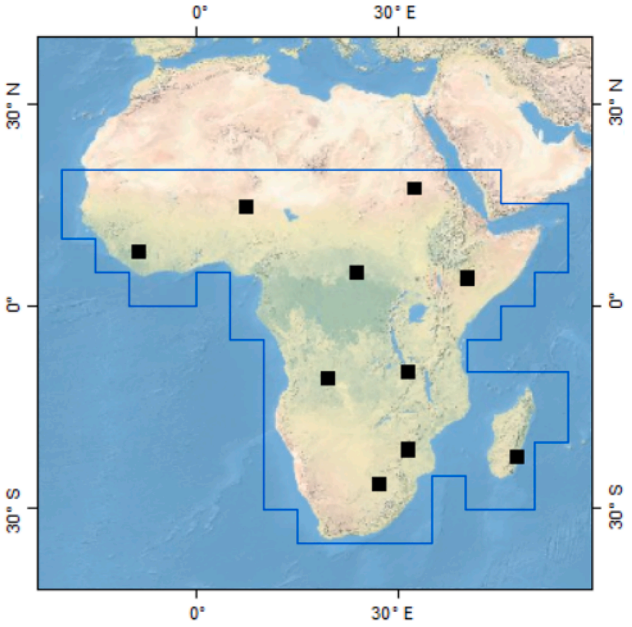


Fig. 1. Map showing the study area (blue frame) and the location of the 2° × 2° validation regions (black squares). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

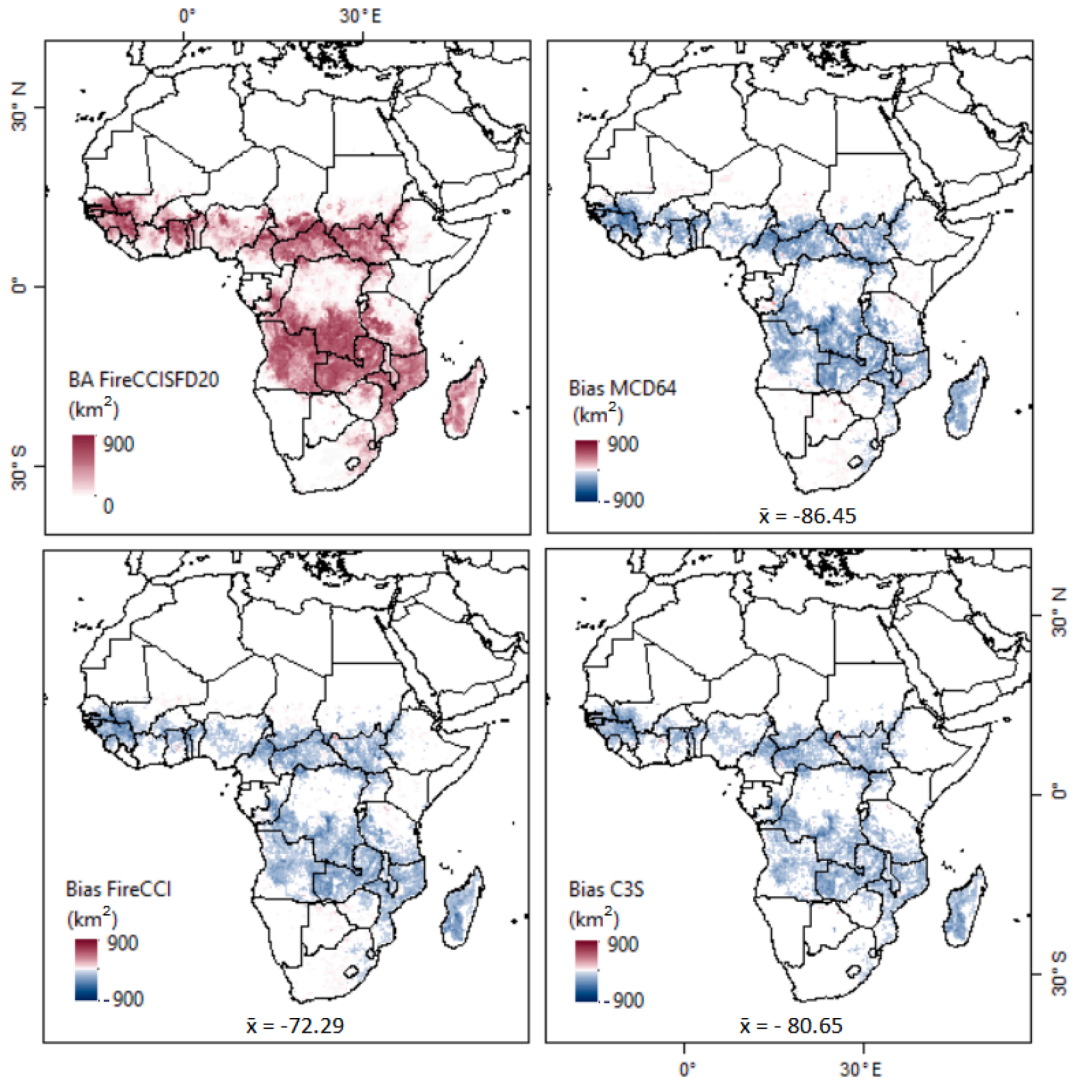


Fig. 2. Spatial patterns of BA in the study region according to the reference BA product (FireCCISFD20) and the response variables (biases of the MCD64, FireCCI and C3S in relation to the FireCCISFD20) at 0.25°.

Table 1
Variables used as predictors of BA biases. The table also shows the data sources, their temporal availability and the units used in this study.

Group	Predictor	Data source	Source resolution	Temporal availability of source	Used units
Coarse BA data	BA MCD64	MCD64CMQ C6	0.25°	2000- (monthly)	km^2
	BA FireCCI	ESA-CCI 5.1.1 cds	0.25°	2001–2019 (monthly)	km^2
	BA C3S	C3S 1.1	0.25°	2017- (monthly)	km^2
BA size	Wind speed	ERA5-10 m u wind	0.25°	1959- (hourly)	ha
	Mean patch area	GLC_FCS30	30 m	1985- (quinquennially)	km^2
	Contagion	GLC_FCS30	30 m	1985- (quinquennially)	%
	Shannon index	GLC_FCS30	30 m	1985- (quinquennially)	Unitless
	CV elevation	SRTM	30″	Unique	%
	Population density	NASA-SEDAC	30″	2000- (quinquennially)	n km^{-2}
	Road density	GRIP4	5′	Unique	m km^{-2}
Persistence	NPP	M*D17A3HGF	500 m	2000- (yearly)	kg C m^{-2} $year^{-1}$
	NDVI	M*D13C2	0.05°	2000- (monthly)	Unitless
	Min. NDVI	M*D13C2	0.05°	2000- (monthly)	Unitless
	Soil water	ERA5-Soil w. lyr1	0.25°	1959- (hourly)	m^3 m^{-3}
	Min. soil water	ERA5–Soil w. lyr1	0.25°	1959- (hourly)-	m^3 m^{-3}
	Annual P	ERA5-Total P	0.25°	1959- (hourly)	mm $year^{-1}$
	Min. P	ERA5-Total P	0.25°	1959- (hourly)	mm $month^{-1}$
Persistence & size	Fraction of forest	GLC_FCS30	30 m	1985- (quinquennially)	%
	Fraction of cropland	GLC_FCS30	30 m	1985- (quinquennially)	%
	Fraction of savanna	GLC_FCS30	30 m	1985- (quinquennially)	%
Maximum BA	Burnable area	GLC_FCS30	30 m	1985- (quinquennially)	km^2
Obstruction	Cloudiness	ERA5-Total cloud	0.25°	1959- (hourly)	%

resolution product for the same year. Likewise, mean biases (mean errors) were computed.

The predictor dataset aimed at explaining the foregoing biases was composed of variables or determinants of the BA estimated from the coarse-resolution product, the BA patch size, the persistence and strength of BA signals, the maximum potential BA, and the obstruction of land surface observation from remote sensors (Table 1). Apart from the BA estimated from the corresponding coarse-resolution product to each predictor, the other variables were obtained from sources available at least for the entire MODIS era (2000-present), thus allowing their inclusion in predictive models to refine the three targeted global BA products.

Assuming that BA patch sizes and shapes are key drivers of errors in BA products, with presumably increased biases in global BA products in those areas where BA patches are smaller (Ramo et al., 2021; Franquesa et al., 2022a; Deshpande et al., 2022), we have computed for 2019 (I) the mean annual values of the 10 m u-component of wind (wind speed) from the monthly data of the ERA5 reanalysis downloaded from the Copernicus Climate Data Store, as high wind speeds might favor fire propagation and thus BA sizes. In addition, we used the 2015 GLC_FCS30 product at 30 m spatial resolution (Zhang et al., 2021) downloaded from the Chinese Academy of Sciences CASEarth data portal (https://data.casearth.cn) to calculate (II) the mean patch area; (III) the contagion, an aggregation and adjacency metric that accounts not only for patch sizes but also for patch shapes and distribution, with values ranging from 0 (uneven distribution of classes) to 100 (maximum clumping and minimum landscape fragmentation); and (IV) the Shannon index of landscape classes, that accounts for the richness and evenness of classes within the analytical unit, increasing as the system entropy and thus landscape heterogeneity increases. The calculation of mean patch area, contagion, and Shannon index was performed using the "sample_lsm" function of the landscapemetrics package (Hesselbarth et al., 2019) after merging all the unburnable land cover classes (impervious, bare areas, water bodies, ice and snow) into a unique category (unburnable), and after reprojecting the data to UTM as recommended by Hesselbarth et al. (2019). As a topographic indicator of landscape heterogeneity potentially affecting BA patch sizes we computed (V) the coefficient of variation (CV) of elevation from the NASA Shuttle Radar Topographic Mission (SRTM) 3 arc-second Digital Elevation Database v4.1 downloaded from CGIAR-CSI (https://cgiarcsi.community/). Likewise, we have considered as limiters of BA patch sizes (VI) the population density acquired at 30 arc-second from the Socioeconomic Data and Applications Center (SEDAC) (https://sedac.ciesin.columbia.edu/) and (VII) road density acquired at 15 arc-min from The Global Roads Inventory Project (Meijer et al., 2018).

Upon the premise that the persistence and strength of BA signals might influence biases among BA products (Rodrigues et al., 2019) because of differences in temporal resolution, spectral resolution and detection algorithms, we included in the predictor dataset (I) the annual Net Primary Productivity (NPP), calculated from composites of the MODIS MOD17A3HGF and MYD17A3HGF Terra and Aqua products, which were downloaded from the NASA's Land Processes Distributed Active Archive Center (LP DAAC) (https://e4ftl01.cr.usgs.gov/); (II) the mean annual Normalized difference vegetation index (NDVI) from monthly composites of the MODIS MOD13C2 and MYD13C2 Terra and Aqua products from LP DAAC, which is an index of vegetation vigor and living fuel amount; and (III) the minimum monthly NDVI from the foregoing composites. From the monthly data of the ERA5 reanalysis we computed (IV) the mean annual volumetric soil water in the uppermost 7 cm of soil; (V) the minimum monthly volumetric soil water in the uppermost 7 cm of soil; (VI) the annual precipitation; and (VII) the minimum monthly precipitation.

As proxies of the persistence and strength of BA signals (Rodrigues et al., 2019), but also related to the BA patch sizes (Zhu et al., 2017), we calculated from the 2015 GLC_FCS30 product (I) the fraction of forest by combining all broadleaved, needleleaved and mixed-leaved forest types; (II) the fraction of cropland by including all rain-fed and irrigated cropland types; (III) and the fraction of savanna, as defined by Chuvieco et al. (2021) by including shrublands and grasslands.

As a geographical constrain to the maximum burnable area and allowable BA bias, we have included in the predictor dataset the total burnable area of each cell in km^2. This variable was calculated by multiplying the total area of each 0.25° cell, which varies with latitude, by the fraction of burnable area obtained from the 2015 GLC_FCS30 product. The fraction of burnable area was calculated as the sum of the fractions of all land cover classes except impervious, bare areas, water bodies, ice and snow.

Finally, we have included in our predictor dataset the cloudiness calculated as the mean annual cloud cover from the ERA5, as proxy of the obstruction of land observations from remote sensing, which might introduce spatial variability in the detection capacity of BA and BA biases among sensors.

When possible, the predictors were acquired at 0.25° spatial resolution (ERA5), and those variables acquired and processed at finer resolutions were resampled to a grid of 0.25° degrees, using the "resample" function of the terra package (Hijmans, 2022) and the average method to obtain mean values, eventually weighted by the pixel fractions within each 0.25° cell when grids were not originally aligned.

2.3. Development of predictive models

Data of the three response variables (BA bias of MCD64, BA bias of Fire CCI and BA bias of C3S with respect the FireCCISFD20), and of all environmental predictors was systematically extracted using one sampling point per 0.25° degrees cell (n = 31076) and imported to the R environment (R Core Team, 2021), where all data analyses were done. Then, the dataset was divided into a training sub-dataset (n = 30453) to build the predictive models and into a validation sub-dataset (n = 623), keeping the data from 10 regions of 2° × 2° for the spatial validation (section 2.4).

Full random forest (RF) models (including all predictors) were built for each response variable, in order to identify the maximum predictability of BA biases. Full RF models were also used to analyze the responses of BA biases to each predictor while holding the rest predictors constant to their median value using the "plotmo" function from the plotmo package (Milborrow, 2022), and these partial-dependence curves were compared to the corresponding univariant local polynomial fitting (loess) curves (where the effect of the other predictors is not neutralized).

For increased efficiency and practical applicability, we also used the training dataset to develop parsimonious RF models, that reach a compromise between the predictive capacity and the number of predictors. We followed a two-step procedure to identify the most suitable predictors (Chavent et al., 2021). First, a hierarchical cluster analysis was performed as a method to eliminate collinear variables (Dormann et al., 2012). The hierarchical cluster analysis and consequent variables selection was performed according to the following steps: (I) a correlation matrix with the absolute values of the Spearman's correlation coefficient was computed, (II) the correlation matrix was converted to a distance matrix by calculating the inverse of the values from the correlation matrix (III) the "hclust" function with the complete agglomeration method was applied to the distance matrix. (IV) Based on the hierarchical cluster analysis, groups of dissimilar variables at a level of 0.3 (i.e., 0.7 of similarity) were identified and clustered in different groups, and (V) from each group the most correlated variable with the response variable was selected according to the absolute value of the Spearman's correlation coefficient. Secondly, the variables that had not a relevant contribution to the predictive capacity of the random forest model, we removed by applying a recursive feature elimination with the "rfe" function of the caret package (Kuhn, 2021), using 10-fold cross-validation and 5 repeats to improve the performance of feature selection, partitioning the dataset into 80% (training) and 20% of data

(validation).

Additionally, we performed RF models only with the predictor showing the strongest relationship with the BA bias according to the absolute values of the Spearman's correlation coefficient to show the magnitude of the decay in predictive capacity when removing the other predictors.

All RF models were implemented with the "randomForest" function of the randomForest package (Liaw and Wiener, 2002). The models were similarly tuned to facilitate inter-comparability, the ntree parameter (number of trees) was fixed to 250, which ensures the quality and stability of results according to mean out-of-bag (OOB) error curves, and is both a large and a computationally feasible value for our datasets (Probst and Boulesteix, 2018); and the mtry parameter (number of predictors sampled for splitting at each node) was fixed to one third of the number of predictors, in accordance to the exploratory results of 10-fold cross validations performed with the functions "trainControl" and "train" from the caret package. The predictive capacity within the training dataset of all RF models was assessed with the proportion of OOB variance explained and with the root mean square error (RMSE). Both OOB variance and RMSE were calculated using the OOB samples with the 63.2% of data for training and, the leftover (36.8%) for validation.

The global importance of each predictor was assessed in the parsimonious RF by the increase in node purity square error (IncNodePurity) calculated with the "importance" function. The local importance for each predictor was also calculated in the parsimonious RF as the average increase in squared OOB residuals when the variable is removed, which allowed to identify the most important predictor for each 0.25° cell across Sub-Saharan Africa.

2.4. Spatial and temporal validation of corrected BA, and practical demonstration

The applicability of the most efficient RF models to obtain refined BA estimates (BA of the coarse product minus predicted biases) was validated in independent study areas (spatial validation) and in a different year (temporal validation). The spatial validation of the most efficient RF models was done by applying the models to the validation data from 2019 (i.e., to the 10 2° grids shown in Fig. 1).

The temporal validation of the most efficient RF models was done by applying the models to the year 2016. For that, we developed a new database by including the response variables and predictors for Sub-Saharan Africa for the year 2016. This database was built using the FireCCISFD11 product created using Sentinel-2A imagery as reference BA data for that year (Roteta et al., 2019). The biases between the reference and the global BA products MCD64 and FireCCI were computed as described in the section 2.3 (the C3S product is not available for 2016). Likewise, all predictors were obtained for the year 2016 as described in the section 2.3, using the MODIS and ERA-5 data from the target year (i.e., 2016), the closest NASA-SEDAC population data (i.e. year 2015), and the immediately preceding GLC_FCS30 data (i.e., year 2015).

The predictive capacity of the most efficient RF models in the spatial and temporal validations was assessed with the adjusted R^2 of the regression performed between the observed and predicted values, with the RMSE, and with the bias (mean error) between the observed and predicted values.

Moreover, to show the applicability of the developed models over time, we computed the corrected BA values (BA of the coarse product minus predicted biases) for Madagascar for the years 2005, 2010, 2015 and 2020. Madagascar was selected as study case as it concentrates a large variety of land cover classes, climatic and topographic conditions, human presence and fire regimes in a relatively small area (587.041 km^2), and the focus of an active ongoing debate about the potential role of fire in landscape degradation (Kull 2004; Phelps et al., 2022). There, we used the total BA from the MCD64 product for the mentioned years, as it is the only analysed global BA product available for that entire period. Likewise, the corresponding predictors of BA biases were obtained as described in the section 2.3, using the MODIS and ERA-5 data of the target years (i.e., 2005, 2010, 2015 and 2020), the closest NASA-SEDAC population data (i.e., years 2005, 2010, 2015 and 2020), and the immediately preceding GLC_FCS30 data (i.e., years 2000, 2005, 2010 and 2015).

3. Results

3.1. Using all environmental predictors

The full RF models explained a proportion of OOB variance directly related to the original error of each product, thus, the lowest OOB explained variance (82.89%) was obtained when predicting the biases in the FireCCI annual BA product (Table 2), which has a mean bias of −72.29 km^2 per 0.25° cell (Fig. 1), whereas the highest OOB variance (86.19%) was explained on the biases in the MCD64 annual BA product, that was the most different product to the reference BA with a mean bias of −86.45 km^2. The RMSE ranged from 42.79 to 46.55 km^2 when predicting biases in the MCD64 product.

The responses of BA biases of to each predictor showed similar responses for the three studied global BA products (Fig. 3). The BA from the global product was the predictor that caused more variation in BA biases, the largest BA underestimations being detected in the frame of 50–300 km^2 (i.e., 5 to 40%) of annual BA per cell (Fig. 3). For the rest of predictors, large differences were detected between the loess fitting and the fitting fixing the other covariates to the median, consequence of the close relation of most variables with BA (Fig. S1). Thus, loess fittings showed BA biases (Fig. 3) and BA (Fig. S1) peaked at high values of burnable area, and Shannon index; at intermediate values of cloudiness, NPP, NDVI, minimum NDVI, soil water, minimum soil water, precipitation wind speed and contagion and CV of elevation; and at low values of minimum precipitation, patch area, road density. However, focusing on the responses to each predictor while holding the rest constant, we found that high levels of cloudiness, NDVI, soil water, minimum soil water, annual precipitation and forest cover clearly contributed to increased BA biases. On the contrary, increases in wind speed, and more clearly mean patch area (in the 0–1 km range) and contagion contributed to better BA estimations.

3.2. Using an optimal subset of predictors

The hierarchical cluster analysis (Fig. 4) performed based on the Spearman's correlation matrix (Fig. S2) grouped the variables in 11 clusters. The clusters including multiple variables were those related to the landscape heterogeneity in terms of land cover (Shannon index, contagion and mean patch area), the annual BA values from coarse resolution imagery (BA from MCD64, FireCCI and C3S), weather-related metrics (annual P, cloudiness, wind speed, minimum soil water and soil

Table 2
Predictive capacity of the random forest models (full models, parsimonious models, and models including as the only predictor the BA from the corresponding global BA product) calculated with the training dataset. OOB: Out-of-bag explained variance, RMSE: Root Mean Squared Error.

Model	Response variable	OOB	RMSE
Full	Bias MCD64	86.19	46.55
	Bias FireCCI	82.89	43.08
	Bias C3S	85.36	42.79
Parsimonious	Bias MCD64	82.30	52.71
	Bias FireCCI	82.42	46.89
	Bias C3S	78.58	48.10
BA	Bias MCD64	43.82	93.71
	Bias FireCCI	41.18	79.71
	Bias C3S	46.46	81.67

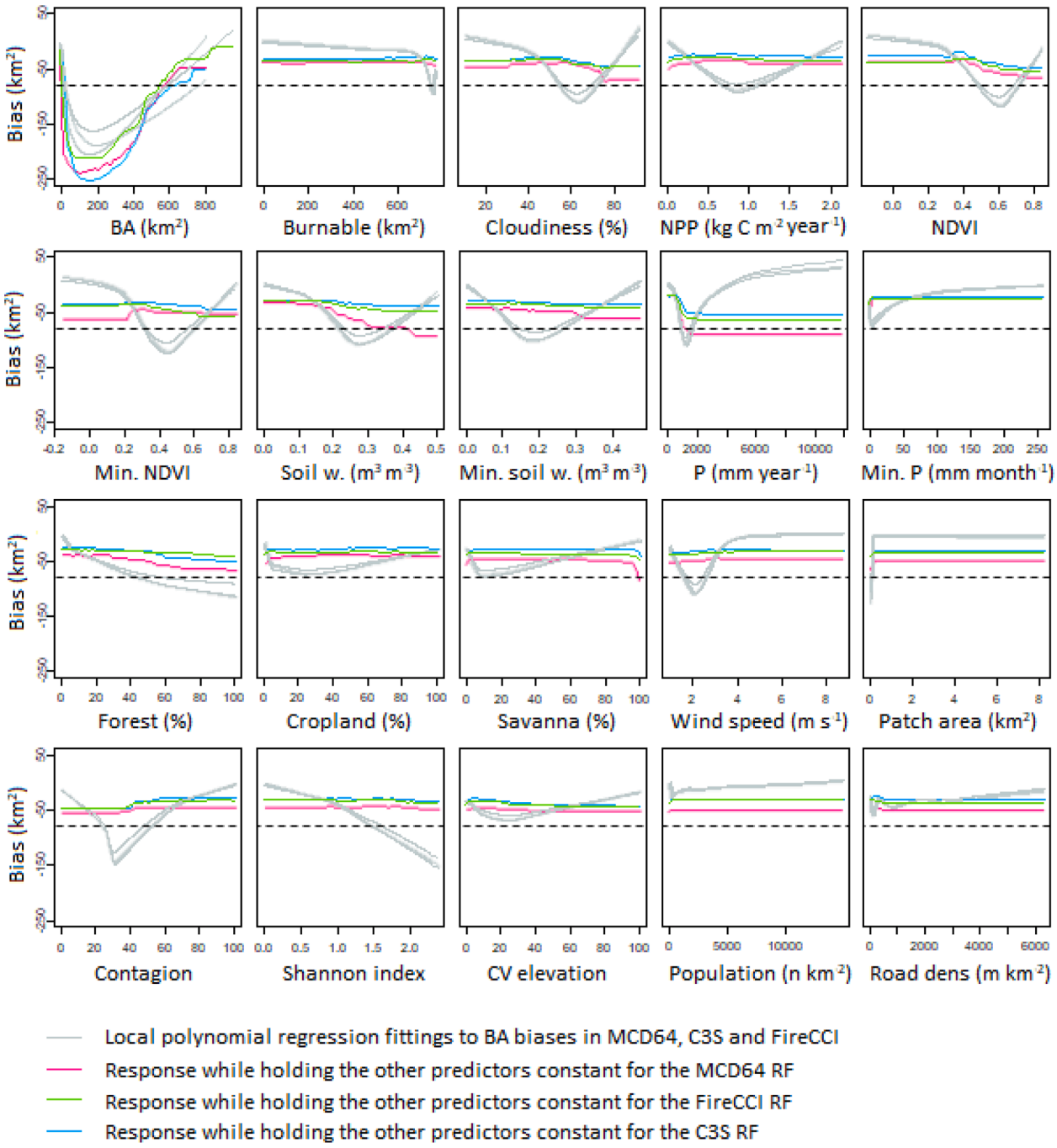


Fig. 3. Responses of BA biases between the three studied global BA products (MCD64, FireCCI and C3S) and the reference BA (FireCCISFD20) to each predictor. Relationships are shown as univariant local polynomial regression fittings and as partial-dependence plots of the full RF model while holding the rest predictors constant to their median.

water), and a cluster related to vegetation greenness (minimum NDVI, NDVI, NPP and fraction of forest). The selection of the most related variable to BA biases from each cluster, and the application of the recursive feature elimination resulted in the optimal set of predictors shown in Fig. 5.

The parsimonious models run with the optimal set of predictors showed similar predictive capacities to those of the full RF models, with explained OOB variances $\geq$ 78.58% and RMSE $\leq$ 52.71 (Table 2). The comparison of the full and parsimonious models with models including only the BA from the corresponding product as predictor revealed that the elimination of most predictors led to large decays in predictive capacity (Table 2).

The removal of collinearity among predictors in the parsimonious models allowed us to honestly rank predictors according to their global variable importance and their local importance. Thus, we found that globally, the annual BA of the corresponding coarse-resolution product was the most important predictor of BA biases for all, the MCD64, FireCCI and C3S products (Fig. 5). The second and third most important predictors for the three models were those coming from the clusters related to vegetation greenness (NDVI) or weather conditions (soil water or annual precipitation). Burnable area was the fourth most important predictor in the three models. The mean patch area, fraction of cropland, population density and minimum monthly precipitation were also important in the three models. The CV of elevation was retained in the

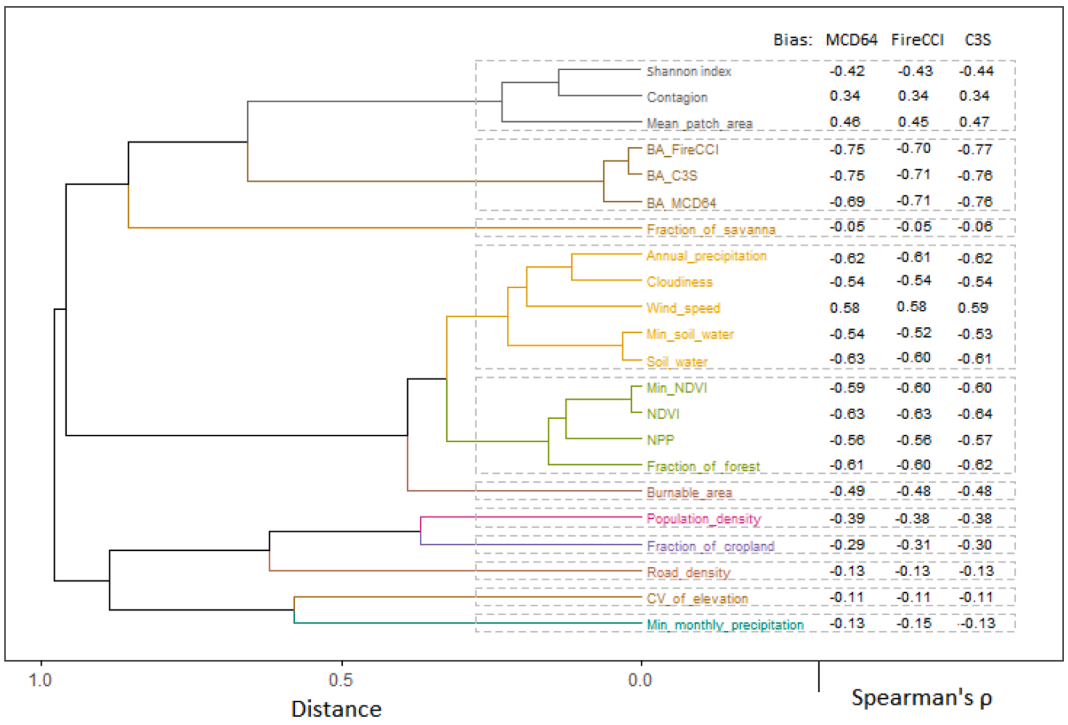


Fig. 4. Hierarchical cluster analysis of the predictor dataset based on distances computed from the absolute values of the Spearman's correlation coefficient (see correlation matrix in Fig. S2), and Spearman's correlation coefficient of each predictor with the biases of the three studied products (MCD64, FireCCI, C3S). The clusters defined at a distance of 0.3 are limited by dashed rectangles.

MCD64 and C3S models, and the fraction of savanna in the FireCCI and C3S.

Focusing on local importance, we found that BA was the most important predictor in ~ 50% of Sub-Saharan Africa. NDVI and precipitation were the most important in many savanna and xeric regions further from the equator at accounting for ~ 20% of Sub-Saharan Africa, and the minimum monthly precipitation was the most important predictor in many moist tropical areas closer to the equator, mainly for the MCD64 product (14% of Sub-Saharan Africa). Other predictors were the most important in particular regions, for instance the fraction of cropland was important where this activity concentrates such as in the band of land between 9 and 14° N.

3.3. Spatial and temporal validation of the corrected BA products

The spatial validation of the BA products with the errors compensated by the parsimonious RF models (Fig. 6) showed large decreases in original biases and RMSE, and increases in R^2. Specifically, the BA from the MCD64 product, which originally underestimated an average of 95.18 km^2 of BA per 0.25° cell in the validation regions, decreased that error to an overestimation of 2.52 km^2 after correction. Likewise, the RMSE decreased from 108.93 km^2 to 61.95 km^2, and the variance explained increased from 72.58% to 91.15%. The biases from the FireCCI and C3S products were slightly less compensated, but they exhibited lower RMSE and higher explained variance than the corrected MCD64 product.

The temporal validation of the BA estimates compensated with the parsimonious RF models for the year 2016 (Fig. 7) showed also large decreases in their original biases. The amelioration was larger for the MCD64 product, which had more margin to correct than the FireCCI. The biases of MCD64 were reduced from −70.40 to 17.37, and those from the FireCCI from −58.14 to 14.86. Likewise, RMSE were reduced more than 25 km^2 per cell in both products, and the variance explained increased at the order of 10–20%.

3.4. Model applicability: the case of Madagasca

The application of the developed parsimonious RF model to the MCD64 BA product from 2005, 2010, 2015 and 2020 in Madagascar is shown in Fig. 8. The figure shows the potential of the developed RF model to spatially characterize BA with corrected biases across different temporal points within the MODIS era, as well as the spatial heterogeneity of BA corrections. Moreover, this study case revealed that 21–25% of the island burned annually in the studied years, being between two and four times higher than reported by the MCD64 product, the only available for all years of the analysed period. The corrected estimate, however, corresponds with estimates based on qualitative observational and archival data suggested an annual burnt area of 15–30% (Kull, 2004). The comparison of the corrected BA spatial patterns and total BA with the reference BA data from the year 2016 (FIRECCISDF11) is shown in Fig. S3. The results shown in Fig.S3 confirm the performance of our model and corroborate that the BA in this island can be more than three times higher than reported by the MCD64, which is in line with the results obtained for our corrected BA in 2015 and 2020 (Fig. 8).

4. Discusion

Although nowadays the use of Sentinel-2 combined with thermal anomalies sensors (VIIRS-375 m or Sentinel-3) has exceptional BA detection capacities (Chuvieco et al., 2022; Franquesa et al., 2022a), the obtention of accurate BA estimates before the Sentinel-2 era remains a challenge in many regions where Landsat temporal resolutions or cloud-free images are not sufficient (Melchiorre and Boschetti, 2018; Ju and Roy, 2008). There, the correction of already observed annual BA, particularly from those sensors covering a relatively long time (e.g. MODIS) is a good alternative to get a more realistic picture of BA over the past few decades. In this context, the methodology we propose provides solutions for those challenges and enables the possibility of developing series of refined global annual BA products since 2000.

The analysis of the responses of annual BA biases from coarse

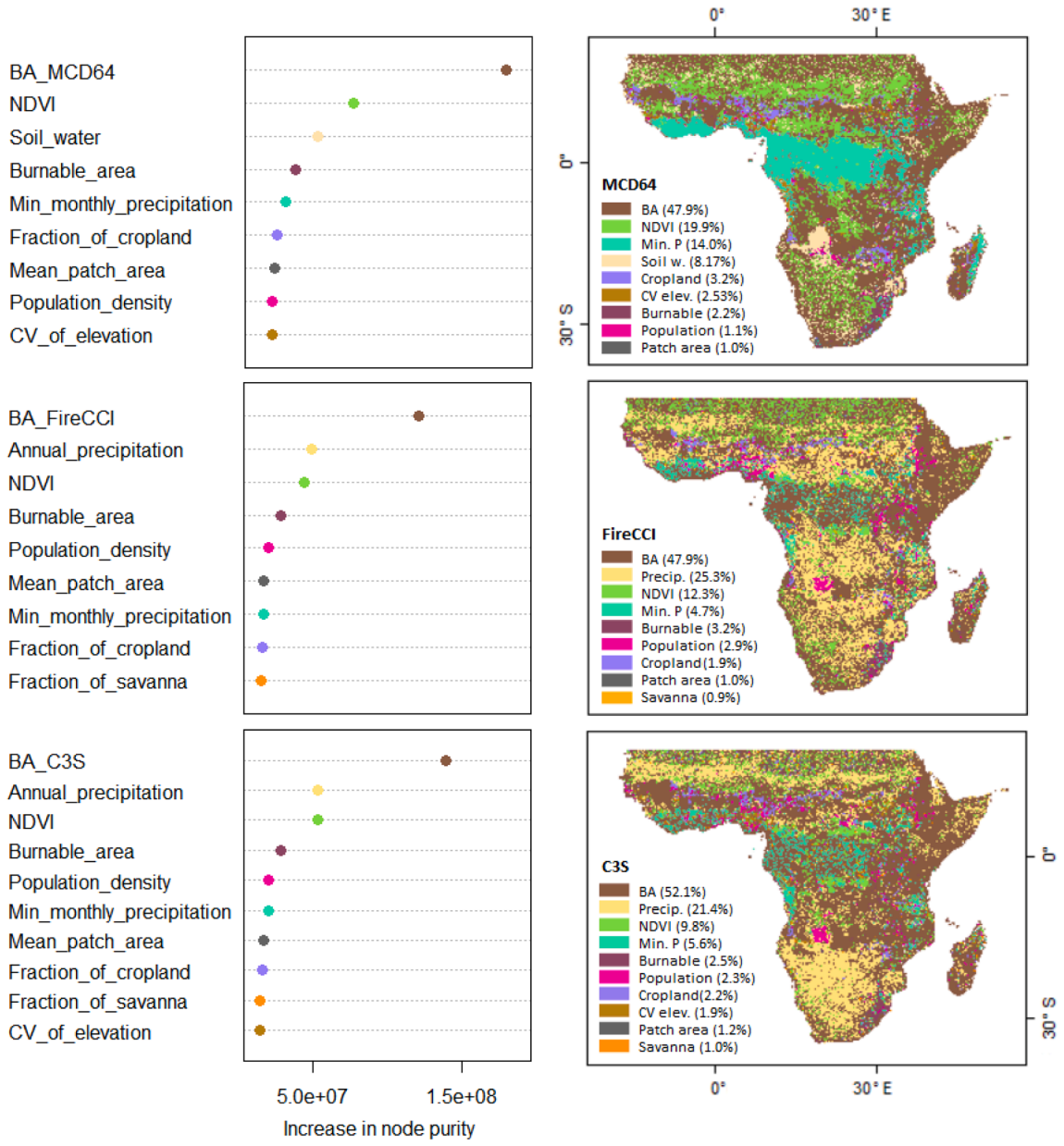


Fig. 5. Global relative importance of the predictors included in the parsimonious RF model (panels on the left), and spatial distribution of the locally most important variables according to the local importance analysis (maps on the right). The percentage values in the maps indicate the extension in the study area where each predictor was the most important.

resolution products (MCD64, FireCCI and C3S) to different predictors, along with our importance analysis, revealed similar behavior for the three products. The annual BA was the most important predictor and governed the bivariate relationships of the others with BA biases. The largest BA underestimations in absolute terms were found in those regions with 5 to 40% of burned land, largely decreasing towards the extremes. Assuming less unburned and BA to mix in the extremes (i.e., regional dominance of one class), our results partially agreed with Franquesa et al. (2022a), which found that 70% of errors from the FireCCI product (only CE were separately analyzed) occur in those pixels containing both burned and unburned land.

After the annual BA, those variables related with vegetation (i.e. from the correlated cluster of NPP, NDVI, minimum NDVI and forest cover) and weather (i.e. from the correlated cluster of annual precipitation, minimum monthly precipitation, cloudiness, soil water and minimum soil water) were the most important. The BA peaked at intermediate values of productivity, greenness and humidity supporting the intermediate fire–productivity hypothesis (Bowman et al., 2009; Pausas and Ribeiro, 2013) and the intermediate rainfall hypothesis (Archibald et al., 2018), both conceived on the assumption that fire is fuel-limited in areas were productivity and moisture are low, and flammability-limited in those where there is too moist. BA biases also peaked at intermediate values of most productivity, greenness and humidity indicators, but fixing the BA and the other predictors to the median we found that high values of NDVI, forest cover, cloudiness, soil water, minimum soil water and annual precipitation clearly contributed to increase BA biases. This can be consequence of some of the following reasons or a combination of them: (I) the high regional greenness might weaken BA spectral signals when only a pixel fraction is burned, which is important in the BA detection algorithms as they are based on local thresholds from vegetation indices (Chuvieco et al., 2018; Giglio et al., 2018); (II) fire intensity and severity in productive and moist areas is usually low, particularly in tropical forests (Fernández-García and Alonso-González, preprint) where surface fire regimes are predominant

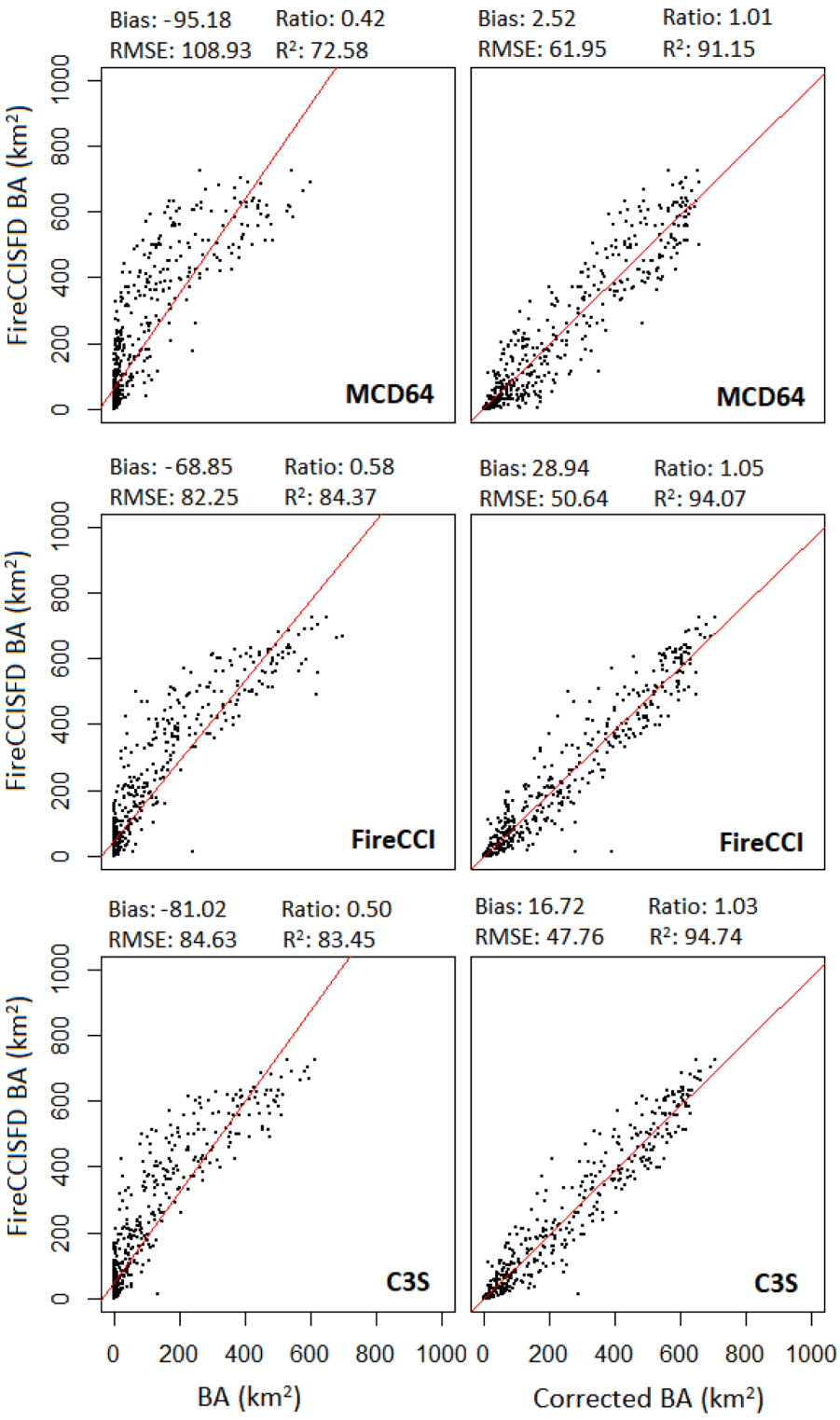


Fig. 6. Scatterplots showing the relationship between the BA from the three global BA products (MCD64, FireCCI and C3S) and the reference BA product of 2019 (FireCCISFD20) in the spatial validation plots. The panels on the left represent the original BA from the global products, and the panels on the right show the BA with errors compensated by the parsimonious RF models. R^2 values are expressed in percentage. RMSE: Root Mean Squared Error. Ratio: fraction between the total BA in the coarse resolution products (original or corrected) and the total BA in the reference product in the validation dataset.

(Bowman et al., 2009). This is intrinsically linked to a weaker spectral change and a greater interference of green canopies (Fernández-García et al., 2018b), contributing to less accurate BA estimates especially at coarse resolution; (III) the rapid fading of BA scars by fast vegetation regeneration and removal of ashes typical in productive and moist areas deteriorate the ability to detect BA (Melchiorre and Boschetti, 2018), which can be aggravated for algorithms seeking for persistent changes, such as the one used in MCD64 (Giglio et al., 2018). This could be linked to the highest importance of greenness and weather variables in predicting MCD64 BA biases, as well as to the largest importance of minimum precipitation in the tropical rainforest regions as compared to the other studied products.

Landscape continuity metrics were positively related to BA, which has been attributed to higher number of anthropogenic ignitions and increased edge flammability (Driscoll et al., 2021). On the contrary, landscape continuity, as bottom-up catalyst of BA patch size (Fernandes et al., 2016), was in general inversely related to BA biases from coarse-resolution products. This supported our initial hypothesis, indirectly

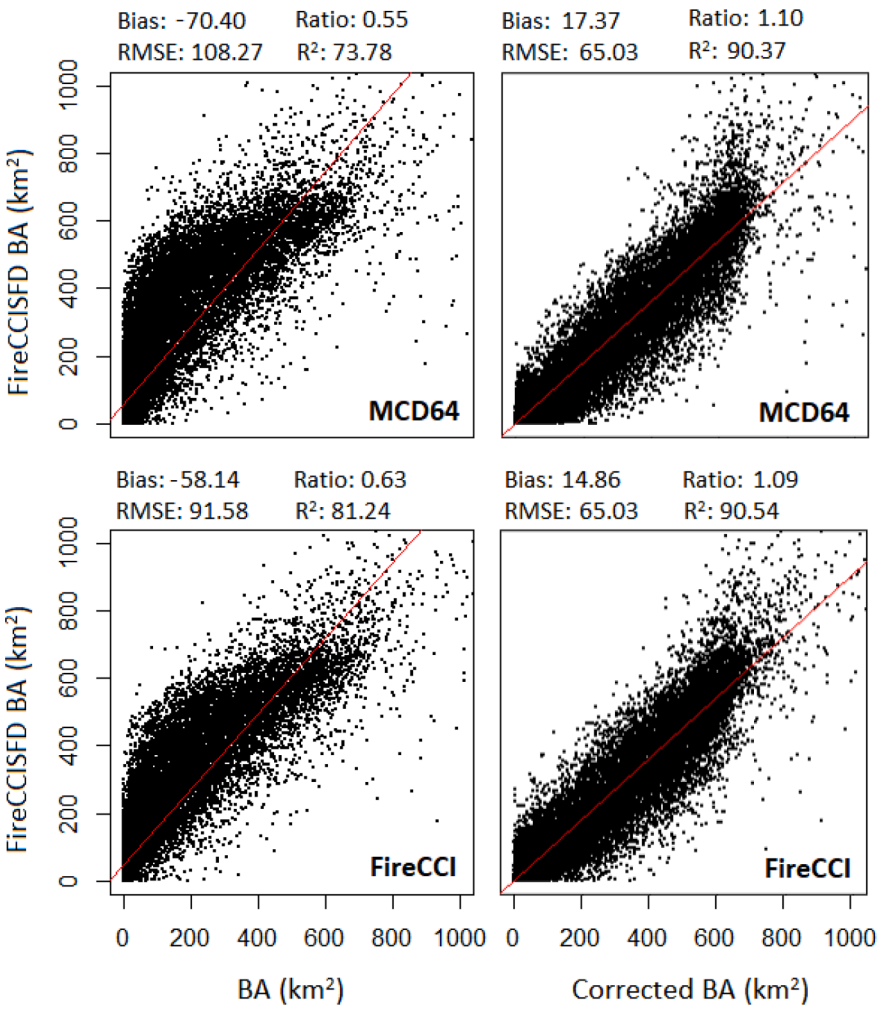


Fig. 7. Scatterplots showing the relationship between the BA from the available global BA products (MCD64, and FireCCI) and the reference BA product of 2016 (FireCCISFD11) in the validation year (2016). The panels on the left represent the original BA from the global products, and the panels on the right show the BA with errors compensated by the parsimonious RF models. R^2 values are expressed in percentage. RMSE: Root Mean Squared Error. Ratio: fraction between the total BA in the coarse resolution products (original or corrected) and the total BA in the reference product in the validation dataset.

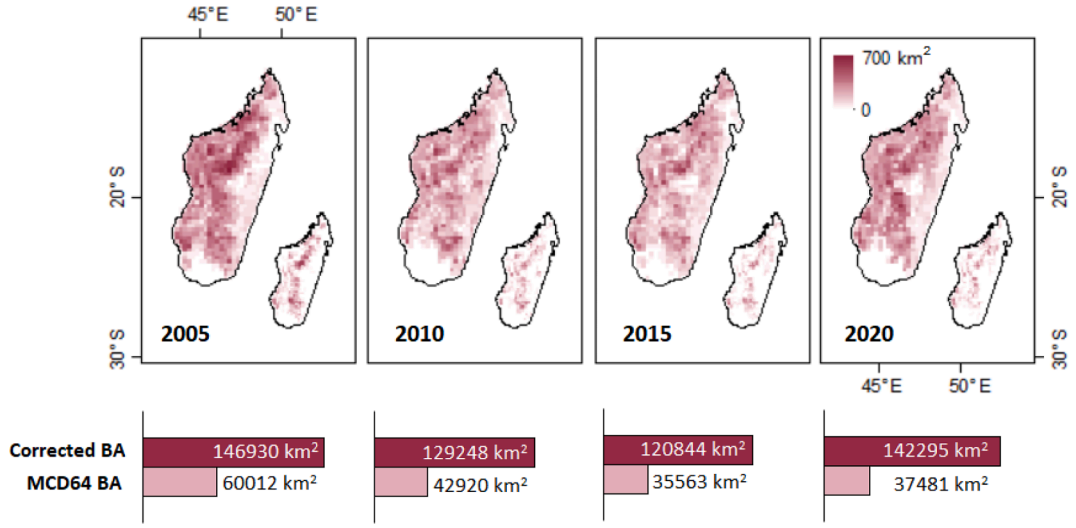


Fig. 8. Spatial patterns of BA in Madagascar according to the MCD64 with BA errors compensated by the parsimonious RF models (corrected BA, big shape of Madagascar) and BA as originally reported by the MCD64 product (MCD64 BA, small shape of Madagascar) for the years 2005, 2010, 2015 and 2020. The bar plots in the bottom indicate the total BA in the island according to both BA products.

showing the limitations of coarse resolution products to detect small fires (Chuvieco et al., 2018; Rodrigues et al., 2019; Ramo et al., 2021; Franquesa et al., 2022a; Deshpande et al., 2022).

Focusing on the predictive capacity of the developed models we found the parsimonious models (using 9–10 predictors) to be optimal for the prediction of BA biases, with decays of < 7% OOB of explained

variance with respect the full models (using 20 predictors). The parsimonious models explained a large proportion of variance (≥78.58%) of BA biases, as well as of corrected BA values (≥91.15% in the spatial validation; ≥85.35% in the temporal validation). These predictive capacities are greater than other recently developed large-scale models of fire ignition (Coughlan et al., 2021), burned area (Abolafia-Rosenzweig et al., 2022) or burn severity (Fernández-García et al., 2022). Despite this, the main asset of our models was their ability of decreasing the biases shown in the spatial validation (e.g. from −95 to near 0 km^2 for the MCD64), and in the temporal validation (from ≤ -58 to ~ 12 km^2). In relation to the latter, the 12% of BA overestimation can be attributed to the different nature of the reference product for 2016 rather than to an underperformance of the model. The reason is that the 2016 reference BA data was built using only one of the Sentinel-2 sensors (2A, 10 days revisit), and active fire data at 1 km spatial resolution (Roteta et al., 2019), thus with poorer sensing capacity than the product of 2019 used for model training. The last is based on the two Sentinel-2 sensors (combined revisit time of 5 days although the detection of BA by both was required to include an area as burned) and active fire data at 375 m (Chuvieco et al., 2022). This difference between 2016 and 2019 reference products can be corroborated by observing the BA bias of the MCD64 BA for 2016 (-70.40 km^2) (Fig. 7) and 2019 (-86.45 km^2) (Fig. 2). According to the foregoing, we can expect our model to provide better estimates over time than reported, if referencing to the standards of the FireCCISFD20 product of 2019.

The low biases of BA data corrected by our models, demonstrate their ability to be used for obtaining accurate regional and continental BA estimates comparable to the state-of-art BA product that has a relative bias of 8.4% (Chuvieco et al., 2022). As an example, the application of the model to Madagascar in 2000, 2010, 2015 and 2020 revealed a BA much larger than reported up to date for these years, with similar spatial patterns than those found for 2016 using the Sentinel-2 at 20 m (Roteta et al., 2019; Phelps et al., 2022). However, the proposed method for the refinement of coarse-resolution BA data might have some limitations. First of all, we selected the set of predictors in the basis of their relevance to explain BA detection errors, but also because of their availability since 2000 enabling the correction of BA biases across the MODIS era, so the model applicability is limited to the year 2000 onwards. Second, the predictor variables were selected prioritizing a high consistency over time, but some of them might not be perfectly consistent over time, as could be the case of the GLC_FCS30 (Zhang et al., 2021) which relies on Landsat imagery that has spectrally similar but not identical resolution and quality (e.g SLC failure; Alexandridis et al., 2013) since 2000. This might lead to potentially slight differences in model performance over time. Third, our model has been trained and tested in Sub-Saharan Africa, so the applicability in other regions must be tested for confident extrapolation. In this sense, we encourage future work to test the proposed models out of Sub-Saharan Africa as the availability of reference BA data increases, and eventually develop more comprehensive algorithms to account for the different environmental conditions and determinants of BA biases worldwide.

5. Conclusions

This study demonstrates the predictability of biases between frequently used coarse-resolution BA products and recently developed state-of-the-art BA products. Annual BA is the main predictor of biases, although other variables such as those related to the persistence of BA signals, landscape fragmentation as proxy of BA patch sizes, and maximum burnable area are also important.

The temporal and spatial validation of corrected BA data from coarse-resolution imagery showed satisfactory results with large decreases in BA bias, opening the possibility of refining annual BA estimates at the regional and continental level.

The application of the developed operational model to Madagascar shows the possibility of reconstructing BA estimates back to 2000, and confirms the large underestimations made by coarse-resolution products over time.

Funding

This study was financed by the Swiss Network for International Studies (SNIS), in the framework of the project “Fire regimes and ecosystem services in African biodiversity hotspots: can fire policies favoring climate change mitigation, biodiversity and local communities converge?”, in which CK and VF-G are the coordinator and the principal member respectively. VF-G is supported by a Margarita Salas fellowship from the Ministry of Universities of Spain, financed with European Union-NextGenerationEU and Ministerio de Universidades funds and granted by the University of León to accomplish his post-doctoral research at the University of Lausanne.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2023.103350.

References

Abolafia-Rosenzweig, R., He, C., Chen, F., 2022. Winter and spring climate explains a large portion of interannual variability and trend in western U.S. summer fire burned area. Environ. Res. Lett. 17, 054030 https://doi.org/10.1088/1748-9326/ac6886.

Alexandridis, T.E., Cherif, I., Kalogeropoulos, C., Monachou, S., Eskridge, K., Silleos, N., 2013. Rapid error assessment for quantitative estimations from Landsat 7 gap-filled images. Remote Sens. Lett. 920–928 https://doi.org/10.1080/2150704X.2013.815380.

Alonso-González, E., Fernández-García, V., 2021. MOSEV: a global burn severity database from MODIS (2000–2020). Earth Syst. Sci. Data 13, 1925–1938. https://doi.org/10.5194/essd-13-1925-2021.

Andela, N., Morton, D.C., Giglio, L., Chen, Y., Van der Werf, G.R., Kasibhatla, P.S., Defries, R.S., Collatz, G.J., Hantson, S., Kloster, S., Bachelet, D., Forrest, M., Lasslop, G., Li, F., Mangeon, S., Melton, J.R., Yue, C., Randerson, J.T., 2017. A human-driven decline in global burned area. Science 356, 1356–1362. https://doi.org/10.1126/science.aal4108.

Archibald, S., Lehmann, C.E.R., Belcher, C.M.M., Bond, W.J., Bradstock, R.A., Daniau, A-L., Dexter, K.G., Forrestel, E.J., Greve, M., Higgins, S.I., Hoffman, W.A., Lamont, B. B., McGlinn, D.J., Moncrieff, G.R., Osborne, C.P., Pausas, J.G., Price, O., Ripley, B.S., Rogers, b.m., Schwilk, D.W., Simon, M.F., Turetsky, M.R., Van der Werf, G.R., Zanne, A.E., 2018. Biological and geophysical feedbacks with fire in the Earth system. Environ. Res. Lett. 13, 033003. Available from: <https://iopscience.iop.org/article/10.1088/1748-9326/aa9ead/meta>.

Belhadj-Khedher, C., Koutsias, N., Karamitsou, A., EI-Melki, T., Ouelhazi, B., Hamdi, A., Nouri, H., Mouillot, F., 2018. A revised historical fire regime analysis in Tunisia (1985–2010) from a critical analysis of the national fire database and remote sensing. Forests 9, 59. doi: 10.3390/f9020059.

Boschetti, l., Roy, D.P., Giglio, L., Huang, H., Zubkova, M., Humber, M.L., 2019. Global validation of the collection 6 MODIS burned area product. Remote Sens. Environ. 235, 111490. doi: 10.1016/j.rse.2019.111490.

Bowman, D.M.J.S., Balch, J.K., Aratxo, P., Bond, W.J., Carlson, J.M., Cochrane, M.A., D'Antonio, C.M.M Defries, R.S., Doyle, J.C., Harrison, S.P., Johnston, F.H., Keeley, J. E., Krawchuk, M.A., Kull, C.A.,Marston, J.B., Moritz, M.A., Prentice, I.C., Roos, C.I., Scott, A.C., Swetnam, T.W., Van der Werf, G.R., Pyne, S.J., 2009. Fire in the earth system. Science 324. doi: 10.1126/science.1163886.

Cascio, W.E., 2018. Wildland fire smoke and human health. Sci. Total. Environ. 624, 586–595. https://doi.org/10.1016/j.scitotenv.2017.12.086.

Chavent, M., Genuer, R., Saracco, J., 2021. Combining clustering of variables and feature selection using random forests. Commun. Stat. 50, 426–445. https://doi.org/10.1080/03610918.2018.1563145.

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 Syst. Sci. Data 10, 2015–2031. https://doi.org/10.5194/essd-10-2015-2018.

Chuvieco, E., Pettinari, L., Koutsias, N., Forkel, M., Hantson, S., Turco, M., 2021. Human and climate drivers of global biomass burning variability. Sci. Total. Environ. 779, 146361 https://doi.org/10.1016/j.scitotenv.2021.146361.

Chuvieco, E., Roteta, E., Sali, M., Stroppiana, D., Boecttcher, M., Kirches, G., Storm, T., Khairoun, A., Pettinari, M.L., Franquesa, M., Albergel, C., 2022. Building a small fire database for Sub-Saharan Africa from Sentinel-2 high-resolution images. Sci. Total. Environ. 845, 157139 https://doi.org/10.1016/j.scitotenv.2022.157139.

Coughlan, R., Giuseppe, F.D., Vitolo, C., Barnard, C., Lopez, P., Drusch, M., 2021. Using machine learning to predict fire-ignition occurrences from lightning forecasts. Meteorol. Appl. 28, e1973.

Deshpande, M.V., Pillay, D., Jain, M., 2022. Detecting and quantifying residue burning in smallholder systems: An integrated approach using Sentinel-2 data. Int. J. Appl. Earth Obs. Geoinf. 108, 102761 https://doi.org/10.1016/j.jag.2022.102761.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Márquez, J. R., Gruber, B., Lafourcade, B., Leitao, Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36(1), 27–46. doi: 10.1111/j.1600-0587.2012.07348.x.

Driscoll, D.A., Armenteras, D., Bennett, A., Brotons, L., Clarke, M.F., Doherty, T.S., Haslem, A., Kelly, L.T., Sato, C.F., Sitters, H., Aquilué, N., Bell, K., Chadid, M., Duane, A., Meza-Elizalde, M.C., Gijohann, K.M., González, T.M., Jambhekar, R., Lazzari, J., Morán-Ordóñez, A., Wevill, T., 2021. How fire interacts with habitat loss and fragmentation. Biol. Rev. 96, 976–998. https://doi.org/10.1111/brv.12687.

Fernandes, P.M., Monteiro-Henriques, T., Guiomar, N., Loureiro, C., Barros, A.M.G., 2016. Bottom-up variables govern large-fire size in Portugal. Ecosystems 19, 1362–1375. https://doi.org/10.1007/s10021-016-0010-2.

Fernández-García, V., Quintano, C., Taboada, A., Marcos, E., Calvo, L., Fernández-Manso, A., 2018a. Remote sensing applied to the study of fire regime attributes and their influence on post-fire greenness recovery in pine ecosystems. Remote Sens. 10, 733. https://doi.org/10.3390/rs10050733.

Fernández-García, V., Santamarta, M., Fernández-Manso, A., Quintano, C., Marcos, E., Calvo, L., 2018b. Burn severity metrics in fire-prone pine ecosystems along a climatic gradient using Landsat imagery. Remote Sens. Enviorn. 2006, 205–217. https://doi.org/10.1016/j.rse.2017.12.029.

Fernández-García, V., Marcos, E., Fulé, P.Z., Reyes, O., Santana, V.M., Calvo, L., 2020. Fire regimes shape diversity and traits of vegetation under different climatic conditions. Sci. Total Environ. 716, 137137 https://doi.org/10.1016/j.scitotenv.2020.137137.

Fernández-García, V., Beltrán-Marcos, D., Fernández-Guisuraga, J.M., Marcos, E., Calvo, L., 2022. Predicting potential wildfire severity across Southern Europe with global data sources. Sci. Total Environ. 829, 154729 https://doi.org/10.1016/j.scitotenv.2022.154729.

Franquesa, M., Stehman, S.V., Chuvieco, E., 2022a. Assessment and characterization of sources of error impacting the accuracy of global burned area products. Remote Sens. Environ. 280, 113214 https://doi.org/10.1016/j.rse.2022.113214.

Franquesa, M., Lizundia-Loiola, J., Stehman, S.V., Chuvieco, E., 2022b. Using long temporal reference units to assess the spatial accuracy of global satellite-derived burned area products. Remote Sens. Environ. 269, 112823 https://doi.org/10.1016/j.rse.2021.112823.

GCOS, 2016. GCOS The Global Observing System for Climate: Implementation Needs. Report number GCOS-200. World Meteorological Organization, Geneva, Switzerland. Available from: <https://library.wmo.int/doc_num.php?explnum_id=3417>.

Giglio, L., Boschetti, L., Roy, D.P., Humber, M.L., Justice, C.O., 2018. The Collection 6 MODIS burned area mapping algorithm and product. Remote Sens. Environ. 217, 72–85. https://doi.org/10.1016/j.rse.2018.08.005.

Hawbaker, T.J., Vanderhoof, M.K., Schmidt, G.L., Beal, Y.-J., Picotte, J., Takacs, J.D., Falgout, J.T., Dwyer, J.L., 2020. The Landsat Burned Area algorithm and products for the conterminous United States. Remote Sens. Environ. 244, 111801 https://doi.org/10.1016/j.rse.2020.111801.

Hesselbarth, M.H.K., Sciaini, M., With, K.A., Wiegand, K., Nowosad, J., 2019. landscapemetrics: an open-source R tool to calculate landscape metrics. Ecography 42, 1648–1657. https://doi.org/10.1111/ecog.04617.

Hijmans, R.J., 2022. terra: Spatial Data Analysis. R package version 1.5-21. Available from: <https://CRAN.R-project.org/package=terra>.

Ju, J., Roy, D.P., 2008. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. Remote Sens. Environ. 112, 1196–1211. https://doi.org/10.1016/j.rse.2007.08.011.

Kuhn, M., 2021. caret: Classification and Regression Training. R package version 6.0-90. Available from: <https://CRAN.R-project.org/package=caret>.

Kull, C.A., 2004. Isle of Fire: The Political Ecology of Landscape Burning in Madagascar. University of Chicago Press, Chicago.

Lasslop, G., Coppola, A.I., Voulgarakis, A., Yue, C., Veraverbeke, S., 2019. Influence of fire on the carbon cycle and climate. Curr. Clim. Change Rep. 5, 112–123. https://doi.org/10.1007/s40641-019-00128-9.

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2, 18–22. Available from: <http://CRAN.R-project.org/doc/Rnews/>.

Liu, Z., Ballantyne, A.P., Cooper, L.A., 2019. Biophysical feedback of global forest fires on surface temperature. Nat. Commun. 10, 214. https://doi.org/10.1038/s41467-018-08237-z.

Lizundia-Loiola, J., Franquesa, M., Boettcher, M., Kirches, G., Pettinari, M.L., Chuvieco, E., 2021. Implementation of the burned area component of the copernicus climate change service: from MODIS to OLCI data. Remote Sens. 13, 4295. https://doi.org/10.3390/rs13214295.

Llorens, R., Sobrino, J.A., Fernández, C., Fernández-Alonso, J.M., Vega, J.A., 2021. A methodology to estimate forest fires burned areas and burn severity degrees using Sentinel-2 data. Application to the October 2017 fires in the Iberian Peninsula. Int. J. Appl. Earth Obs. Geoinf. 95, 102243 https://doi.org/10.1016/j.jag.2020.102243.

Meijer, J.R., Huijbregts, M.A.J., Schotten, K.C.G.J., Schipper, A.M., 2018. Global patterns of current and future road infrastructure. Environ. Res. Lett. 13, 064006 https://doi.org/10.1088/1748-9326/aabd42.

Melchiorre, A., Boschetti, L., 2018. Global analysis of burned area persistence time with MODIS data. Remote Sens. 10, 750. https://doi.org/10.3390/rs10050750.

Milborrow, S., 2022. plotmo: Plot a Model's Residuals, Response, and Partial Dependence Plots. R package version 3.6.2. Available from: <https://CRAN.R-project.org/package=plotmo>.

Padilla, M., Wheeler, J., Tansey, K., 2018. ESA CCI ECV Fire Disturbance: D4.1.1. Product Validation Report, Version 2.1. Tech. Rep. Available from: <https://climate.esa.int/media/documents/Fire_cci_D4.1.1_PVR_v2.1.pdf>.

Pausas, J.G., Ribeiro, E., 2013. The global fire–productivity relationship. Global Ecol. Biogeog. 22, 728–736. https://doi.org/10.1111/geb.12043.

Phelps, L., Andela, N., Gravey, M., Davis, D.S., Kull, C.A., Douglass, K., Lehmann, C.E.R., 2022. Madagascar's fire regimes challenge global assumptions about landscape degradation. Glob. Change Biol. 00, 1–17. https://doi.org/10.1111/gcb.16206.

Probst, P., Boulesteix, A., 2018. To tune or not tune the number of trees in random forest. J. Mach. Learn. Res. 18, 1–18. https://www.jmlr.org/papers/volume18/17-269/17-269.pdf.

Ramo, R., Roteta, E., Bistinas, I., van Wees, D., Bastarrika, A., Chuvieco, E., van der Werf, G.R., 2021. African burned area and fire carbon emissions are strongly impacted by small fires undetected by coarse resolution satellite data. PNAS 118, e2011160118. Available from: <https://www.pnas.org/doi/epdf/10.1073/pnas.2011160118>.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available from: <https://www.R-project.org/>.

Roces-Díaz, J.V., Santín, C., Martínez-Vilalta, J.M., Doerr, S.H., 2022. A global synthesis of fire effects on ecosystem services of forests and woodlands. Front. Ecol. Environ. 20, 170–178. https://doi.org/10.1002/fee.2349.

Rodrigues, J.A., Libonati, R., Pereira, A.A., Nogueira, J.M.P., Santos, F.L.M., Peres, L.F., Rosa, A.S., Schroeder, W., Pereira, J.M.C., Giglio, L., Trigo, I.F., Setzer, A.W., 2019. How well do global burned area products represent fire patterns in the Brazilian Savannas biome? An accuracy assessment of the MCD64 collections. Int. J. Appl. Earth Obs. Geoinf. 78, 318–331. https://doi.org/10.1016/j.jag.2019.02.010.

Roteta, E., Bastarrika, A., Padilla, M., Storm, T., Chuvieco, E., 2019. Development of a Sentinel-2 burned area algorithm: Generation of a small fire database for sub-Saharan Africa. Remote Sens. Environ. 222, 1–17. https://doi.org/10.1016/j.rse.2018.12.011.

van Dijk, D., Shoaie, S., van Leeuwen, T., Veraberbeke, S., 2021. Spectral signature analysis of false positive burned area detection from agricultural harvests using Sentinel-2 data. Int. J. Appl. Earth Obs. Geoinf. 97, 102296 https://doi.org/10.1016/j.jag.2021.102296.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., Mi, J., 2021. GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery. Earth Syst. Sci. Data 13, 2753–2776. https://doi.org/10.5194/essd-13-2753-2021.

Zheng, B., Ciais, P., Chevallier, F., Chuvieco, E., Chen, Y., Yang, H., 2021. Increasing forest fire emissions despite the decline in global burned area. Sci. Adv. 7, eabh2646. https://www.science.org/doi/full/10.1126/sciadv.abh2646.

Zhu, C., Kobayashi, H., Kanaya, Y., Saito, M., 2017. Size-dependent validation of MODIS MCD64A1 burned area over six vegetation types in boreal Eurasia: large underestimation in croplands. Sci. Rep. 7, 4181. https://doi.org/10.1038/s41598-017-03739-0.