ISPRS Journal of Photogrammetry and Remote Sensing 142 (2018) 94–105

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Review Article

Burned area estimations derived from Landsat ETM+ and OLI data: Comparing Genetic Programming with Maximum Likelihood and Classification and Regression Trees

Ana I.R. Cabral[a,*], Sara Silva[b,c], Pedro C. Silva[a], Leonardo Vanneschi[d], Maria J. Vasconcelos[a]

[a] Forest Research Centre, School of Agriculture, University of Lisbon, Tapada da Ajuda, 1349-017 Lisbon, Portugal
[b] BioISI – Biosystems & Integrative Sciences Institute, Department of Informatics, Faculty of Sciences, University of Lisbon, 1749-016 Lisbon, Portugal
[c] CISUC, Department of Informatics Engineering, University of Coimbra, 3030-290 Coimbra, Portugal
[d] NOVA IMS, Universidade Nova de Lisboa, 1070-312 Lisbon, Portugal

ARTICLE INFO

Keywords:
Burned area mapping
Genetic Programming
Savana woodlands
Classification and Regression Trees
Maximum Likelihood
Landsat ETM+/OLI

ABSTRACT

Every year, large areas of savannas and woodlands burn due to natural conditions and land management practices. Given the relevant level of greenhouse gas emissions produced by biomass burning in tropical regions, it is becoming even more important to clearly define historic fire regimes so that prospective emission reduction management strategies can be well informed, and their results Measured, Reported, and Verified (MRV). Thus, developing tools for accurately, and periodically mapping burned areas, based on cost advantageous, expedite, and repeatable rigorous approaches, is important. The main objective of this study is to investigate the potential of novel Genetic Programming (GP) methodologies for classifying burned areas in satellite imagery over savannas and tropical woodlands and to assess if they can improve over the popular and currently applied methods of Maximum Likelihood classification and Classification and Regression Tree analysis. The tests are performed using three Landsat images from Brazil (South America), Guinea-Bissau (West Africa) and the Democratic Republic of Congo (Central Africa). Burned areas were digitized on-screen to produce mapped information serving as surrogate ground-truth. Validation results show that all methods provide an overestimation of burned area, but GP achieves higher accuracy values in two of the three cases. GP is the most versatile machine learning method available today, but still largely underused in remote sensing. This study shows that standard GP can produce better results than two classical methods, and illustrates its versatility and potential in becoming a mainstream method for more difficult tasks involving the large amounts of newly available data.

1. Introduction

In tropical regions, large areas of savanna and woodlands burn every year. Occurring mainly during the dry season when herbaceous vegetation has dried out, fires are one of the main drivers of ecosystem transformation or maintenance (Bucini and Lambin, 2002), also releasing gases and particles into the atmosphere (Smith et al., 2007). In fact, estimates show that burning of savannas and woodlands in Sub-Saharan Africa accounts for more than 50% of the total global emissions from biomass burning during any typical year (Williams et al., 2012).

Land management practices induced by human activities are at the base of the majority of fire occurrences in the tropics. Shifting cultivation, agricultural expansion, deforestation and harvesting are some of the practices involving fires that may contribute to partial or complete destruction of vegetation cover, depending on fire intensity and combustion efficiency (Bucini and Lambin, 2002; Daldegan et al., 2014). Significant intensification of fire frequency or avoidance of fire occurrence can negatively affect existing ecosystems and have impacts on vegetation composition, landscape patterns, habitat types, and soil erosion processes, which in turn affect hydrological processes and the carbon cycle. Therefore, accurate and multi-temporal burned area maps are important tools that can help fire and land managers understand and assess the impacts of specific interventions, while informing landscape management strategies.

Multi-temporal data records of fire distribution, extent, and timing, correspond to historical activity data, which together with vegetation emission factors, support the quantification of emissions. The establishment of emission baselines against which the results of subsequent

* Corresponding author.
E-mail addresses: anaicabral@isa.ulisboa.pt, anaicabral70@gmail.com (A.I.R. Cabral), sara@fc.ul.pt (S. Silva), pcsilva@isa.ulisboa.pt (P.C. Silva), lvanneschi@novaims.unl.pt (L. Vanneschi), maria.perestrelo@gmail.com (M.J. Vasconcelos).

https://doi.org/10.1016/j.isprsjprs.2018.05.007
Received 9 August 2017; Received in revised form 2 May 2018; Accepted 13 May 2018

vegetation and fire management actions can be compared is essential for the Measuring, Reporting and Verification (MRV) activities necessary in carbon accounting procedures. Thus, accurately and frequently mapping burned areas over large extents, using cost advantageous, periodic, and expedite approaches is very desirable.

In the last decade, several methodologies based on remote sensing techniques have been developed and applied to recurrently map burned areas in tropical ecosystems. Some were based on coarse spatial resolution satellite data such as Moderate Resolution Imaging Spectroradiometer (MODIS), Along Track Scanning Radiometer (ATSR)/Advanced ATSR (AATSR), Satellite pour l'Observation de la Terre (SPOT) Vegetation (VGT) and National Oceanic and Atmospheric Administration (NOAA)/Advanced Very High Resolution Radiometer (AVHRR) (Brivio and Maggi, 2003; Giglio et al., 2009; Silva et al., 2005; Grégoire et al., 2003; Zhang et al., 2015). These are adequate for global and regional scale studies but are insufficient for local applications where higher detail is needed and medium to high resolution sensors are preferable for accurately mapping burned areas (Stroppiana et al., 2012).

The higher resolution sensors Landsat TM (Thematic Mapper), ETM + (Enhanced Thematic Mapper) and OLI (Operational Land Imager) are a valuable source of information and have been widely used in the development of automated methods to detect burned areas (Bastarrika et al., 2011; Chen et al., 2016; Daldegan et al., 2014; Hudak et al., 2004; Júnior et al., 2014; Korontzi et al., 2003; Hawbaker et al., 2017; Meddens et al., 2016; Laris, 2005; Liu et al., 2018; Melchiori et al., 2014; Matricardi et al., 2013; Oumar, 2015; Smith et al., 2007; Stroppiana et al., 2012; Trisakti et al., 2016). Even though classification of burned areas using Landsat images provide satisfactory results with classical approaches (Morton et al., 2011), the spectral similarities between burnt surfaces and other land cover categories, such as, water bodies, shadows, and mixed water-vegetation, still introduce spectral confusion and overlap with other classes. Therefore, it is important to explore new methods capable of increasing the discrimination between burns and other landscape features, minimizing the uncertainties (Giglio et al., 2010; Jain, 2007).

This study aims at investigating if there are comparative advantages in using Genetic Programming (GP) – one of the most powerful and underused flavors of machine learning – for identifying and mapping burned areas in Landsat ETM+/OLI imagery when compared to Maximum Likelihood classification (MLK) and Classification and Regression Tree analysis (CART) – two classical classification methods. Our research is conducted over three study areas located in Brazil, Guinea-Bissau, and the Democratic Republic of Congo. The respective tropical territories are subject to frequent and extensive fires, mainly due to human activity.

The merit of each approach is assessed by calculating the overall accuracy, *Dice* and *kappa* coefficients, and omission and commission errors over a representative sample grid of points extracted from the images. According to Padilla et al. (2014), measures such as the *Dice* coefficient that are focused on a single category (i.e. burned), are the most appropriate in the validation of Burned area products. Additionally, the agreement of the classifications with surrogate ground-truth burned area maps is calculated based on precision and recall measures (Powers, 2007). Surrogate ground-truth is obtained from visual interpretation and on-screen digitizing of burned area perimeters over the entire images. In order to assess the similitude of the overall landscape structure obtained from the on-screen digitizing with that obtained from the classifications, a set of landscape metrics are also derived and compared.

Several authors applied MLK and CART to map burned areas (Chen et al., 2016; Henry, 2008; Meddens et al., 2016; Sá et al., 2003; Sertel and Alganci, 2016; Silva et al., 2003, Thariqa et al., 2016; Verlinden and Laamanen, 2006), however, very few studies exist for GP (Silva et al., 2010). Djerriri and Mimoun (2015) successfully applied a new approach combining unsupervised classification and GP to automatically extract burned areas from Landsat 8 imagery. Also, Brumby et al. (2001) obtained encouraging results when applying GP to extract wildfire scars from Landsat 7 imagery, but found some confusion with dark cloud shadows and bare ground/rock outcrops. More recently, a different type of GP, called Geometric Semantic Genetic Programming (GSGP) (Vanneschi, 2017), was used by Castelli et al. (2015) for identification of burned areas. Although GSGP is a very powerful method, it does not provide readable models. Even though very few applications of GP for classification/data extraction of remote sensing images can be found in the literature, GP has been successfully used in several other areas, e.g., modeling and regression, image and signal processing, time series prediction, control, medicine, biology and bioinformatics, and even arts and entertainment (Poli et al., 2008). GP often yields results that are not merely academically interesting, but competitive with the work developed by humans (Koza, 2010). It is the master algorithm of evolutionary computation, and the only one with the potential to emulate all the other machine learning approaches (GP can evolve decision trees, neural networks, Bayesian networks, and almost anything else one can think of) (Domingos, 2015).

New sensors, such as those on board of the European Union (EU) Sentinel satellites[1] provide free full global coverage and high frequency optical and radar imagery. The EU Copernicus program, which also aims at providing environmental monitoring services for South America and Africa,[2] can become a driver for the systematic and high periodicity production of high resolution burned area maps over tropical regions. Therefore, methodological developments that contribute to improve operational processes while improving output accuracy may increase the usefulness of products and facilitate their respective diffusion. Recent studies have shown the feasibility of using distributed GP in long running systems dealing with big data (Hodjat et al., 2014).

2. Study areas and data

2.1. Study areas

One study area located in Brazil and two in Africa were chosen to test the performance of the burned area mapping methods: the southeastern Amazonian region of Brazil, the Coastal western region of Guinea-Bissau, and the central eastern region of Congo; each corresponding to one complete Landsat image as shown in Fig. 1.

The first area, located in eastern Amazonia, in southeastern Pará, Brazil (BRZ site) lies to the south of the Amazon River which is drier than the central and western parts of the Amazon, with annual rainfall between 1500 mm and 2000 mm and average temperatures ranging from 23 °C to 30 °C. Forests types range from lowland Amazon forest (tall trees to 40 m in height) in the north through submontane dense and open forests in the south (Olson et al., 2001). The Landsat image covers one of the most degraded regions in Amazonia, in the frontier with a drier and more populated zone where intense forest degradation driven by agriculture and cattle raising is occurring, mainly along the roads.

The second area is located in Guinea-Bissau (GB site), which is characterized by a marshy coastal plain with dry to moist (North to South) tropical climate. There are two marked seasons, a dry season between November and May, and a wet season between June and October. Total annual rain values vary from 2400 to 2600 mm in the Southwest region, and from 1200 to 1400 mm in the Northeast region (Marinho, 1946). The monthly average temperature ranges from 25.9 and 27.1 °C (Catarino, 2004). The vegetation consists of mangroves on the coast, and gradually becomes composed of mainly dry forest and savanna inland. The extent of natural vegetation patches has been decreasing in the last decade mainly due to the intensification of

[1] http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Overview4.
[2] http://www.copernicus.eu/.

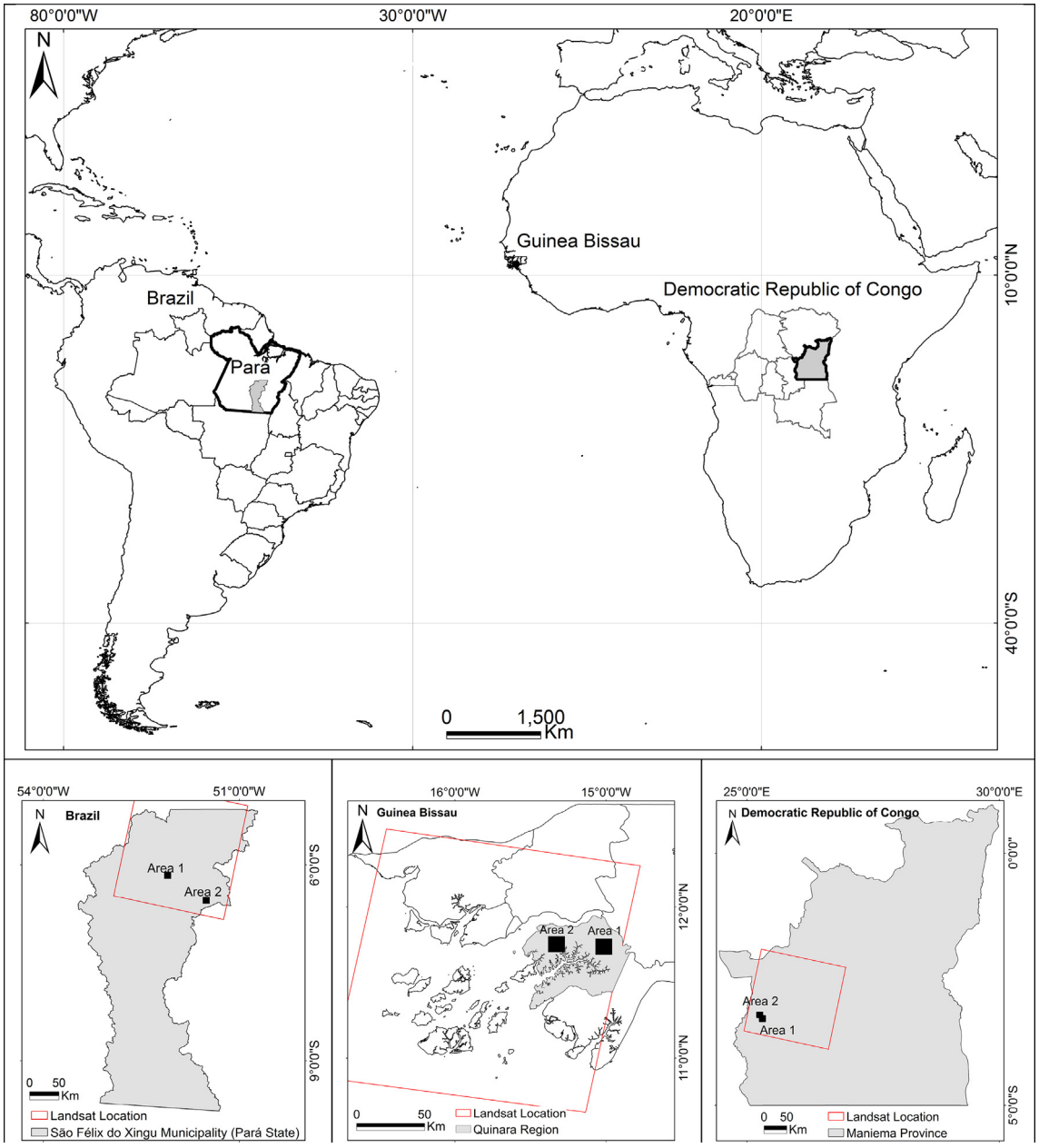

Fig. 1. Location of the three study areas in Brazil (BRZ), Guinea-Bissau (GB) and Democratic Republic of Congo (DRC) and in South America and Africa continent. Area 1 and Area 2 are locations used for assessing details in classifier performance.

subsistence livelihood practices, such as shifting cultivation, slash and burn agriculture, illegal logging and fire (MDRARNAGB, 2009), and more recently with conversion to permanent cashew tree plantations (Catarino et al., 2015).

The third study area is located in central-eastern Democratic Republic of Congo (DRC site) and it corresponds to a humid tropical climate with two distinct seasons, a dry season (with temperatures ranging between 18 and 27 °C) from June to August, and a rainy season (with temperatures ranging between 22 and 33 °C) from September to May. It is characterized by a vegetation transition zone, with congolian lowland forest in the north and miombo woodlands in the south. In the south west region, the population pressure has conducted to the degradation of the miombo woodlands, consequence of the demand for food and fuel (USAID, 2010). Although historically low, deforestation has been increasing due to several human practices, in particular, slash and burn agriculture, and charcoal production (Ickowitz et al., 2015).

2.2. Landsat data

Two cloud free Landsat 8 OLI images and one Landsat 7 ETM+ image were obtained for Brazil, Democratic Republic of Congo and Guinea Bissau, respectively. The details of each Landsat image are given in Table 1. All images were freely downloaded from the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Data Center (EDC).[3] GLOVIS datacenter provides images processed with a Standard Terrain Correction (Level 1T). This level of correction provides systematic geometric and radiometric corrections based on ground control points and a Digital Elevation Model (DEM) for topographic accuracy.

All images were acquired in the dry season to ensure the presence of burned areas and to maximize the discrimination among the vegetation

[3] http://glovis.usgs.gov/.

Table 1
Site, sensor, date and path and row of each image used in this study.

Site	Satellite (sensor)	Acquisition date DD/MM/YY	Scene or point identifier	
			Path	Row
Brazil (BRZ)	Landsat OLI	28/02/2015	225	64
Guinea-Bissau (GB)	Landsat ETM+	13/05/2002	204	52
Democratic Republic of Congo (DRC)	Landsat OLI	08/06/2013	175	62

types. Landsat images were geometrically corrected to UTM (Universal Transverse Mercator), Zone 22 South (BRZ), Zone 28 North (GB) and Zone 35 South (DRC), WGS84 (World Geodetic System 84) datum and a spatial resolution of 30 m. Bands 1 to seven of the Landsat 8 OLI images and Landsat ETM+ were used in this study to test the classifiers. Panchromatic bands from ETM+ and OLI sensors were not used since they have different spatial resolution, as well as, the OLÍs thermal bands in order to test the algorithms capacity in detecting burned areas without thermic spectral information.

3. Methodology

3.1. Reference dataset

Visual inspection of the combination of Landsat ETM+ bands 7, 4 and 3 and Landsat OLI bands 7, 5 and 4 allows a clear visual depicting of burned areas (Pereira et al., 1999). Using the 7-4-3 or 7-5-4 combination for display, according to the sensor, all visible polygons of burned areas perimeters were manually delimited on-screen to constitute a burned area ground-truth map for each of the study areas. These burned areas, and their complementary non-burned polygons, constitute the binary base map against which the maps produced by image classification shall be compared.

For training and testing the classifiers, reference data are assembled for each image. The reference sets consist of the spectral values associated with pixels extracted from the original images together with the corresponding class labels (burned/non-burned) obtained from the on-screen digitized maps. These reference data sets consist of: (a) 4872 pixel samples (2053 burned and 2819 unburned) for the BRZ site, (b) 3637 pixel samples (1889 burned and 1748 unburned) for the GB site, (c) 2849 pixel samples (877 burned and 1972 unburned) for the DRC site. Each observation in the data set (a sample) consists of the DN values for the seven spectral bands (seven explanatory variables) at a pixel location and the corresponding target value, burned (1) and unburned (0) (dependent variable).

Due to the characteristics of some of the classification methods (next section) a transformation of the data was applied to obtain a data distribution in the range 0–1 for both explanatory and dependent variables. For the Landsat ETM+ data, represented in 8-bits, a simple scaling to the interval [0 1] was applied by dividing each DN by its maximum allowed value, 255. Regarding the Landsat OLI sensor, it records pixels in 12-bits, which translates into 4096 potential grey levels for each band. However, the images are delivered as 16-bit values, scaled to 55,000 grey levels.[4] Looking at the values in each reference set, one can observe that most values are concentrated in a very narrow interval, particularly for the bands 1-4 (as can be clearly observed for the BRZ site in Fig. 2, left).

For that reason, the values of each band were mapped by a linear stretching and scaling into the interval [0,1] in such a way that its input limits correspond to the minimum and maximum values found for the band, excluding the outliers which fall outside the range [0,1] (Fig. 2, right).

[4] https://landsat.usgs.gov/landsat-8.

3.2. Classification approaches

Two classical methods to map burned areas using Landsat data are evaluated: Classification Trees (Breiman et al., 1984) and the Maximum Likelihood classifier (Richards and Jia, 2005; Theodoridis et al., 2009). Due to their popularity and success, these two methods are regarded as standards over which newly developed methods should improve. The new method proposed is based on GP, a young and successful paradigm of evolutionary computation developed by Koza (1992) which can be used for an array of different tasks, including supervised classification. A brief description of each method is presented below.

Classification trees are a non-parametric method based on a binary recursive partitioning which has been applied successfully to remote sensing data for burned area mapping in Africa (Silva et al., 2005; Silva et al., 2003; Pereira et al., 2000; Sá et al., 2001). They consist of a set of hierarchical *if-then* rules induced by identification of patterns of the predictor variables corresponding to a given class (Cabral et al., 2006). The software used to induce the classification trees was the Salford Systems CART package (Steinberg and Colla, 1997). In the generation of a single classification tree, several parameters were considered: (1) the Gini index criterion for node splitting (Breiman et al., 1984), (2) equal classification error costs for burned and unburned classes, (3) equal class prior probabilities, and (4) terminal nodes with a minimum of 20 observations. Linear combinations were employed to deal more effectively with the underlying data patterns (Cabral et al., 2006). The selection of the best classification tree size (optimal tree) was performed using the cost-complexity pruning technique, to avoid overfitting with the training data. For every generated classification tree, the cost complexity technique was based on a test sample. This technique selects an optimal compromise between the number of tree nodes and misclassification rate, and penalizes very large trees (Breiman et al., 1984; Hayes et al., 2015; Silva et al., 2003).

The MLK approach is a parametric method based on the Bayes' rule (Richards and Jia, 2005; Theodoridis et al., 2009). It computes the conditional probability $P(\omega|\boldsymbol{x})$ that a pixel is correctly assigned to a spectral class ω (burned or unburned) given that its feature vector (defined by the scaled and/or stretched DN in the seven bands) takes the value $\boldsymbol{x}$, and then assigns the pixel to the class with the highest likelihood. This conditional probability can be expressed, according to Bayes' rule, as

$$P(\omega|\boldsymbol{x}) = P(\omega)p(\boldsymbol{x}|\omega)/P(\boldsymbol{x}).$$

Here $P(\omega)$ denotes the *a priori* probability that a class ω occurs in the study area (estimated as the proportion of training samples in class ω) and $P(\boldsymbol{x})$ is the probability of finding a pixel of any class with feature vector $\boldsymbol{x}$ (which is omitted from the classifier since it is the same for all classes). The samples in each spectral class ω are assumed to follow a multivariate normal distribution with class conditional density function $p(\boldsymbol{x}|\omega)$. The corresponding class signatures (mean μ_ω and covariance matrix Σ_ω) are estimated using the Maximum Likelihood method on each set of training samples assigned to ω. The class ω with the highest likelihood $P(\omega|\boldsymbol{x})$, or equivalently, the highest $log\ P(\omega|\boldsymbol{x})$, corresponds to the class with the highest score for the so-called discriminant function,

$$g_\omega = logN_\omega - \frac{1}{2}(\boldsymbol{x}-\mu_\omega)^t\Sigma_\omega^{-1}(\boldsymbol{x}-\mu_\omega) - \frac{1}{2}log|\Sigma_\omega|,$$

where N_ω is the number of training samples in class ω and $|\Sigma_\omega|$ denotes the determinant of the covariance matrix Σ_ω. In order to be able to compare the outcomes among the three classification approaches no discrimination threshold was imposed on the MLK outputs, and therefore all pixels were classified either as burned or unburned. All computations were directly implemented in the statistical software R.

Genetic Programming is a non-parametric method for the

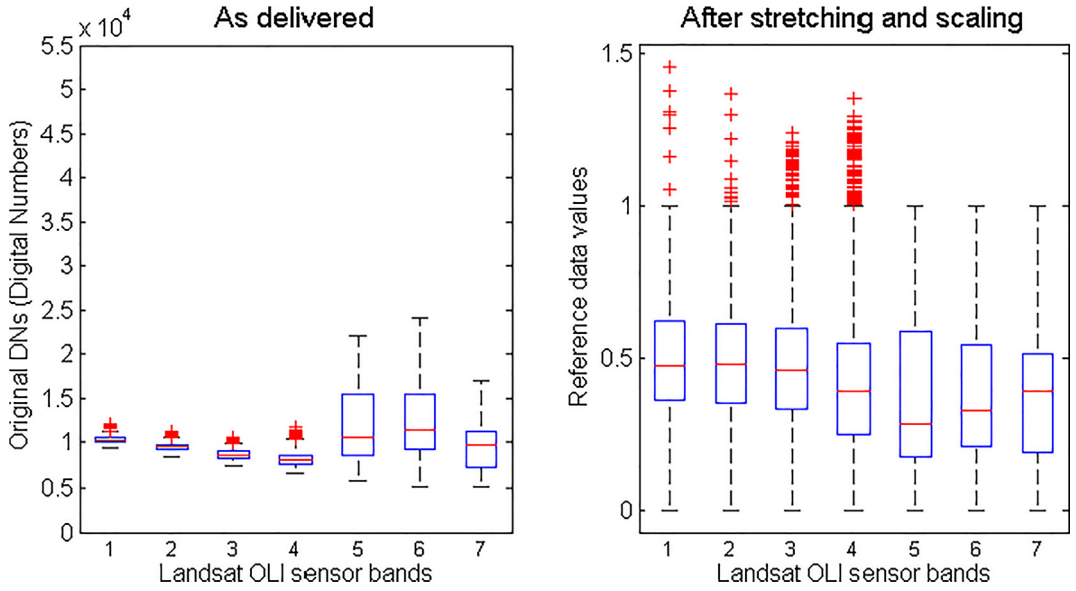

Fig. 2. Boxplots of original DNs as delivered (left) and values obtained after stretching and scaling (right), on each of the Landsat OLI sensor bands, for the BRZ site reference data. On each box, the central mark is the median, the edges of the box are Q1 and Q3 (the 25th and 75th percentiles, respectively), and the whiskers extend to 1.5 times the interquartile range above Q3 and below Q1. The outliers are the points falling outside this interval, represented by crosses.

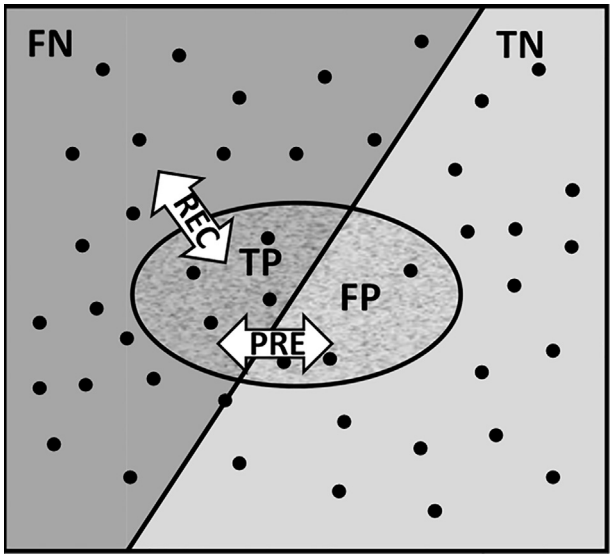

Fig. 3. Graphical representation of the Precision/Recall concepts. The dark and light grey areas in the square correspond to pixels classified as unburned (negative), whereas the ellipsoid shaded areas correspond to pixels classified as burned (positive). The former is further divided into the set of true negative (TN) and false negatives (FN) while the latter is decomposed into true positives (TP) and false positives (FP). In this framework, the Precision = TP/(TP + FP) and Recall = TP/(TP + FN) measures can be viewed as the proportions of shaded areas depicted in the figure.

automated learning of computer programs, using Darwinian selection and Mendelian genetics as sources of inspiration (Koza, 1992; Poli et al., 2008). Starting from an initial population of randomly created programs representing the potential solutions to a given problem, it evaluates the fitness of each, quantifying how well the program solves the problem. New generations of programs are iteratively created by selecting parents based on their fitness, and breeding them using genetic operators like crossover and mutation, where pieces of code are swapped and modified, respectively. Because fitter individuals are selected more often and given the chance to pass their best characteristics to their offspring, the population tends to improve in quality along successive generations. Theoretically, GP can solve any problem whose candidate solutions can be evaluated and compared, making it a widely applicable technique.

The individuals in GP are most commonly represented as parse trees, but depending on the particular set of initial elements used to build them, different types of programs are evolved, giving rise to different classification methods. Even though qualitative ingredient sets, such as rules, could be used, for simplicity we used the four basic arithmetic operators only (addition, subtraction, multiplication and division). Thus, the programs evolved are arithmetic expressions combining the different variables, outputting real values that are expected to be close to the target values 0 or 1. To the output of such an expression, a cutoff value in the continuous output range between 0 and 1, is applied in order to obtain a binary classification (1 or 0, respectively, burned or unburned). Since GP searches for classifiers whose output values stand as close as possible to the expected 0 or 1 classification, 0.5 is generally regarded as an adequate cutoff value.

Table 2
Description of the landscape metrics.

Structural category	Landscape metric	Abbreviation	Description	Units/Value
Area/Density/Edge	Number of patches	NP	The number of patches	No units; > 1
	Mean patch size	MPS	The average mean surface of the patches	Hectares; > 0
	Largest Patch Index	LPI	Percentage of the landscape area occupied by the largest patch	No units; 0–100%
Shape	Mean Shape Index	MSI	Ratio between the perimeter of a patch and the perimeter of the simplest patch in the same area (Indicator of class shape complexity)	No units; > =1
	Landscape Shape Index	LSI	Complexity of landscape structure	No units; > =1

Table 3
Number of pixel samples for Brazil, Guinea-Bissau and Democratic Republic of Congo sites.

		Image sites		
		Brazil	Guinea-Bissau	Democratic Republic of Congo
Number of pixel samples	Burned	27	43	59
	Unburned	3498	564	3471
	Total	3525	607	3530

A common and pernicious problem affecting GP methods is the *bloat* phenomenon, i.e., an excess of code growth without a corresponding improvement in fitness (Silva and Costa, 2009), which often stagnates the evolution and causes the proposed solutions to become too complex and difficult to interpret. In order to avoid bloat in GP, the most advanced state-of-the-art bloat control method was used, Operator Equalisation (OpEq) (Silva et al., 2012). For the DRC site (Landsat ETM + data), the Mutation-based variant of OpEq (*MutOpEq*) was used. This variant was the one producing the shorter solutions in earlier work (Silva et al., 2010), and its search dynamics was found to achieve good results with minimal computational effort (Silva and Vanneschi, 2011). However, for the BRZ and GB sites (Landsat OLI data) the more powerful variant of OpEq (*DynOpEq*) had to be used in order to produce competitive results in the same number of generations.

3.3. Experimental setup and selection of classifiers

To assess the learning and generalization ability of the different classification techniques, 30 random partitions were created from the reference data set collected as explained in Section 3.1; 70% used for training and the remaining 30% used for testing. Each of the three classification methods ran 30 times, once over each training and corresponding testing set. The classifier achieving the highest accuracy among the 30 classifiers was selected to generate the final map used for overall performance assessment. A more detailed description is given below for each of the classification approaches.

For generating the best classification tree, four steps were implemented in each run (Breiman et al., 1984): (1) A large initial tree T_0 was grown using only the training samples, according to several defined parameters as mentioned in Section 3.2, (2) The branches of T_0 were iteratively pruned, in order to obtain a sequence of optimally nested subtrees, (3) The misclassification costs of each subtree were estimated from the independent testing samples, (4) The subtree with smallest misclassification cost (highest accuracy) was selected as the best subtree (optimal tree size). This sequence of steps was repeated for the 30 partitions, with all classifiers reaching overall accuracies around 99% at the three sites, and the classifier attaining the smallest error chosen to generate the final burned map.

In the MLK case, the class signatures, mean μ_ω and covariance matrix Σ_ω, were estimated using the Maximum Likelihood method,

$$\hat{\mu}_{ML} = \frac{1}{N_\omega}\sum_{i=1}^{N_\omega} \boldsymbol{x}_{i,\omega},\ \hat{\Sigma}_{ML} = \frac{1}{N_\omega}\sum_{i=1}^{N_\omega} (\boldsymbol{x}_{i,\omega}-\hat{\mu}_{ML})(x_{i,\omega}-\hat{\mu}_{ML})^t,$$

where $\boldsymbol{x}_{i,\omega}$, $i = 1, ...,N_\omega$, denotes the set of training samples belonging to ω, with the number of training samples per class approaching or surpassing 100 times the number of feature variables (7 spectral bands), as recommended by Swain and Davis (1978). In all cases the classifiers were able to learn from their training sets with overall accuracies above or close to 98% for the BRZ site, 99% for the GB site and 96% for the DRC site.

For the GP methods (Silva, 2009), each run employed a population of 500 individuals and was allowed to evolve for 200 generations. The parameter settings were defined according Koza (1992), except with respect to the following details. The selection of the parents of the next generation was made with lexicographical tournaments (Luke and Panait, 2002) of size 10. The offspring were created using standard subtree crossover, and no mutation (Koza, 1992). Fitness was measured as the Root Mean Squared Error (RMSE) between expected and predicted values on the training set. For each run, the solution returned was the one among all the individuals in the last generation that achieved the lowest RMSE. It was observed that all the 30 different classifiers (30 partitions) of each of the three different data sets were able to learn their training sets and to generalize on the respective test sets, with accuracy values rounding 99% very well balanced between the two classes.

3.4. Accuracy assessment

The burned area map derived from each classifier in each study area was compared with the corresponding surrogate ground-truth fire perimeters map using two types of map-to-map comparisons.

The first type, consisted of a direct comparison between classified burned areas with fire perimeters digitized on-screen, and included in the ground truth map. The agreement between the landscape structures of the two maps (classified and ground truth) was also assessed by means of the Precision-Recall measures (Powers, 2007) and of several landscape metrics (McGarigal et al., 2002). The Precision measure is the fraction of truly burned pixels (from the ground truth) with respect to the number of pixels classified as burned and it is related with commission errors. Recall is the ratio of true burned pixels caught by the classifier with respect to the total number of truly burned pixels (Fig. 3), and it is related with omission errors. The Precision and Recall indices were computed through cross-tabulation of the number of pixels lying in the same class in the burned area maps and in the ground-truth map versus the number of pixels lying in distinct classes.

The similarity between the landscape structure derived from each classification method and the landscape structure observed in the ground-truth maps was assessed using several landscape metrics (Table 2), resorting to the FRAGSTATS v4.2.597 software (a spatial pattern analysis program for categorical maps) (McGarigal et al., 2002).

For the second map-to-map comparison type, a new systematic grid of 5000 points with a random origin was generated. This validation grid is based upon the minimum bounding box containing the satellite image area at each site. All points located at the *no Data* region were removed. For each site, burned and unburned classes were identified overlapping the validation grid over the corresponding Landsat image. The final number of burned and unburned samples pixels per site are shown in Table 3.

For each sample, a pixel-by-pixel comparison over the validation grid was performed to calculate the overall accuracy, *Dice* and *kappa* coefficient (K), and omission and commission errors (Foody, 2002; Padilla et al., 2014) for all classifiers and sites.

4. Results and discussion

Visual comparison between the classified maps and the satellite images shows distinct performance patterns according to the site location and classification method adopted. A detail of fire perimeter agreement between the ground-truth maps and the classified burned areas is shown for Area 1 and Area 2 of each of the three locations in Fig. 4.

The Precision and Recall values obtained for each burned area map are reported in Table 4. These values are also depicted as points in a plot to facilitate their interpretation (Fig. 5).

The joint analysis of the results of Table 4 and Fig. 5 clearly reveals that the Recall values have a narrower range of variation (between 72.0 and 90.2), whereas the Precision values (between 9.1 and 73.4) exhibit greater dispersion. This means that the three classification methods have more consistent and satisfactory behavior in regard to omission errors than to commission errors. In fact, while all methods could detect

high percentages of burned pixels, they all overestimate burned areas. The best area agreement was observed for MLK in the BRZ site, and all methods had a similar performance in the GB site. GP (both simple and complex) achieved the best agreement in DRC.

The poor results observed for CART and GP at the BRZ site are attributable to the existence of a mislabeled polygon in the reference data set (Silva et al., 2017). This polygon, identified after observing the results, consists of 20 pixels of forest that were mislabeled as burned, representing less than 0.5% of the reference set. Being a probabilistic approach, the MLK method was able to adequately deal with this small

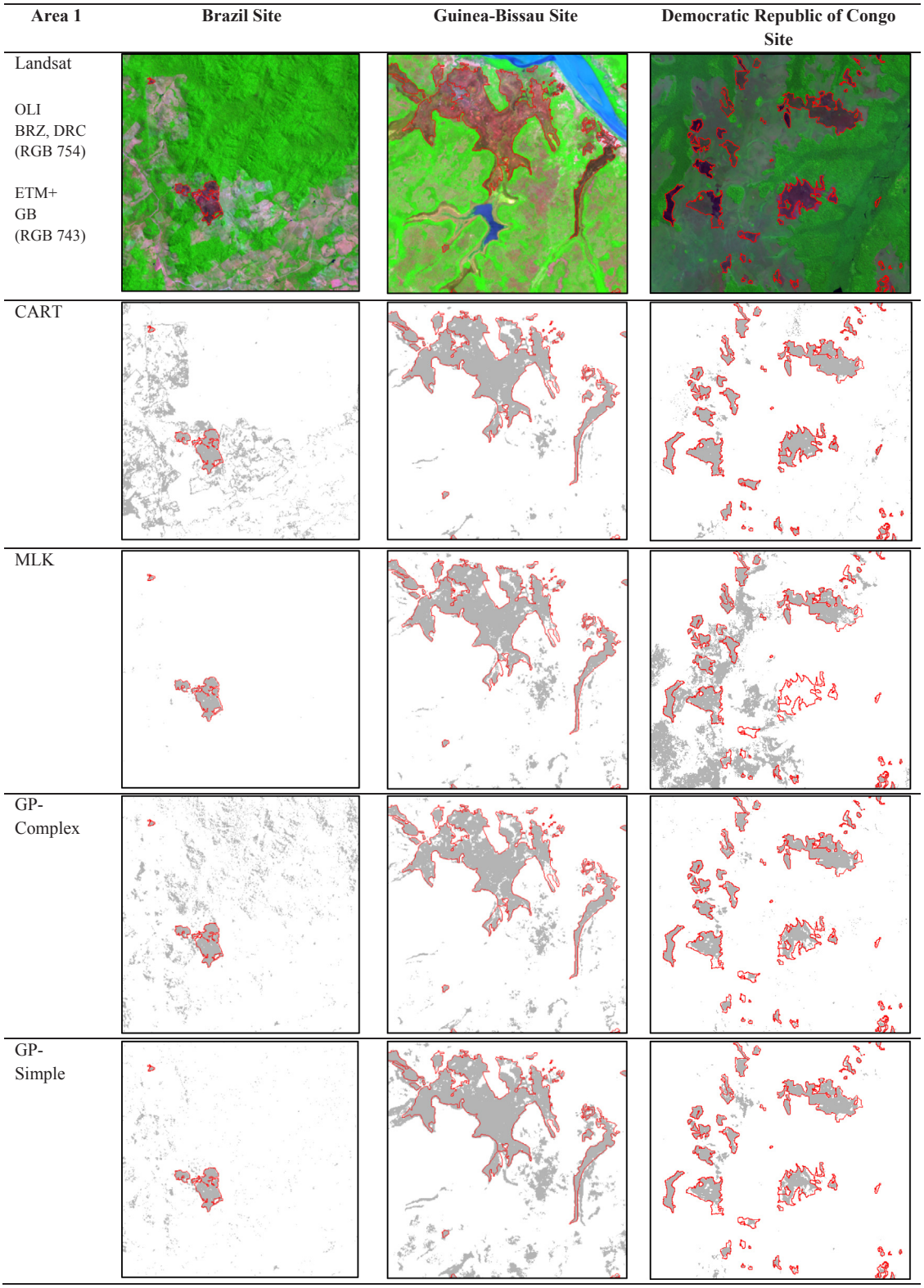

Fig. 4. Landsat OLI (RGB 754) and Landsat ETM+ (RGB743) color composites crops corresponding to areas 1 and 2, located at the small squares of Fig. 1, together with the burned area maps (in grey) derived from the three classification methods for each study site. The red line polygons delimit the ground-truth fire perimeters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

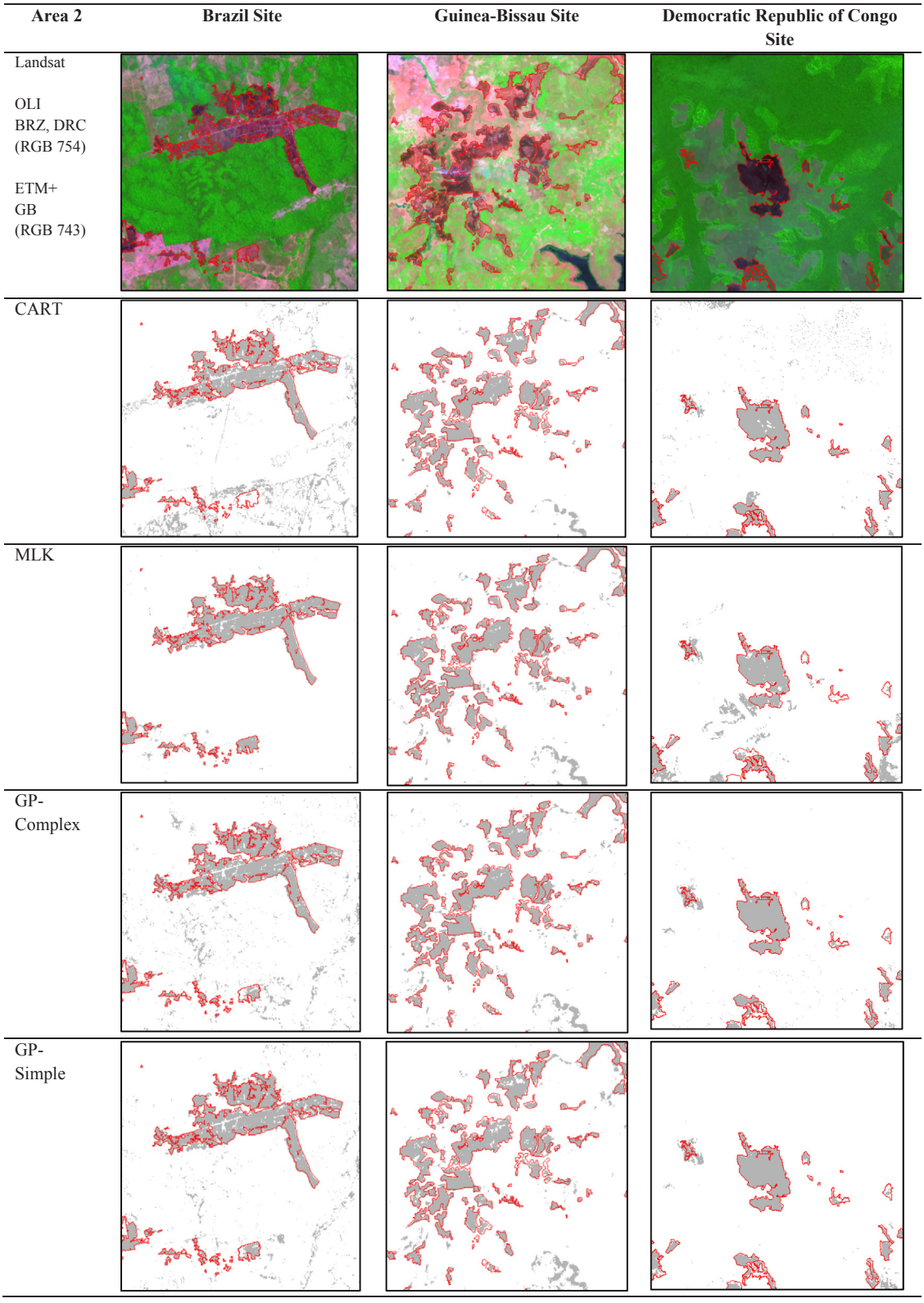

Fig. 4. (*continued*)

percentage of error in the training set, since they hardly affected the class signature estimates. However, CART and, quite surprisingly, also GP, were highly affected by the error. An extensive analysis of how a robust method like GP can be so strongly affected by less than 0.5% of mislabeled samples, is work in progress (Silva et al., 2017). It was already found that, in the first 30 generations of the evolutionary process, GP ignores the errors as it focuses on the easier task of improving the fitness over the correctly labelled samples. As the improvement of fitness becomes harder, it eventually also learns the incorrectly labelled samples. A much better understanding of this special type of hidden overfitting was obtained, and a simple solution has been proposed (Silva et al., 2017), one that may be easily applied to upcoming studies.

Table 4
Precision and Recall values obtained comparing baseline burned area polygons with burned area maps at the Brazil, Guinea-Bissau and Democratic Republic of Congo sites with each method.

Methods	Brazil		Guinea-Bissau		Democratic Republic of Congo	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)	Precision (%)	Recall (%)
CART	9.1	76.8	65.5	81.4	22.9	81.6
MLK	73.4	90.2	54.1	85.2	36.8	72.0
GP-Complex	16.4	80.3	53.7	85.3	60.1	78.2
GP-Simple	29.1	82.5	45.8	85.1	62.7	76.2

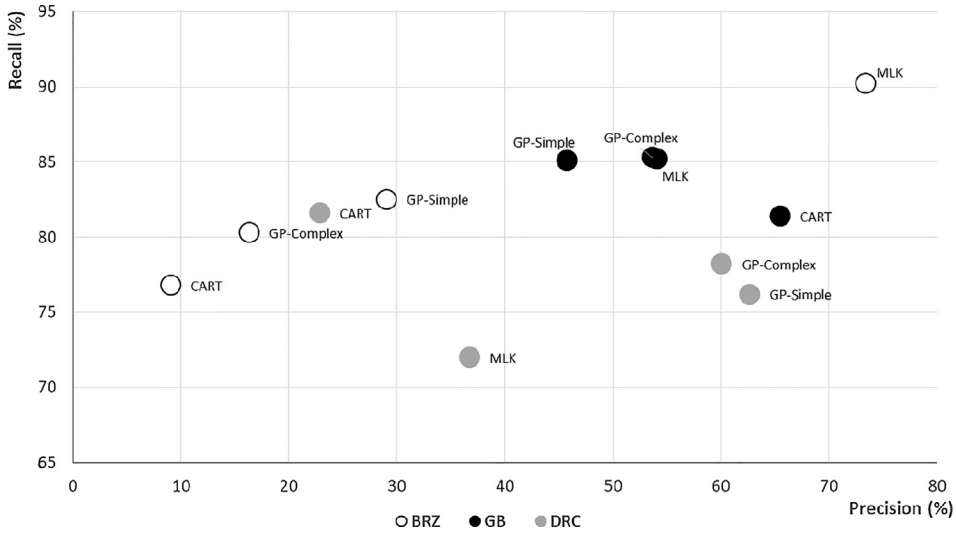

Fig. 5. Plot of Precision against Recall values for each method and site.

Table 5
Summary of the classification accuracies for each classifier for Brazil, Guinea-Bissau and Democratic Republic of Congo study areas.

Study site	Methods	Accuracy			Class	Error	
		Overall (%)	Kappa	Dice		Omission (%)	Commission (%)
Brazil	CART	95.5	0.23	0.24	Burned	7.4	82.3
					Unburned	4.5	0.1
	MLK	99.7	0.84	0.84	Burned	11.1	20.0
					Unburned	0.1	0.1
	GP-Complex	98.4	0.49	0.50	Burned	0.0	67.1
					Unburned	0.0	1.6
	GP-Simple	98.7	0.51	0.52	Burned	7.4	64.2
					Unburned	1.3	0.1
Guinea-Bissau	CART	96.5	0.67	0.69	Burned	46.5	4.2
					Unburned	0.9	3.4
	MLK	96.7	0.70	0.71	Burned	41.9	7.4
					Unburned	0.4	3.1
	GP-Complex	97.5	0.78	0.79	Burned	34.9	0.0
					Unburned	0.0	2.6
	GP-Simple	97.0	0.73	0.75	Burned	37.2	6.9
					Unburned	0.4	2.8
Democratic R of Congo	CART	96.7	0.47	0.48	Burned	10.2	67.3
					Unburned	3.1	0.2
	MLK	98.2	0.59	0.60	Burned	20.3	52.0
					Unburned	1.5	0.4
	GP-Complex	99.7	0.91	0.91	Burned	3.4	13.6
					Unburned	0.3	0.1
	GP-Simple	99.8	0.93	0.94	Burned	5.1	8.2
					Unburned	0.1	0.1

For now, we remain focused on the standard GP method.

Accuracy values, *Dice* and *kappa* coefficients and omission and commission errors, obtained by cross tabulating the pixels of the validation grid with the classified maps are reported in Table 5. A closer inspection of Table 5 shows that the GP methods at the DRC and GB sites present the highest values of overall accuracy and *Dice* and *kappa* coefficients. However, for the GB site, the observation of the few misclassified pixels revealed that about half of them were common to all

Table 6
Landscape metrics for Ground Truth Map and for each burned area map for Brazil, Guinea-Bissau and Democratic Republic of Congo. NP-Number of patches; MPS-Mean Patch Size; LPI-Largest Patch Index; MSI-Mean Shape Index.

Study site	Methods	Area (ha)	Landscape metrics			
			Area/Density/Edge			Shape
			NP	MPS	LPI	MSI
Brazil	Ground Truth Map	23 766	1379	3844.8	70.2	1.7
	CART	198 773	15,574	340.4	66.9	2.4
	MLK	**29 234**	**2099**	**2525.9**	**70.1**	**1.8**
	GP-Complex	116 043	14,631	362.4	69.0	2.1
	GP-Simple	67 391	7959	666.2	69.7	2.1
Guinea-Bissau	Ground Truth Map	56 559	3142	1638.4	70.1	1.9
	CART	**70 210**	**6078**	**846.9**	**70.1**	**1.8**
	MLK	89 029	6187	832.0	64.4	1.9
	GP-Complex	85 669	8415	611.8	70.1	1.9
	GP-Simple	105 181	8763	587.5	70.1	1.9
Democratic R of Congo	Ground Truth Map	42 534	1227	4016.9	69.8	1.4
	CART	151 940	18,078	272.6	68.9	2.2
	MLK	83 251	4252	1159.2	69.0	2.2
	GP-Complex	55 412	3528	1397.0	69.6	1.9
	GP-Simple	**51 780**	**3190**	**1545.1**	**69.7**	**1.9**

classifiers and correspond to pixels lying in wet grasslands around the sea arms that go inland in the coastal areas. These pixels are often confused with burns (Pereira et al., 1999), due to their similar spectral signatures. In the GB case, none of the three methods has actually learned these errors, which speaks strongly of their robustness. Nevertheless, the *Dice* and *kappa* values obtained by CART at the three sites was clearly low and below the results reported in some studies using the same type of data (Meddens et al., 2016; Thariqa et al., 2016). However, the MLK accuracy performance was only marginally inferior when compared with other published works (Chen et al., 2016; Oumar, 2015). Higher accuracy results have been obtained by different authors but most using low resolution imagery (Edwards et al., 2018; Pereira et al., 2000) or multi-temporal approaches (Sá et al., 2003; Silva et al., 2003, 2005) which are not directly comparable.

The commission and omissions errors varied considerably depending on the classifier and study site. The higher commission errors occurred at the BRZ site for CART, followed by the two GPs (corresponding to the low Precision values). These are most likely related to the labelling errors detected in the training set, as already pointed out. At the DRC site, CART and MLK were the worst performers presenting high percentages of omission and commission errors. This is in accordance with the low Precision scores attained by both classification methods at the DRC site, with MLK not being able to detect around 20% of the burned area. This fact is probably related with parameter estimation errors of the class signatures due to the smaller number of training samples that were available for this site.

The landscape metrics calculated for the ground-truth map and for the burned area maps, derived from the three classification methods are reported in Table 6. The results show that no approach was consistently better for all study areas. For the BRZ site, the best approximation in terms of the number of patches, complexity shape and size of the largest patch, was obtained for MLK followed by GP-Simple. The values of the Mean Shape Index (MSI) were not significantly different between sources, indicating similar shapes between the ground truth map and the classified burned patches. All other approaches presented more fragmented areas with higher shape complexity. For the GB site, CART yields the highest agreement between ground-truth and the classified burned area maps, but with higher landscape fragmentation, lower patch sizes, and simpler shape. The most fragmented patterns were also generated by the GP and MLK methods. At the DRC site, GP-Simple and GP-Complex classifications result in the best landscape structure agreement between the classification and the ground-truth map. Nevertheless, the burned maps produced by GP-Simple and GP-Complex exhibit higher fragmentation and disaggregation of the burned patches.

The results presented above reveal that no single method is better than the others. They also reveal that the implementation of standard GP was able to produce some of the best results, indicating that it is worthwhile to invest some effort in the exploitation of other GP characteristics to further improve its applicability. With the exception of linear combinations in CART, both CART and MLK were also used in their standard format, leaving room for some additional explorations in terms of the *a priori* probabilities for each class, and the choice of the best threshold in MLK. In turn, GP is more versatile and, based on these results, will be molded in several different ways in upcoming studies. For example, the fitness function used to evaluate each candidate solution could be specialized in optimizing not only the overall accuracy, but also the precision and recall measures, or the *Dice* and *kappa* coefficients, or a combination of all of them in a multi-objective setting. The set of choices is actually very large and does not facilitate a full exploration. In any case, it is not expected that a single setting will return the best results in all sites and for all the quality assessment measures. Instead, the major improvements may arrive from the development of non-standard GP elements that provide the method with added robustness to labelling errors (Silva et al., 2017), as discussed above.

5. Conclusions

The present study has compared three different methods (GP, CART and MLK) for detecting burned areas in three different sites (BRZ, GB and DRC). The results have shown that, depending on the study area and sensor type, the three methods achieved different accuracies. Nevertheless, the accuracies of the burned area maps produced by the GP methods were always higher than those produced by CART, and only at the BRZ site GP performed worse than MLK, being affected by a small percentage of mislabeling errors.

GP revealed to be a very versatile supervised classification method, obtaining some of the best results with practically no tuning of its immense array of possible configurations that remain open to further experiment. The sensitivity revealed by this method to a small percentage of labelling errors was a surprise, and prompted further studies with the development of a new solution that is still work in progress and will be used in upcoming studies.

Despite the relatively high *Dice* and *kappa* (higher than 0.90 in some cases) and accuracy levels (higher than 99% in some cases) of the different classifiers, many fire scars were missed, and other areas were incorrectly mapped as fires (for example, shadows due to the topographic effect), even in the best performing case. Each study site is characterized by different land cover types, and some of these types show similar spectral values with those representing burned sites, which can introduce classification errors. Also, the well-known fact that the spectral signal of fire scars has a fast decay in savannas may explain some of the difficulties of accurately mapping burned areas in a single date approach. Further research is needed to establish the degree of improvement that can be achieved, in particular with relatively unexplored methods like GP, in a multi-temporal, data heavy and high-resolution study.

Acknowledgments

The projects leading to this work have received funding from the European Union's Horizon 2020 Research and innovation programme under the Marie Skłodowska - Curie grant agreement No. 691053, and from the FCT/MCTES/PIDDAC (Portuguese Foundation for Science and

Technology) which funded strategic projects UID/AGR/002389/2013 (CEF) and UID/MULTI/04046/2013 (BioISI).

References

Bastarrika, A., Chuvieco, E., Martin, M.P., 2011. Mapping burned areas from Landsat TM/ETM+ data with a two-phase algorithm: balancing omission and commission errors. Remote Sens. Environ. 115, 1003–1012. http://dx.doi.org/10.1016/j.rse.2010.12.005.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Belmont, California, pp. 173.

Brivio, P.A., Maggi, M., 2003. Mapping burned surfaces in Sub-Saharan Africa based on multi-temporal neural classification. Int. J. Remote Sens. 24, 4003–4018. http://dx.doi.org/10.1080/0143116031000103835.

Bucini, G., Lambin, E.F., 2002. Fire impacts on vegetation in Central Africa: a remote-sensing-based statistical analysis. Appl. Geogr. 22, 27–48. http://dx.doi.org/10.1016/S0143-6228(01), 00020-0.

Brumby, S.P., Theiler, J., Perkins, S., Harvey, N.R., Szymanski, J.J., 2001. Genetic programming approach to extracting features from remotely sensed imagery. In: Fourth International Conference on Image Fusion, FUSION 2001, Montreal, Quebec, Canada.

Cabral, A.I.R., Vasconcelos, M.J.P., Pereira, J.M.C., Martins, E., Bartholomé, E., 2006. A land cover map of Southern hemisphere Africa based on SPOT-4 Vegetation data. Int. J. Remote Sens. 6 (20), 1053–1074. http://dx.doi.org/10.1080/01431160500307409.

Castelli, M., Vanneschi, L., Popovic, A., 2015. Predicting burned areas of forest fires: an artificial intelligence approach. Fire Ecol. 11 (1), 106–118. http://dx.doi.org/10.4996/fireecology.1101106.

Catarino, L.M.F., 2004. Fitogeografia da Guiné-Bissau. Provas de doutoramento em Engenharia Agronómica. Instituto Superior de Agronomia, Universidade Técnica de Lisboa, pp. 440.

Catarino, L., Menezes, Y., Sardinha, R., 2015. Cashew cultivation in Guinea-Bissau – risks and challenges of the sources of a cash crop. Sci. Agricola 72 (5), 459–467. http://dx.doi.org/10.1590/0103-9016-2014-0369.

Chen, W., Moriya, K., Sakai, T., Koyama, L., Cao, C.X., 2016. Mapping a burned area from Landsat TM data by multiple methods. Geomatics, Nat. Hazards Risk 7 (1), 384–402. http://dx.doi.org/10.1080/19475705.2014.925982.

Daldegan, G.A., de Carvalho, O.A., Guimarães, R.F., Gomes, R.A.T., Ribeiro, F. de F., McManus, C., 2014. Spatial patterns of fire recurrence using remote sensing and GIS in the Brazilian Savanna: Serra do Tombador Nature Reserve, Brazil. Remote Sens. 6, 9873–9894. http://dx.doi.org/10.3390/rs6109873.

Djerriri, K., Mimoun, M., 2015. Genetic programming and one-class classification for discovering useful spectral transformations. IGARSS 425–428.

Domingos, P., 2015. The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Basic Books, 352 pp.

Edwards, A.C., Russel-Smith, J., Maier, S.W., 2018. A comparison and validation of satellite-derived fire severity mapping techniques in fire prone north Australian savannas: extreme fires and tree stem mortality. Remote Sens. Environ. 206, 287–299. http://dx.doi.org/10.1016/j.rse.2017.12.038.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201. http://dx.doi.org/10.1016/S0034-4257(01), 00295-4.

Giglio, L., Loloda, T., Roy, D.P., Quayle, B., Justice, C.O., 2009. An active-fire based burned area mapping algorithm for the MODIS sensor. Remote Sens. Environ. 113, 408–420. http://dx.doi.org/10.1016/j.rse.2008.10.006.

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

Grégoire, J.-M., Tansey, K., Silva, J., 2003. The GBA2000 initiative: developing a Global Burned Area Database from SPOT-VEGETATION imagery. Int. J. Remote Sens. 24 (6), 1369–1376. http://dx.doi.org/10.1080/0143116021000044850.

Hayes, T., Usami, S., Jacobucci, R., McArdle, J.J., 2015. Using Classification and Regression Trees (CART) and Random Forests to analyze attrition: results from two simulations. Psychol. Aging 30 (4), 911–929. http://dx.doi.org/10.1037/pag0000046.

Henry, M.C., 2008. Comparison of single- and multi-date Landsat data for mapping wildfire scars in Ocala National forest, Florida. Photogr. Eng. Remote Sens. 74 (7), 881–891. http://dx.doi.org/10.14358/PERS.74.7.881.

Hodjat, B., Hemberg, E., Shahrzad, H., O'Reilly, U.-M., 2014. Maintenance of a long running distributed genetic programming system for solving problems requiring big data. In: Genetic Programming Theory and Practice XI. Springer, pp. 65–83. http://dx.doi.org/10.1007/978-1-4939-0375-7_4.

Hudak, A.T., Fairbanks, D.H.K., Brockett, B.H., 2004. Trends in fire patterns in a southern African savanna under alternative land use practices. Agric. Ecosyst. Environ. 101, 307–325. http://dx.doi.org/10.1016/j.agee.2003.09.010.

Ickowitz, A., Slayback, D., Asanzi, P., Nasi, R., 2015. Agriculture and deforestation in the Democratic Republic of the Congo: A synthesis of the current state of knowledge. In: Occasional paper 119, Bogor, Indonesia, CIFOR.

Jain, A.K., 2007. Global estimation of CO emissions using three sets of satellite data for burned area. Atmos. Environ. 41, 6931–6940. http://dx.doi.org/10.1016/j.atmosenv.2006.10.021.

Júnior, A.C.P., Oliveira, S.L.J., Pereira, J.M.C., Turkman, M.A.A., 2014. Modeling fire frequency in a cerrado savanna Protected area. PLOS ONE 9 (7), e102380. http://dx.doi.org/10.1371/journal.pone.0102380. 10.1371/journal.pone.0102380.

Hawbaker, T.J., Vanderhoof, M.K., Beal, Y.-J., Takacs, J.D., Schmidt, G.L., Falgout, J.T., Williams, B., Fairaux, N.M., Caldwell, M.K., Picotte, J.J., Howard, S.M., Stitt, S., Dwyer, J.L., 2017. Mapping burned areas using time-series of Landsat data. Remote Sens. Environ. 198, 504–522. http://dx.doi.org/10.1016/j.rse.2017.06.027.

Korontzi, S., Justice, C.O., Scholes, R., 2003. Influence of timing and spatial extent of savanna fires in southern Africa on atmospheric emissions. J. Arid Environ. 54, 395–404. http://dx.doi.org/10.1006/jare.2002.1098.

Koza, J.R., 1992. Genetic Programming – On the Programming of Computers by Means of Natural Selection. MIT Press, 813 pp.

Koza, J.R., 2010. Human-competitive results produced by genetic programming. Genet. Program Evolvable Mach. 11 (3–4), 251–284. http://dx.doi.org/10.1007/s10710-010-9112-3.

Laris, P.S., 2005. Spatiotemporal problems with detecting and mapping mosaic fire regimes with coarse-resolution satellite data in savanna environments. Remote Sens. Environ. 99, 412–424. http://dx.doi.org/10.1016/j.rse.2005.09.012.

Liu, J., Heiskanen, J., Maeda, E.E., Pellikka, P.K.E., 2018. Burned area detection based on Landsat time series in savannas of southern Burkina Faso. Int. J. Appl. Earth Observ. Geoinform. 64, 210–220. http://dx.doi.org/10.1016/j.jag.2017.09.011.

Luke, S., Panait, L., 2002. Lexicographic parsimony pressure. In: Proceedings of the Genetic and Evolutionary Computation Conference, Morgan Kaufmann Publishers Inc., pp. 829–836.

Marinho, T., 1946. Esboço do clima da Guiné Portuguesa. An. Junta Invest. Coloniais 1, 153–190.

Matricardi, E.A.T., Skole, D.L., Pedlowski, M.A., Chomentowski, W., 2013. Assessment of forest disturbances by selective logging and forest fires in the Brazilian Amazon using Landsat data. Int. J. Remote Sensing 24 (4), 1057–1086. http://dx.doi.org/10.1080/01431161.2012.717182.

McGarigal, K., Cushman, S.A., Neel, M.C., Ene, E., 2002. FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. University of Massachusetts. <http://www.umass.edu/landeco/research/fragstats/fragstats.html> (accessed 21 March 2018).

MDRARNAGB, 2009. Estratégia e plano de acção nacional para a biodiversidade. Programa das Nações Unidas para o Desenvolvimento, Projecto GBS/97/G31/1G/9, 161 pp.

Meddens, A.J.H., Kolden, C.A., Lutz, J.A., 2016. Detecting unburned areas within wildfire perimeters using Landsat and ancillary data across the northwestern United States. Remote Sens. Environ. 186, 275–285. http://dx.doi.org/10.1016/j.rse.2016.08.023.

Melchiori, A.E., Setzer, A.W., Morelli, F., Libonati, R., Cândido, P. de A., Jesús, S.C., 2014. A Landsat-TM/OLI algorithm for burned areas in the Brazilian Cerrado – preliminary results. Adv. For. Fire Res. 1302–1311. http://dx.doi.org/10.14195/978-989-26-0884-6_143.

Morton, D.C., Defries, R.S., Nagol, J., Junior, C.M.S., Kasischke, E.S., Hurtt, G.C., Dubayah, R., 2011. Mapping canopy damage from understory fires in Amazon forests using annual time series of Landsat and Modis data. Remote Sens. Environ. 115, 1706–1720. http://dx.doi.org/10.1016/j.rse.2011.03.002.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassen, K.R., 2001. Terrestrial ecoregions of the worlds: a new map of life on earth. BioSciences 51 (11), 933–938. http://dx.doi.org/10.1641/0006-3568(2001)051[0933:teotwa]2.0.co;2.

Oumar, Z., 2015. Fire scar mapping for disaster response in KwaZulu-Natal South Africa using Landsat 8 imagery. South African J. Geomat. 4 (3), 309–316 ISSN: 2225-8531.

Padilla, M., Stehman, S.V., Chuvieco, E., 2014. Validation of the 2008 MODIS-MCD45 global burned area product using stratified random sampling. Remote Sens. Environ. 144, 187–196. http://dx.doi.org/10.1016/j.rse.2014.01.008.

Pereira, J.M.C., Sá, A.C.L., Sousa, A.M.O., Silva, J.M.N., Santos, T.N., Carreiras, J.M.B., 1999. Spectral characteristics and discrimination of burnt areas. Remote sensing of large wildfires in the European Mediterranean Basin. In: Emilio Chuvieco (Ed.), Springer, pp. 123-138. http://doi.org/10.1007/978-3-642-60164-4_7

Pereira, J.M.C., Vasconcelos, M.J.P., Sousa, A.M., 2000. A rule-based system for burned area mapping in temperate and tropical regions using NOAA/AVHRR imagery. Adv. Glob. Change Res. 3, 215–232.

Poli, R., Langdon, W.B., McPhee, N.F., 2008. A Field Guide to Genetic Programming. With contributions by J. R Koza http://www.gp-field-guide.org.uk/. (Accessed 26 March 2018).

Powers, D.M.W., 2007. Evaluation: From precision, recall and f-factor to ROC, informedness & correlation. Scholl of Informatics and Engineering, Technical report SIE-07-001, Australia, 24p.

Richards, J.A., Jia, X., 2005. Remote Sensing Digital Image Analysis. An Introduction. Springer-Verlag, Berlin Heidelberg, pp. 439.

Sá, A.C.L., Pereira, J.M.C., St. Aubyn, A., 2001. Comparação entre a regressão logística e as árvores de classificação na cartografia de áreas ardidas com imagens do satélite Landsat 5 TM. In: A Estatística em Movimento: Atas do VIII Congresso Anual da Sociedade Portuguesa de Estatística, M. Neves et al. (Ed.), Lisboa, pp. 387–394.

Sá, A.C.L., Pereira, J.M.C., Vasconcelos, M.J.P., Silva, J.M.N., Ribeiro, N., Awasse, A., 2003. Assessing the feasibility of sub-pixel burned area mapping in miombo woodlands of northern Mozambique using MODIS imagery. Int. J. Remote Sens. 24 (8), 1783–1796. http://dx.doi.org/10.1080/01431160210144750.

Sertel, E., Alganci, U., 2016. Comparison of pixel and object-based classification for burned area mapping using SPOT-6 images. Geomatics, Nat. Hazards Risk 7 (4), 1198–1206. http://dx.doi.org/10.1080/19475705.2015.1050608.

Silva, J.M.N., Pereira, J.M.C., Cabral, A.I., Sá, A.C.L., Vasconcelos, M.J.P., Mota, B., Grégoire, J.-M., 2003. An estimate of the area burned in southern Africa during the 2000 dry season using SPOT-VEGETATION satellite data. J. Geophys. Res. 108 (D13), 8498–8509. http://dx.doi.org/10.1029/2002JD002320.

Silva, J.M.N., Sá, A.C.L., Pereira, J.M.C., 2005. Comparison of burned area estimates derived from SPOT-Vegetation and Landsat ETM+ data in Africa: influence of spatial

pattern and vegetation type. Remote Sens. Environ. 96, 188–201. http://dx.doi.org/10.1016/j.rse.2005.02.004.

Silva, S., Costa, E., 2009. Dynamic Limits for Bloat Control in Genetic Programming and a review of past and current bloat theories. Genet. Program Evolvable Mach. 10 (2), 141–179. http://dx.doi.org/10.1007/s10710-008-9075-9.

Silva, S., 2009. GPLAB-A Genetic Programming Toolbox for MATLAB, version 3.0. <http://gplab.sourceforge.net/> (accessed 26 March 2018).

Silva, S., Vasconcelos, M.J., Melo, J.B., 2010. Bloat Free Genetic Programming versus Classification Trees for Identification of Burned Areas in Satellite Imagery. Applications of Evolutionary Computation. Springer, pp. 272–281.

Silva, S., Vanneschi, L., 2011. The importance of being flat – studying the program length distributions of operator equalisation. In: Riolo, R., Vladislavleva, E., Moore, J.H. (Eds.), Genetic Programming Theory and Practice IX, Genetic and Evolutionary Computation. Springer, pp. 211–233. http://dx.doi.org/10.1007/978-1-4614-1770-5_12.

Silva, S., Dignum, S., Vanneschi, L., 2012. Operator equalization for bloat free genetic programming and a survey of bloat control methods. Genet. Program Evolvable Mach. 13 (2), 197–238. http://dx.doi.org/10.1007/s10710-011-9150-5.

Silva, S., Vanneschi, L., Cabral, A.I.R., Vasconcelos, M.J., 2017. A semi-supervised genetic programming method for dealing with noisy labels and hidden overfitting. Swarm Evol. Comput. 39, 323–338. http://dx.doi.org/10.1016/j.swevo.2017.11.003.

Smith, A.M.S., Drake, N.A., Wooster, M.J., Hudak, A.T., Holden, Z.A., Gibbons, C.J., 2007. Production of Landsat ETM+ reference imagery of burned areas within Southern African savannahs: comparison of methods and application to MODIS. Int. J. Remote Sensing 28 (12), 2753–2775. http://dx.doi.org/10.1080/01431160600954704.

Steinberg, D., Colla, P., 1997. CART-Classification and Regression Trees: A Supplementary Manual for Windows. Salford Systems Inc., San Diego.

Stroppiana, D., Bordogna, G., Carrara, P., Boschetti, M., Brivio, P.A., 2012. A method for extracting burned areas from Landsat TM/ETM+ images by soft aggregation of multiple spectral indices and a region growing algorithm. ISPRS J. Photogr. Remote Sensing 69, 88–102. http://dx.doi.org/10.1016/j.isprsjprs.2012.03.001.

Swain, P.H., Davis, S.M., 1978. Remote Sensing: The Quantitative Approach. McGraw Hill, New York, pp. 396.

Thariqa, P., Sitanggang, I.S., Syaufina, L., 2016. Comparative analysis of spatial decision tree algorithms for burned area of Peatland in Rokan Hilir Riau. Telkomnika 14 (2), 684–691. http://dx.doi.org/10.12928/TELKOMNIKA.v14i1.3540.

Theodoridis, S., Koutroumbas, K., 2009. Pattern Recognition. Academic Press, Elsevier, pp. 961.

USAID, 2010. Democratic Republic of Congo: Biodiversity and Tropical Forestry Assessment (118/119). United States Agency International Development, Final report, pp. 209.

Trisakti, B., Nugroho, U.C., Zubaidah, A., 2016. Technique for identifying burned vegetation area using Landsat 8 data. Int. J. Remote Sens. Earth Sci. 13 (2), 121–130.

Williams, J.E., Weele, M.V., Velthoven, P.F.J.V., Scheele, M.P., Liousse, C., Werf, G.R.V.D., 2012. The impact of uncertainties in African biomass burning emission estimates on modeling global air quality, long term range transport and tropospheric chemical lifetimes. Atmosphere 3, 132–163. http://dx.doi.org/10.3390/atmos3010132.

Verlinden, A., Laamanen, R., 2006. Long term fire scar monitoring with remote sensing in northern Namibia: relations between fire frequency, rainfall, land cover, fire management and trees. Environ. Monit. Assess. 112, 231–253. http://dx.doi.org/10.1007/s10661-006-1705-1.

Vanneschi, L., 2017. An introduction to geometric semantic genetic programming. In: Schütze, O., Trujillo, L., Legrand, P., Maldonado, Y. (Eds.), NEO 2015. Studies in Computational Intelligence. vol. 663. Springer, pp. 3–42. http://dx.doi.org/10.1007/978-3-319-44003-3_1.

Zhang, R., Qu, J.J., Liu, Y., Hao, X., Huang, C., Zhan, X., 2015. Detection of burned areas from mega-fires using daily and historical MODIS surface reflectance. Int. J. Remote Sens. 36 (4), 1167–1187. http://dx.doi.org/10.1080/01431161.2015.1007256.