remote sensing

MDPI

Article

Mapping Burned Areas in Tropical Forests Using a Novel Machine Learning Framework

Varun Mithal [1,†], Guruprasad Nayak [1,†], Ankush Khandelwal [1], Vipin Kumar [1,*], Ramakrishna Nemani [2] and Nikunj C. Oza [2]

1 Department of Computer Science and Engineering, University of Minnesota, 4-192 Keller Hall, Minneapolis, MN 55455, USA; varunmithal@gmail.com (V.M.); nayak013@umn.edu (G.N.); ankush.kwal@gmail.com (A.K.)

2 NASA Ames Research Center, Moffett Field, Mountain View, CA 94035, USA; rama.nemani@nasa.gov (R.N.); nikunj.c.oza@nasa.gov (N.C.O.)

* Correspondence: kumar001@umn.edu

† These authors contributed equally to this work.

Received: 21 November 2017 ; Accepted: 3 January 2018 ; Published: 6 January 2018

Abstract: This paper presents an application of a novel machine-learning framework on MODIS (moderate-resolution imaging spectroradiometer) data to map burned areas over tropical forests of South America and South-east Asia. The RAPT (RAre Class Prediction in the absence of True labels) framework is able to build data adaptive classification models using noisy training labels. It is particularly suitable when expert annotated training samples are difficult to obtain as in the case of wild fires in the tropics. This framework has been used to build burned area maps from MODIS surface reflectance data as features and Active Fire hotspots as training labels that are known to have high commission and omission errors due to the prevalence of cloud cover and smoke, especially in the tropics. Using the RAPT framework we report burned areas for 16 MODIS tiles from 2001 to 2014. The total burned area detected in the tropical forests of South America and South-east Asia during these years is 2,071,378 MODIS (500 m) pixels (approximately 520 K sq. km.), which is almost three times compared to the estimates from collection 5 MODIS MCD64A1 (783,468 MODIS pixels). An evaluation using Landsat-based reference burned area maps indicates that our product has an average user's accuracy of 53% and producer's accuracy of 55% while collection 5 MCD64A1 burned area product has an average user's accuracy of 61% and producer's accuracy of 27%. Our analysis also indicates that the two products can be complimentary and a combination of the two approaches is likely to provide a more comprehensive assessment of tropical fires. Finally, we have created a publicly accessible web-based viewer that helps the community to visualize the burned area maps produced using RAPT and examine various validation sources corresponding to every detected MODIS pixel.

Keywords: MODIS; burned area mapping; tropical forests; machine learning

1. Introduction

Forest fires are known to generate a significant flux of greenhouse gases and particulate matter into the atmosphere and also contribute to several ecological effects such as the loss of animal habitat and biodiversity [1]. In the tropical forests, fires are often associated with active deforestation fronts and linked to illegal establishment of industrial timber, oil palm, soy, and tea and coffee plantations [2]. Forest fire mapping from satellite data offers opportunities for providing timely information on the implementation of sustainable forest management, which is critical for making sound policy decisions for protecting forests. Furthermore, multi-annual burned area products are also needed to improve the understanding of the relationship between climate, vegetation and fires [3]. As a result, there has been

an increase in demand for automated and reliable tools to monitor forest fires from earth observing satellite data [4].

Existing satellite-based techniques for burn area assessment can be grouped into two broad categories-active fire (hotspot) detection and post-fire burned area mapping. Hotspot detection approaches use thermal energy associated with burning of biomass to map active (ongoing) fires with the purpose of real-time fire management. A number of papers have used active fire data as a proxy to report the burned area estimates [5–7]. However, hotspot detection methods are known to have a high omission error rate because they tend to miss burned pixels due to obstruction by clouds and smoke as well as due to limited satellite diurnal sampling (i.e., satellite overpass occurred when the fires were not burning) [8]. Moreover, active fire often overestimates burned area in regions with a large proportion of small, sub-pixel fires [9]. In contrast, post-fire burned area mapping techniques consider satellite observations of the land surface over a longer temporal interval around the burn date to create more reliable historical maps of burned areas [8,10–14]. Note that post-fire mapping techniques are relatively more robust to issues due to cloud cover or smoke from fires because often burn scars remain detectable in the spectral observations for several months after the burn date. This paper presents a new post-fire burned area product that is particularly relevant for tropical forests where both of these issues are very common.

Satellite-based post-fire burned area mapping algorithms face two key challenges. First, the relationship between the explanatory variables (spectral features) and target variable (burned/unburned) changes with spatial regions and land cover [8]. Therefore, training a single classification model to distinguish burned pixels from unburned pixels and applying it across different land cover classes and geographies can result in poor performance. To address this issue, existing approaches train separate customized models for each land cover and spatial region. However, this requires annotated training samples in each land cover and geographical region, which is infeasible due to the considerable human effort involved in collecting training samples using ground and aerial surveys. To address the challenge of lack of annotated training data for different spatial regions and land classes, previous approaches such as [8,9] have used Active Fire hotspots for selecting training samples. However, active fire hotspots are only imperfect surrogates for burned areas; therefore, these previous studies also used hand-crafted "cleaning" rules while selecting training pixels from active fire hotspots to ensure that the training samples are accurate. While such "cleaning heuristics" are helpful in improving the performance of classification models, the spectral diversity of burned pixels can make such "cleaning heuristics" very brittle in some regions and land classes.

The second challenge is that the unburned locations considerably outnumber the burned locations. Hence even with a small false positive rate, a traditional classification algorithm can result in a significant number of spurious burned areas. This issue has been addressed by previous approaches using the concept of identifying confident burns by using co-occurrence of active fire hotspots and burn scars in spectral observations [8]. Moreover, since landcover changes can occur due to other reasons such as logging, requiring the presence of active fire hotspots ensure that landcover changes other than fire are excluded.

This paper presents an application of a novel machine-learning framework to map burned areas over tropical forests of South America and South-east Asia. This 3-stage framework RAPT (RAre Class Prediction in the absence of True labels) [15] is able to build data adaptive classification models using noisy training labels (See Section 3.1 for details). The RAPT framework was used to build a burned area product derived from MODIS Surface Reflectance 8-day composite product MOD09A1 [16] and MODIS active fire product MOD14A2 [17] for the tropical forests in Amazon and South-east Asia. An evaluation was performed using Landsat-based reference burned area maps that were constructed by following the procedure in [18] (see Section 3.2 for details).

Finally, we have created a publicly accessible web-based viewer for visualizing our burned area product http://z.umn.edu/fireviewer/. The viewer shows the RAPT events corresponding to a user-selected MODIS tile and year as event polygons. For each polygon, the Landsat composites

before and after the date of event can be queried. Polygon-level statistics such as total number of MODIS pixels, number of pixels with an active fire hotspot, and the number of pixels identified by collection 5 MCD64A1 are reported. The MODIS pixels belonging to a particular event polygon can also be viewed by selecting the polygon. In addition, relevant information about these MODIS pixels (such as their Enhanced Vegetation index series, Normalized Burn Ratio index series, and MODIS land cover classification labels) can be queried by selecting (clicking) one of the pixels.

2. Material: Study Area and Input Data

This study considers the burned areas in forests located in tropical regions in South America and South-east Asia. We do not consider the African tropics in this study since the forest fires in that region mainly occur at a subpixel level in forested pixels that border other landcover classes such as grasslands and savannas, making it difficult to build a high accuracy classifier for them using 500 m MODIS data. The burned area detection algorithm uses the 500 m MODIS Surface Reflectance 8-day composite product MOD09A1 [16,19], MODIS active fire product MOD14A2 [17,19], and MODIS MOD12Q1 land cover classification product [19,20]. The MODIS products are defined on global sinusoidal grids in fixed geolocated tiles approximately 10 degree by 10 degree in size and are publicly available from Land Processes Distributed Active Archive Center. The MODIS products used in this study are from the MODIS 5 collection. Moreover, collection 5 MODIS MCD64A1 [8,19] is used as an existing state-of-art burned area product for comparative study. Here we always refer to collection 5 MODIS MCD64A1 as MCD64 product.

The RAPT framework is applied to the forested locations in 16 MODIS tiles in South America and South-east Asia from 2001 to 2014. Figure 1 shows the MODIS tiles covered in the region of study.

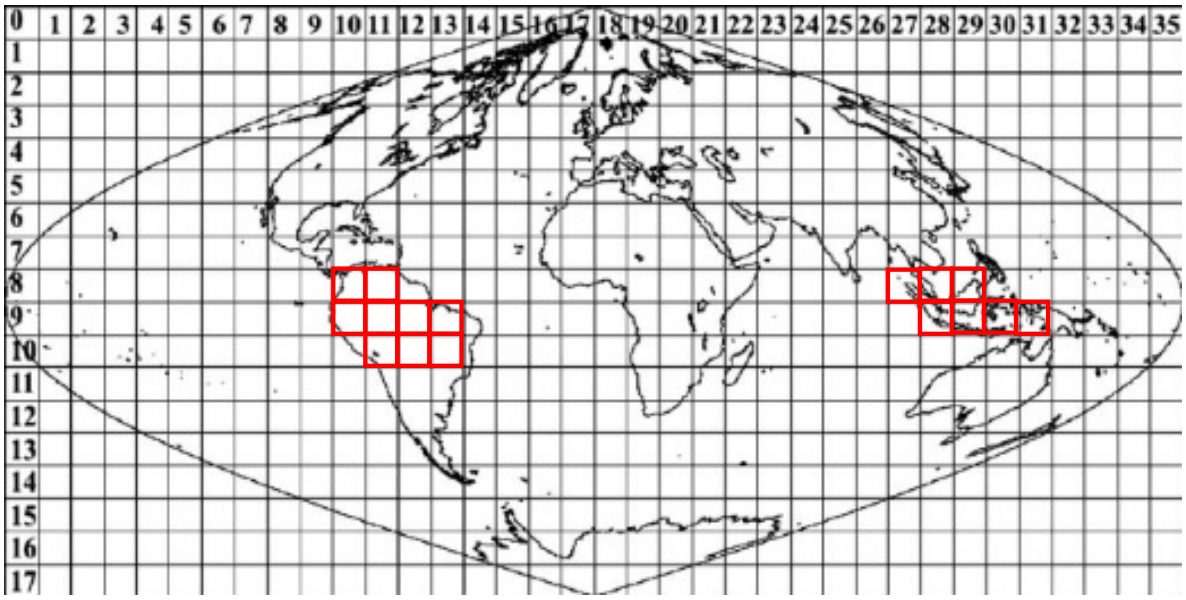


Figure 1. Figure shows the MODIS tiles where we report burned areas.

3. Methods: RAPT Framework and Validation Procedure

3.1. RAPT Framework

Our burned area detection framework makes use of the information in MODIS active fire hotspots and multispectral surface reflectance data for global-scale burned area detection. There are two key innovations used in our algorithm that distinguish it from previous approaches. First, the algorithm uses a new training procedure [15] that is robust to presence of noise in the labels of training samples, as a result of which our approach does not rely on "cleaning heuristics". Second, the classifier used in

our approach [21] takes into account the yearly context of the burn scar signature instead of just a per time step signature. The longer temporal context helps improve the user's accuracy of the detected burned area while maintaining a higher coverage. The detection algorithm proceeds through the three stages described below (Figure 2 shows flowchart of the RAPT framework).

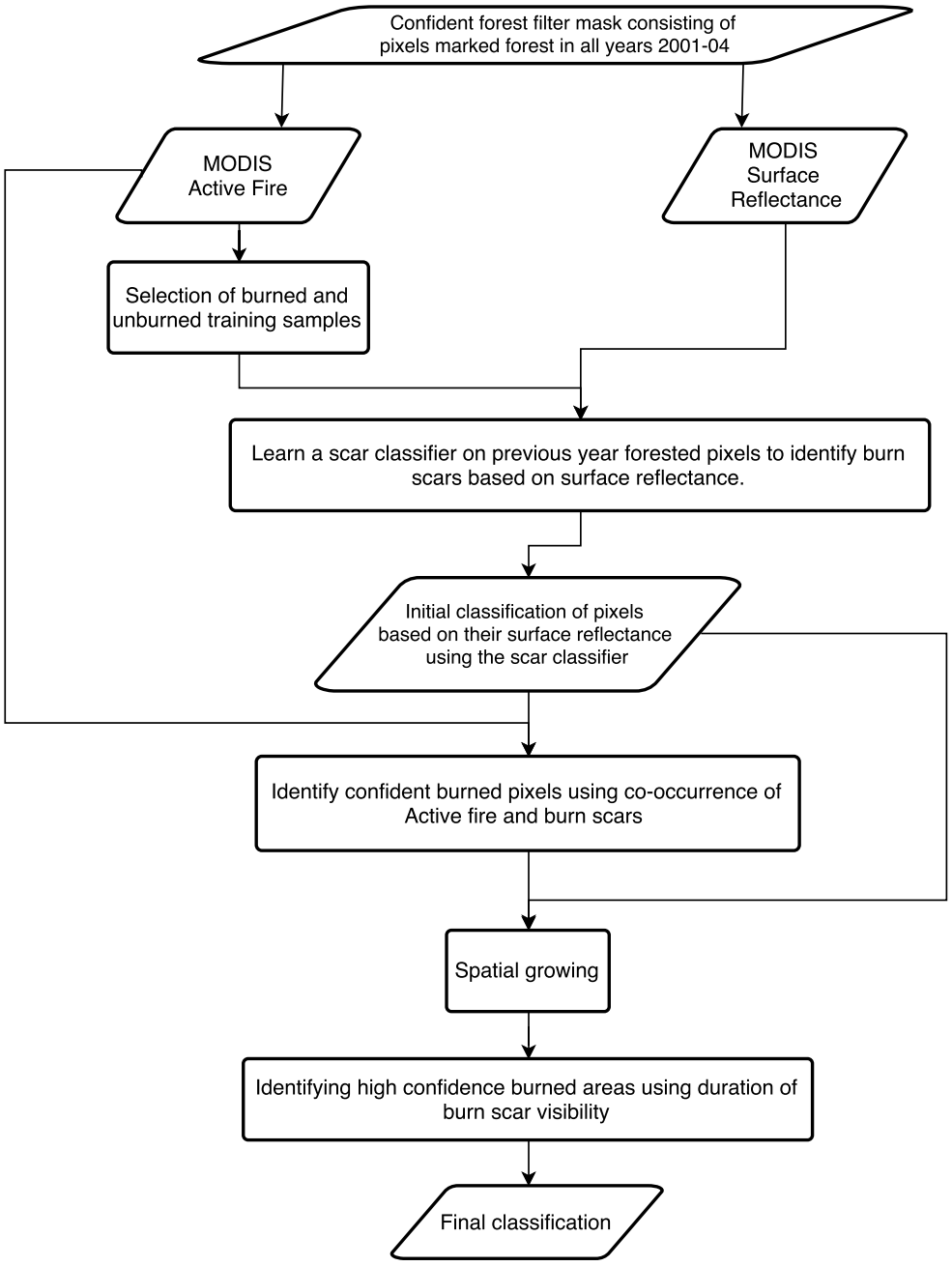


Figure 2. Flowchart showing the steps of the RAPT framework.

3.1.1. Stage 1: Identifying Burn-Scars from Spectral Data

In this step, a classification model is trained to identify the burned pixels using the 7-dimensional feature space consisting of all the seven bands of MODIS multispectral reflectance data. Active fire hotspots were used as training labels to learn the classification model. Specifically, 20,000 unburned (with no Active Fire) and 20,000 pixels with Active fire are used as training instances for the classifier (in case 20,000 pixels with Active Fire are not available, we take however many such samples are

available). This classifier is then applied on all pixels to produce the initial classification for each 500 m pixel annually. Burned pixels identified by this classifier are treated as candidate burned pixels, which are then refined in the subsequent stages to give the final burned areas.

Selecting training pixels: We make use of the active fire hotspots to select the training samples for burned and unburned classes to train the scar classifier. In particular, we select positive training samples for burned class from pixels with an active fire hotspot, if they form a spatial cluster of size greater than 10 pixels. This spatial pruning is done to eliminate active fire pixels belonging to sub-pixel burns. Negative training samples (unburned class) are selected from pixels that do not have an active fire hotspot. Pixels with an active fire observation on a given date are more likely to show a burn scar than those without an active fire. However, active fire hotspots are only a weak surrogate for burn-scars and a training data set created using active fire observations is often contaminated with noise in the labels of the training samples.

Computing posterior class probabilities: We use a neural network model [21] as our classifier to model the probability that a given pixel belongs to the burned class. The model takes the spectral features corresponding to the 46 time steps present in a year as input and outputs a probability that the location belongs to the burned class for the year. The design of this neural network model is motivated by the following two characteristics of fire events. First, a burn event leads to a burn scar that is often distinguishable in the spectral features. Second, the burn events (and hence the scars) follow a temporal context and occur in a fire season that is specific to each geography and land cover. Our neural network model SeqRep makes use of these characteristics of fire events to identify burned pixels using multispectral observations.

Figure 3 shows the design of the neural network. The network assumes that the input data is observed over $T = 46$ time steps and has $D = 7$ dimensional feature vector corresponding to every time step. The input layer of the neural network takes has $T \times (D + 1)$ nodes in the first layer, corresponding to the D features and a bias term for each time step. The model has a hidden layer with T nodes $\{h_1, \cdots, h_T\}$. Each one of these hidden nodes corresponds to a time step. Unlike a regular neural network with one hidden layer, the hidden layer nodes h_t corresponding to time step t is connected only to input (first layer) nodes for that time step. Thus, there is a correspondence between every hidden layer node and a time step. Moreover, the weights $\{\beta_0, \beta_1, \cdots, \beta_D\}$ connecting the input layer node at a time step to the corresponding hidden layer node are assumed to be shared across time steps. Finally, there is an output node O that is a function of all the per time step predictions captured in the T nodes in the hidden layer. Weights $\{w_0, w_1, \cdots w_T\}$ connect the T hidden layer nodes along with a bias term to the output node O. Thus, the neural network has $(T + D + 2) = 55$ weights in all. The hidden layer nodes can be viewed as outputs from a classifier that maps the features at a time step (e.g., reflectance values) to the per time step predictions (e.g., How scarred does the location look on this time step?). The vector of these per time step predictions are then combined to output the prediction for the whole time series (e.g., does the location look scarred at the right time steps?). A back propagation algorithm is used to select the model parameters of the neural network model. The parameter update steps for training this model are provided in the Appendix A.

Note that since active fire typically precedes the date when burn scar is visible at a location, using it as training label only informs the classifier of the presence of a burn scar at the location at some time during the year (and not the exact date of it). SeqRep jointly learns the spectral signature of the burn scar and the timing of the fire season for the region using the MODIS multispectral data and the active fire label for the year provided as input. Thus, although the classifier models fire seasons, it does not require a prior knowledge of the timing of the fire season and hence, can be easily applied to new regions and landcovers.

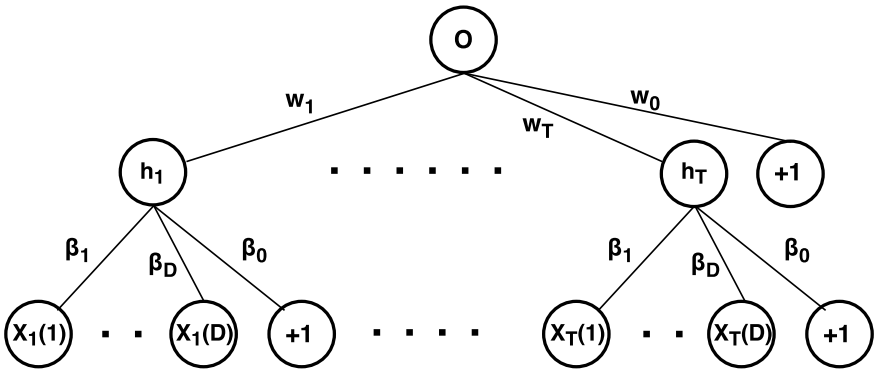


Figure 3. SeqRep model.

Fixing decision threshold: The final classifier is written in the following form:

$$y = \begin{cases} 1 & \text{if } P[y = 1|\mathbf{x}; \mathbf{w}] \geq \gamma \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{x}$ is the spectral feature vector, $\mathbf{w}$ is the classification model parameter and y is the target class (burned/unburned). The threshold γ determines the trade-off between the errors of omission and commission, which varies as γ is swept from 0 to 1. In case of burned area mapping a suitable classification objective is to jointly maximize the user's accuracy and producer's accuracy of the burned class. Therefore, in the RAPT framework we select γ to maximize the product of user's accuracy and producer's accuracy of the burned class. Note that selecting the decision threshold γ by maximizing the product of the user's and producer's accuracy ensures that the errors of omission and commission are relatively balanced. The selection of decision threshold γ is a challenge in our problem setting because the training data created using active fire hotspots is plagued with noise in training labels, therefore estimation of user's and producer's accuracy corresponding to different choices of γ is tricky. In particular, if we ignore the noise in labels of training samples, it may lead to incorrect estimation of user's and producer's accuracy resulting in selection of a sub-optimal threshold value, due to which the classification model will show higher rates of omission and commission compared to the optimal model. In our detection method, we use the algorithm developed in [15] that addresses the problem of training classification models when training samples suffer from label noise in the context of imbalanced class problems.

3.1.2. Stage 2: Identifying Confident Burned Pixels

In the second stage, the predictions of the scar classifier are combined with the active fire hotspots to assign either confident burn or unlabeled class to every pixel in each year. We use a conservative combination step that assigns a confident burn only if both the burn-scar and active fire hotspot are present at the pixel for the given year. This strict criterion ensures that the confident burn pixels have a smaller number of commission errors compared to both the scar classifier and active fire hotspots individually.

3.1.3. Stage 3: Spatial Growing and Final Classification

The confident burned areas, identified in the previous step, typically cover only a small fraction of the total burned pixels. This is especially true in the tropics where active fire hotspots have a low coverage due to poor data quality (e.g., obstruction by clouds) and limited frequency of satellite overpass. Therefore, in the third step of RAPT we improve the coverage of burned areas by leveraging the spatial context of confident burned pixels and candidate burned pixels. More specifically, a spatial growing method is used, which includes the pixels that show a burn-scar but are not detected as confident burns as part of final burned areas if they are spatially connected to some confident burned pixels within a spatial distance of 5 MODIS pixels. This is a manually selected threshold, but the

performance is not very sensitive to it. In fact, our experimental results indicated that if we choose the distance threshold between 5 to 10 pixels, the results remained unchanged for most tiles and years. However, having this threshold as low as 5 is helpful in ensuring that large spatial regions incorrectly classified as burned in Stage 1 do not get classified as burned at the end of Stage 3 just due to a spurious active fire hotspot that coincides with the onset of event in Stage 1.

3.1.4. Generating Global Product: Additional Data Processing Considerations

To account for the spectral diversity of burn scars, we train multiple classification models, each focusing on a smaller, homogeneous partition of the data, grouped according to geographical region. Training a separate, customized classifier for each homogeneous group is known to improve the burned area detection performance, as it allows for automated adaptation of model parameters to the specifics of fire occurrence in the biome and spatial region of interest [8]. Specifically, in our detection framework, we create separate groups corresponding to each MODIS tile. Moreover, burned area results are reported for only stable forest pixels. A pixel is considered to belong to stable forest in a particular year if it is labeled as forest by MODIS land cover in the previous year as well as in the first 4 years (i.e., 2001–2004).

3.2. Validation Procedure

We evaluated the RAPT burned area product by comparing it against the reference validation maps produced from Landsat images using publicly available BAMS software developed by [10]. We used Landsat 5 images obtained from (https://earthexplorer.usgs.gov/) to generate reference maps used for validation. In this subsection, we present the sampling design used for selecting Landsat scenes for evaluation, the procedure for generating the reference maps for the selected Landsat scene and the accuracy measures used.

3.2.1. Sampling Design

Burned areas are only a small portion of the forested regions, and as a consequence a simple random sampling is not effective when the focus is on validating burned areas. Therefore, we used a proportional sampling scheme similar to the one used in [18], in which the probability of selecting a Landsat scene is proportional to the number of active fire hotspots in the forested regions around the date of that scene. This biases the sampling towards Landsat scenes with higher burned activity, and therefore is helpful in obtaining sufficient fire events in the reference maps generated to robustly estimate omission errors of the burned area products being evaluated.

Since the selection of Landsat scenes is driven by the number of active fire hotspots in the scene in a given year, if a region has seen multiple fires across years then the corresponding Landsat scene can get selected multiple time in our 15 year period. For example, in our set of reference maps in Table 1, scenes 118,062 and 213,070 were selected in multiple years.

Once a Landsat scene was sampled, the date of the fire event was estimated by computing the median of active fire hotspot dates observed in that scene for that year. Finally, pre-event and post-event images for this scene were selected around the date of the fire event according to the following criteria:

1. The images should have low scene level cloud cover fraction according to Landsat metadata (less than 25%).
2. Post-event image should be within 100 days of the date of burn. However, if the selected post-event Landsat image has more than 25% total cloud cover, then the Landsat scene was excluded for that year.
3. Pre-event image should be within 100 days of the date of burn. However, if the selected image has more than 25% cloud cover, then we looked for a clean image in the previous year in the same season. If a Landsat image with less than 25% cloud cover can still be not found, then we exclude the scene for this year.

Note that while comparing the burned area output from the algorithms with the reference maps, only the burned pixels reported between the dates of the pre-event and post-event images are considered as burned. For example, if an unburned pixel according to the reference map is identified by the algorithm outside these dates, it is not considered a commission error. Pre-event and post-event images corresponding to all the reference maps used in study have been added in the supplementary material. Along with these images, we also show pre-event images that are strictly within 100 days of the post image (although cloudy and hence are not suitable for creating reference maps) to provide visual verification that the post-images are within 100 days of the fire event.

3.2.2. Cloud Mask

We used an object based cloud detection method described in [22,23] to mask any pixels impacted by clouds and cloud shadows. For Landsat L1T images, Digital Number (DN) values were first converted to Top of Atmosphere (TOA) reflectance and Brightness Temperature (Celsius degree). Then, multispectral bands, Brightness Temperature and derived indices (NDVI, NDSI) were used to create a potential cloud layer and a potential cloud shadow layer by creating rules based on cloud and cloud shadow physical properties. Cloud layer is then segmented and the geometric relationships are used to match the potential cloud shadow layer, to create the final cloud and cloud shadow mask.

3.2.3. Reference Data

The pre-event and post-event images were first processed using BAMS to obtain a feature set for all pixels. BAMS uses known indices described in literature for burn scar detection. Specifically, BAMS code makes use of 10 spectral variables (the post-fire BAIM (Burned Area Index), NBR (Normalized Burned Ratio), MIRBI (Mid Infrared Burned Index), GEMI (Global Environment Monitoring Index), and NDVI (Normalized Difference Vegetation Index) indexes, plus the temporal differences of those five indexes) [10]. Then several burned pixels were selected by a manual annotator from the Landsat images that were used by BAMS software as training samples for building a classification model. BAMS then applied this classification model to generate the Landsat-based burned areas for the scene, which were used as fire polygons to serve as the reference maps. The performance of BAMS can get impacted by presence of clouds. Hence a cloud detection method described in Section 3.2.2 was used to remove pixels that are affected by clouds, cloud shadows and haze. After excluding invalid pixels, a reference map was considered only if it had sufficiently many valid pixels in order to ensure robust estimation of performance measures. In our experiment, we excluded reference maps that had less than 1000 valid pixels at MODIS scale (each Landsat image has approximately 130,000 MODIS pixels).

As MODIS and Landsat data are at different scales, a mapping between MODIS and Landsat pixels was derived, and then used to assign a fraction burned to each MODIS pixel based on the number of Landsat pixels in it that are reported as burned in the reference maps. The error matrix is completed using these fractions. In particular, for every MODIS pixel flagged as burned by our algorithm, we add the fraction of Landsat pixels belonging to this pixel that are assigned to burn class by BAMS to the true positives and the remaining fraction to the false positives. Similarly, for every MODIS pixel flagged as unburned by our algorithm, we add the fraction of Landsat pixels belonging to this pixel that are assigned to burn class by BAMS to the false negatives and the remaining fraction to the true negatives.

3.2.4. Accuracy Measures

Several measures have been used in previous studies to quantify accuracy of burned area products. The choice of the evaluation measure is guided by the objectives of the end user. In our study, we followed the commonly used approach that uses reference maps to generate pixel-level error matrices for the burned area product. A number of accuracy measures can be derived from error matrix. For skewed class problems such as burned area mapping, it is common to assess the quality of a classification product in terms of the user's accuracy and producer's accuracy. User's accuracy

indicates the probability that a pixel reported as burned is actually burned. Producer's accuracy indicates the probability that an actually burned pixel is reported by the detection algorithm. Note that there is a trade-off between user's and producer's accuracy, as one measure can be made extremely high at the cost of the other. Therefore, often combined measures are used to combine information in user's and producer's accuracy into a single measure for comparing multiple burned area products. One of the commonly used measure is the Dice coefficient (DC). DC is defined as $\frac{2*TP}{(2*TP+FP+FN)}$, where TP is number of true positives, FP is number of false positives and FN is number of false negatives in the error matrix. In this study, we report the user's accuracy, producer's accuracy and Dice coefficient for RAPT and MCD64 computed over the study region using the Landsat-generated reference maps.

4. Results: Validation Using Landsat-Based Reference Maps and Comparison with MCD64 Product

In this section we validate the burned area maps generated using the RAPT algorithm. For reference, we also compare our detection with the MODIS product MCD64. The MCD64 product is the primary source of the Global Fire Emissions Database (GFED) Version 4 [24]. In [25], the authors did a comparative evaluation of 6 burned area products including MCD64 [8], MCD45 [26], Geoland2 [27] and 3 products from the European Space Agency's (ESA) Climate Change Initiative (CCI) program [28]. They concluded that MCD64 was the most accurate burned area product. Hence, in this study, we provide a comparison against MCD64 product to establish the utility of our product.

For the ease of comparison we scaled RAPT and MCD64 to annual products, i.e., every year each 500 m pixel in the product was assigned to either burned or unburned class depending on whether it is flagged as burned in at least one of the dates in the corresponding year. On comparing the two products for a given year, each pixel belongs to one of the following categories-reported as burned by both products, reported as burned only by RAPT and unburned by MCD64, reported as burned only by MCD64 and unburned by RAPT, or reported as unburned by both products. Table 2 reports the number of burned pixels belonging to each of the three categories- RAPT (only), Common (i.e., both RAPT and MCD64), and MCD64 (only) aggregated over the 14 years. We observe that RAPT identifies about 60% of burned areas reported by MCD64. However, more importantly, RAPT identifies 2,071,378 MODIS pixels (approximately 520 K sq. km.), which is about three times as many burned areas compared to MCD64. We present the annual burned area reported by RAPT, MCD64 and Active fire for each MODIS tile in the supplementary section.

Results

Table 1 shows the user's accuracy, producer's accuracy and Dice coefficient for RAPT and MCD64 burned area products computed using the Landsat reference maps for 19 Landsat scenes. The average user's accuracy for both RAPT and MCD64 product is greater than 50% (average for RAPT is 53% and for MCD64 is 61%). However, the average producer's accuracy for RAPT is about 55%, while for MCD64 it is 27%. It is clear that RAPT has significantly higher producer's accuracy compared with MCD64 for these scenes. A somewhat higher user's accuracy for MCD64 suggests that MCD64 identifies strong burns while RAPT identifies weaker burn events as well. Note that both RAPT and MCD64 products have a producer's accuracy of 0 for two scenes (ID 11 and ID 18), i.e., they did not identify any burned pixels in these scenes. This is because the aforementioned scenes have only a small number of burned pixels according to the Landsat reference maps. Also note that for these scenes, user's accuracy is undefined because when a method does not detect any burned pixel, it has both 0 true positives as well as 0 false positives. In the following section, we examine the Landsat images and reference maps in context of the errors of omission and commission to understand the nature of errors made by RAPT algorithm.

Table 1. Landsat validation results for 19 Landsat-based reference maps. The number of burned and unburned pixels are at 500 m MODIS scale.

ID	Tile	Landsat Scene	Pre-Date	Post-Date	Cloud Cover	User's Acc. RAPT	Prod's Acc. RAPT	Dice Co-ef. RAPT	Overall Acc. RAPT	User's Acc. MCD	Prod's Acc. MCD	Dice Co-ef. MCD	Overall Acc. MCD	Number of Burned Pixels	Number of Unburned Pixels
1	h11v09	230066	18 June 2010	8 October 2010	2	0.63	0.63	0.63	0.99	0.86	0.37	0.52	0.99	359	34,408
2	h12v09	230066	18 June 2010	8 October 2010	2	0.63	0.44	0.52	0.99	0.77	0.02	0.03	0.99	855	103,964
3	h11v10	233068	7 June 2010	27 September 2010	28	0.61	0.29	0.4	0.93	0.91	0.16	0.27	0.93	5941	69,437
4	h12v09	227065	12 June 2004	2 October 2004	28	0.79	0.6	0.68	0.98	0.87	0	0.01	0.96	3646	92,248
5	h12v10	229070	26 June 2004	16 October 2004	12	0.78	0.26	0.39	0.95	0.83	0.35	0.49	0.95	3534	52,315
6	h11v10	232070	16 June 2010	6 October 2010	9	0.29	0.74	0.42	0.85	0.34	0.33	0.33	0.90	3606	46,017
7	h29v09	118062	17 June 2004	21 September 2004	26	0.35	0.57	0.44	0.97	0.41	0.63	0.5	0.97	1110	53,239
8	h29v09	118062	14 May 2009	18 August 2009	41	0.32	0.3	0.31	0.99	0.24	0.26	0.25	0.99	329	35,745
9	h11v10	231070	13 July 2005	1 October 2005	74	0.88	0.61	0.72	0.99	0.87	0.42	0.56	0.99	652	29,259
10	h11v10	231070	25 June 2010	2 December 2010	83	0.18	0.99	0.31	0.94	0.18	0.46	0.26	0.96	198	14,056
11	h12v09	228066	6 June 2005	12 October 2005	85	NaN	0	NaN	1	NaN	0	NaN	1	4	17,616
12	h11v10	002067	8 September 2004	13 October 2005	29	0.7	0.31	0.43	0.92	0.79	0.04	0.07	0.90	4391	38,524
13	h11v09	002067	8 September 2004	13 October 2005	4	0.58	0.41	0.48	0.89	0.78	0.12	0.2	0.88	5445	37,084
14	h11v10	232068	4 July 2005	9 November 2005	31	0.28	0.84	0.42	0.97	0.21	0.09	0.13	0.99	854	76,257
15	h11v09	002066	9 July 2005	13 October 2005	4	0.43	0.57	0.49	0.97	0.59	0.15	0.24	0.98	3368	129,531
16	h12v09	225066	28 April 2010	21 October 2010	20	0.56	0.68	0.62	0.98	0.54	0.49	0.52	0.98	2441	101,855
17	h30v09	110062	11 July 2001	18 October 2002	40	0.42	0.67	0.51	0.91	0.5	0.61	0.55	0.93	1245	16,808
18	h10v09	006066	31 May 2004	23 September 2005	74	NaN	0	NaN	1	NaN	0	NaN	1	93	31,097
19	h11v08	232058	4 April 2006	20 March 2007	13	0.49	0.41	0.45	0.95	0.62	0.06	0.1	0.95	2046	37,003

Table 2. Table reports the number of burned pixels (at 500 m. spatial resolution) corresponding to RAPT (only), Common and MCD64 (only) for each tile in the region of study.

MODIS Tile	RAPT Only	Common	MCD64 Only
h10v08	13,040	326	3344
h11v08	51,858	7744	6011
h10v09	23,686	695	1299
h11v09	194,456	40,055	16,976
h12v09	266,621	107,571	35,199
h13v09	85,442	11,139	5405
h11v10	342,575	109,382	50,166
h12v10	254,563	116,862	97,935
h13v10	6738	4607	2514
h27v08	12,170	1373	3256
h28v08	121,498	20,450	11,927
h29v08	38,930	1749	10,859
h28v09	76,559	13,479	22,401
h29v09	94,143	39,015	32,004
h30v09	11,532	1493	1155
h31v09	1469	158	6919
Total	1,595,280	476,098	307,370

5. Discussion of Results

5.1. Reasons for Higher Producer's Accuracy of RAPT Compared to MCD64

The higher producer's accuracy of RAPT relative to MCD64 can be due to the following differences between the two products in their algorithm characteristics-

1. MCD64 requires manual selection of parameters-MCD64 uses a number of hand-crafted rules and parameters during pre-processing to select good quality training samples (for example, an active fire hotspot is included as burned training sample only if the loss in vegetation exceeds a manually fixed threshold), and during classification to detect high quality (precise) burned area (for example, a manually fixed threshold on posterior probability is used to identify confident burned locations). Since the values of these thresholds are not selected in a data-adaptive fashion, the choice of these parameters can be too conservative for the tropics.
2. RAPT balances omission and commission errors-The RAPT framework does a principled selection of decision threshold on the class probability to maximize the product of producer's and user's accuracy of the burned area product. This improves the detection performance compared to the previous approaches that relied on a manually fixed threshold that (i) requires human effort and (ii) may not jointly optimize user's and producer's accuracy, i.e., increase one at the expense of the other.
3. RAPT makes use of a seasonality-aware classifier-The classifier used in RAPT approach takes into account the yearly context of the burn scar signature instead of just a per time step signature. The longer temporal context helps improve the precision of the detected burned area while maintaining a higher coverage. Moreover, use of a longer temporal context to identify fire scars makes RAPT more robust to poor data quality, as scars often remain detectable for multiple months. On the other hand, MCD64 misses burned locations due to lack of good quality reflectance data around the time of burn event. This problem is particularly relevant for tropical areas with high average percent cloud cover and where burning results in the release of large quantities of particulate matter into the atmosphere. MCD64 product declares some pixels on certain dates to be unburned because the data quality on those pixels in a 20 day neighborhood was too poor to make predictions. The product includes 5 bits of quality flags at monthly scale for each pixel in the MODIS scene. Specifically, when bit 1 takes a value of 1, it indicates that there was sufficient valid data in the reflectance time series for the grid cell to be processed by MCD64. We consider this bit in MCD64 for 3 time steps- the time step corresponding to the only RAPT

burned event for the pixel and the next two time steps. In Table 3 we report the fraction of only RAPT burned pixels that have 0, 1, 2 or 3 time steps with good quality flags according to the bit 1 in MCD64 product in each MODIS tile. We observe that about 19% of only RAPT burned pixels have 0 good quality time step around the burn date, and about 20% of only RAPT burned pixels have only 1 good quality time step. Similarly, in Table 4 we report the fraction of MCD64 burned pixels that have 0, 1, 2 or 3 time steps with good quality flags according to the bit 1 in MCD64 product in each MODIS tile. We observe that about 88% of MCD64 burned pixels have 2 or more good quality time step around the burn date. Thus, overall, 39% of all locations detected by only the RAPT algorithm have poor quality data (0 or only 1 good quality time step). On the other hand, only 12% all the locations detected by MCD64 have poor quality data, which suggests that quality plays a role in the performance of MCD64.

Table 3. Table shows the fraction of Only RAPT burned pixels with 0, 1, 2 or 3 good quality time steps out of the 3 time steps from date of burn event set.

Tile	RAPT Only	Fraction of Locations with 0 Good Quality Timesteps	Fraction of Locations with 1 Good Quality Timesteps	Fraction of Locations with 2 Good Quality Timesteps	Fraction of Locations with 3 Good Quality Timesteps
h10v08	13,040	0.46	0.26	0.10	0.18
h11v08	51,858	0.06	0.25	0.13	0.57
h10v09	23,686	0.21	0.23	0.17	0.39
h11v09	194,456	0.07	0.12	0.23	0.58
h12v09	266,621	0.26	0.34	0.22	0.19
h13v09	85,442	0.51	0.20	0.13	0.15
h11v10	342,575	0.19	0.25	0.15	0.41
h12v10	254,563	0.06	0.08	0.15	0.71
h13v10	6738	0.07	0.20	0.17	0.57
h27v08	12,170	0.06	0.08	0.13	0.73
h28v08	121,498	0.23	0.17	0.20	0.41
h29v08	38,930	0.11	0.12	0.18	0.59
h28v09	76,559	0.25	0.21	0.20	0.33
h29v09	94,143	0.31	0.25	0.17	0.27
h30v09	11,532	0.00	0.03	0.07	0.90
h31v09	1469	0.28	0.16	0.20	0.36
Total	1,595,280	0.1876	0.2045	0.1789	0.4290

Table 4. Table shows the fraction of MCD64 burned pixels with 0, 1, 2 or 3 good quality time steps out of the 3 time steps from date of burn event set.

Tile	MCD64	Fraction of Locations with 0 Good Quality Timesteps	Fraction of Locations with 1 Good Quality Timesteps	Fraction of Locations with 2 Good Quality Timesteps	Fraction of Locations with 3 Good Quality Timesteps
h10v08	3670	0.00	0.20	0.25	0.55
h11v08	13,755	0.00	0.14	0.16	0.71
h10v09	1994	0.00	0.07	0.33	0.60
h11v09	57,031	0.00	0.07	0.24	0.69
h12v09	142,770	0.00	0.16	0.41	0.43
h13v09	16,544	0.00	0.18	0.29	0.54
h11v10	159,548	0.00	0.10	0.21	0.68
h12v10	214,797	0.00	0.07	0.17	0.76
h13v10	7121	0.00	0.23	0.12	0.66
h27v08	4629	0.00	0.10	0.17	0.74
h28v08	32,377	0.00	0.13	0.21	0.66
h29v08	12,608	0.00	0.05	0.11	0.84
h28v09	35,880	0.00	0.27	0.30	0.42
h29v09	71,019	0.00	0.20	0.23	0.56
h30v09	2648	0.00	0.05	0.12	0.82
h31v09	7077	0.00	0.05	0.14	0.81
Total	783,468	0	0.12	0.24	0.64

To illustrate this further, Figure 4 shows a region in Indonesia where MCD64 missed several burned pixels due to poor data quality around burn date. Figure 4d shows a burned area map with only RAPT, common and only MCD64 detection. We notice that this small region has mostly only RAPT burned pixels (i.e., MCD64 has very little detection here). The burn scar corresponding to the region detected by RAPT can be seen from Landsat images in Figure 4b,c, thus suggesting presence of a burn event. Figure 4e shows a color coded map for quality bit according to MCD64 over these RAPT events, and it is clear from this figure that MCD64 considered the data quality to be bad around the date of these RAPT events.

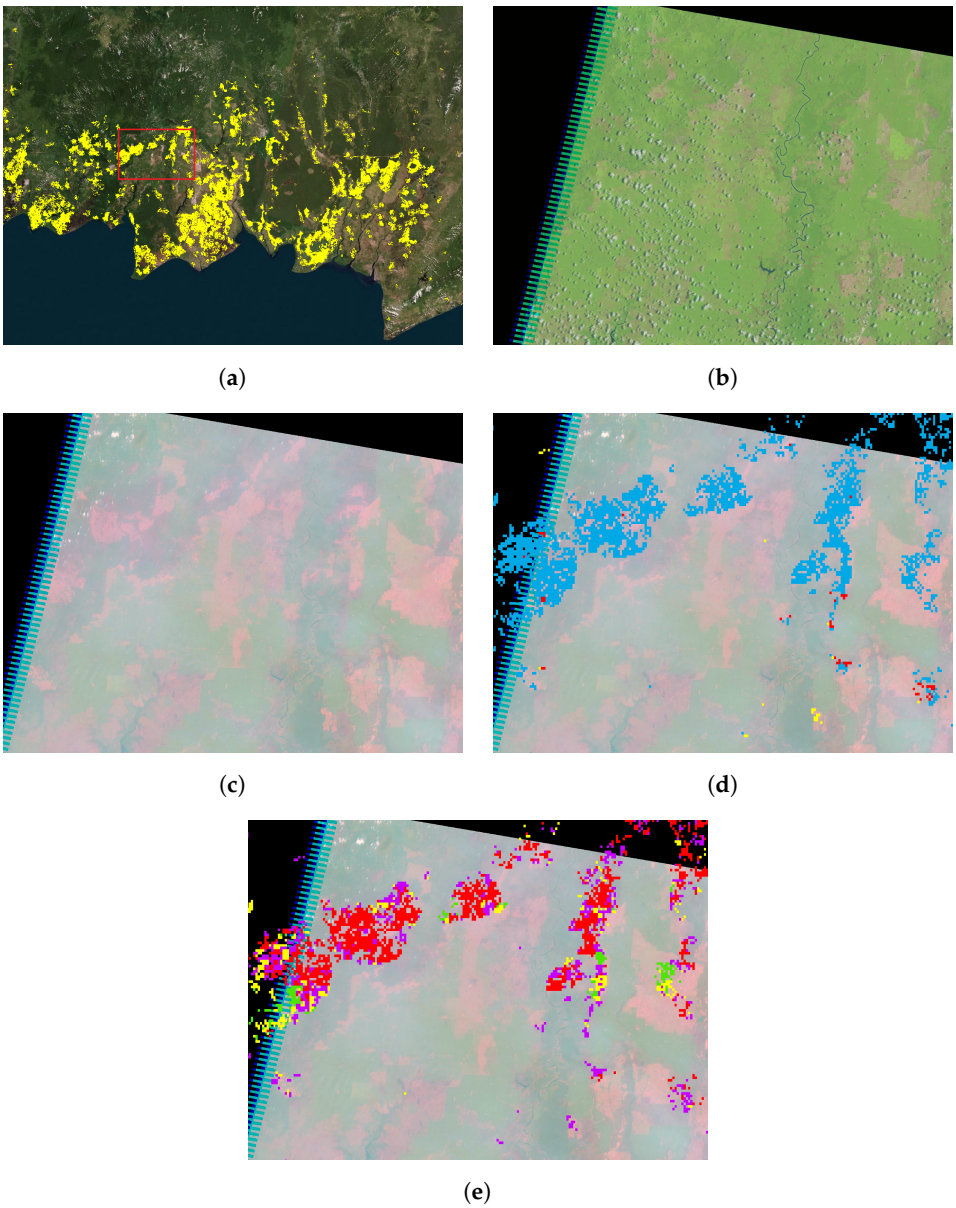
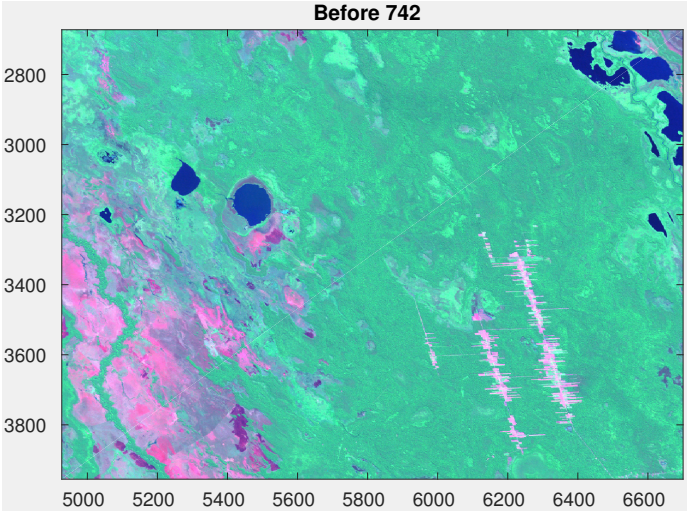


Figure 4. Figure shows a region in Indonesia where RAPT identified several burned pixels missing in MCD64. Figure also shows the data quality as provided by MCD64 product around the burn date for the identified burned pixels. (**a**) Burned area (in yellow) in a region of SE Asia. The red rectangle is the focus area for this figure; (**b**) Before burn Landsat 5 image on 10 March 2006; (**c**) After burn Landsat 5 image on 5 November 2006; (**d**) Burned area comparison map (*blue* is only RAPT, *red* is common, *yellow* is only MCD64); (**e**) Color coded quality map according to MCD64 bit 1 flag (*red* is 0/3 time steps with good quality, *purple* is 1/3 time steps with good quality, *yellow* is 2/3 time steps with good quality, *green* is 3/3 time steps with good quality).

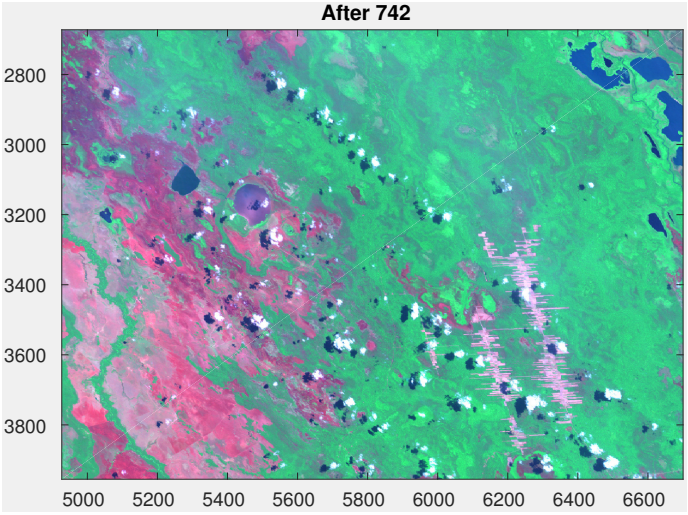
5.2. Sources of Errors in the RAPT Product

5.2.1. Errors of Omission

1. For regions where RAPT is missing large patches of burned events that are present in the reference maps, the Landsat image composites often show a burn scar that is quite diffused and lighter in intensity (see Figures 5 and 6). These lighter intensity burn scars are most likely due to burns for which the vegetation quickly recovered in a few months. The RAPT algorithm tends to miss fires that have a fast vegetation recovery, and therefore quick recovering fire events correspond to the major source of omission errors in the RAPT product.
2. The RAPT algorithm trains a separate classification model for each MODIS tile to account for the geographical variations in the spectral characteristics of burned and unburned pixels. However, for some MODIS tiles there may be considerable variability within a tile itself. In such cases, RAPT classifier in Stage 1 often learns to identify the dominant type of burn spectral signature, and as a consequence has a high omission error rate on the burned pixels of other non-dominant burn spectral signatures.

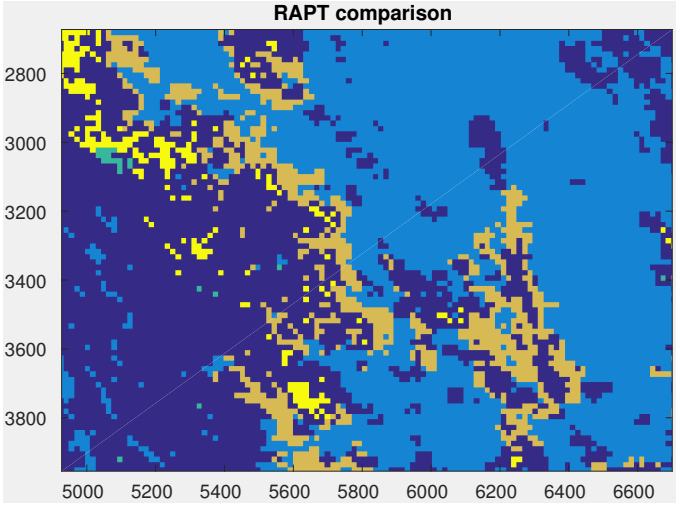


(**a**) Before burn false color composite image on 16 June 2010

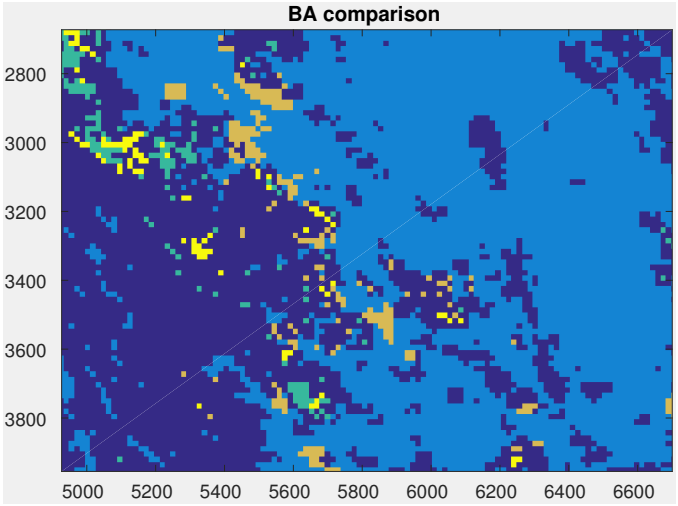


(**b**) After burn false color composite image on 6 October 2010

Figure 5. *Cont.*



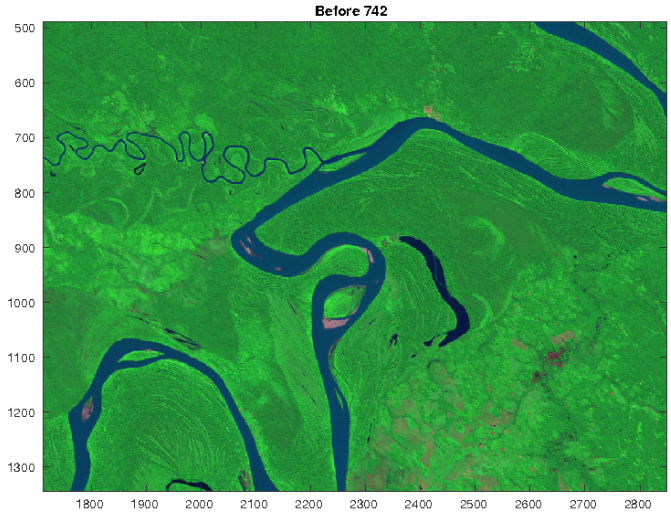
(**c**) Comparison map: RAPT



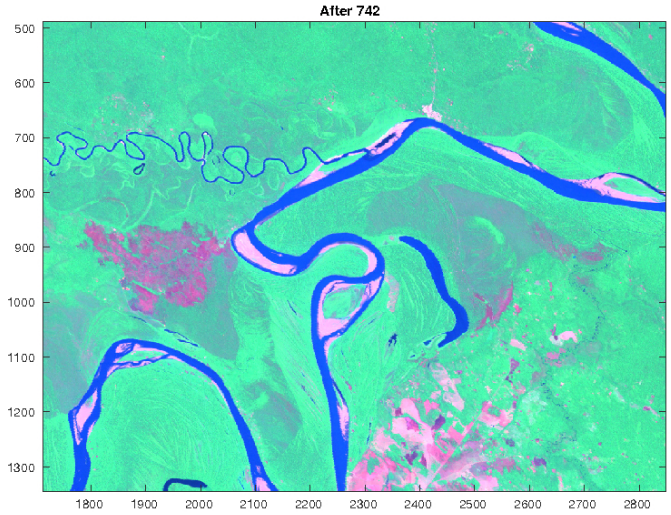
(**d**) Comparison map: MCD64

Figure 5. Figure shows an example of errors of commission based on Landsat-based validation maps. The top row shows before and after false color composites (using bands 7,4,2) of a region with detected burned area (WRS path-232, WRS row-070, MODIS tile name-h11v10, year 2010). The bottom row shows the burned area detected by RAPT and MCD64 relative to Landsat-based ground truth generated using the BAMS scheme. Each comparative image is color coded as yellow for true positives, orange for false positives, green for false negatives and light blue for true negatives. Masked regions (due to detectable cloud cover or non-forest landcovers) are colored dark blue in the comparative images. Note that the region shown has 1000 $\times$ 1600 Landsat pixels.

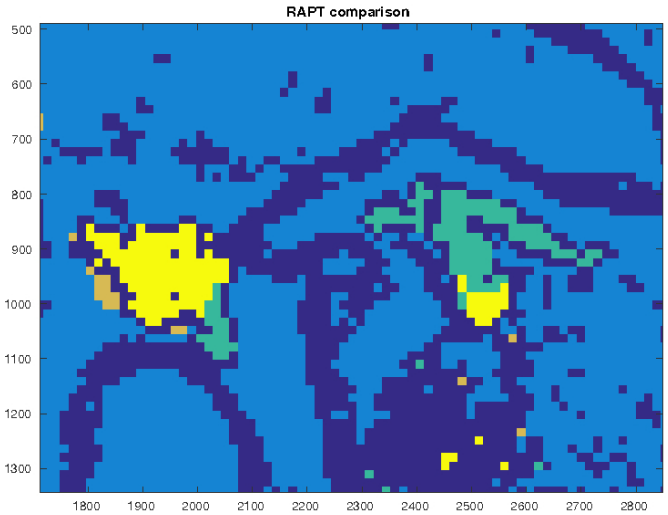
3. Another source of omission errors in RAPT is the dependence on Active Fire signal to identify confident burned pixels. We observe that some burn events, which are identified by Stage 1 classifier, get incorrectly eliminated by RAPT algorithm in Stage 2 because of complete absence of Active fire in the pixels of the burn event. Since none of the pixels belonging to these events have an Active fire presence, there are no confident burned pixels identified in Stage 2. As a result, the event cannot be recovered back in Stage 3 (spatial growing) and gets completely missed in RAPT output. This type of error typically impacts burn events of small spatial extent, as the probability of complete absence of Active fire signal is very small for large burn events.



(**a**) Before burn false color composite image on 7 June 2010

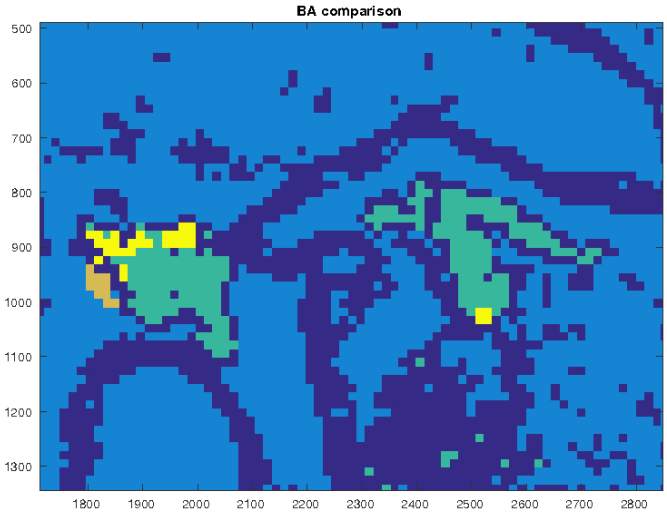


(**b**) After burn false color composite image on 27 September 2010



(**c**) Comparison map: RAPT

Figure 6. *Cont.*



(**d**) Comparison map: MCD64

Figure 6. Figure shows an example of errors of omission based on Landsat-based validation maps. The top row shows before and after false color composites (using bands 7,4,2) of a region with detected burned area (WRS path-233, WRS row-068, MODIS tile name-h11v10, year 2010). The bottom row shows the burned area detected by RAPT and MCD64 relative to Landsat-based ground truth generated using the BAMS scheme. Each comparative image is color coded as yellow for true positives, orange for false positives, green for false negatives and light blue for true negatives. Masked regions (due to detectable cloud cover or non-forest landcovers) are colored dark blue in the comparative images. Note that the region shown has 900 × 1200 Landsat pixels.

5.2.2. Errors of Commission:

1. RAPT algorithm trains a classification model to distinguish between burned and unburned pixels in the forest land class. Therefore, when it is applied to MODIS pixels that are only partially forest and partially some other land class such as cropland, then it may spuriously detect the pixel as burned. We notice that some of the commission errors were present in such pixels of mixed land cover types.
2. The RAPT product is based on the ability of a statistical model to distinguish between burned and unburned pixels based on their spectral data. Soil, smoke and other atmospheric conditions add considerable variability to the spectral characteristics of burned pixels, and sometime result in spurious burn events if there is presence of Active Fire signal in the spatial vicinity.

5.3. Complementarity of RAPT and MCD64 Products

Although Dice coefficient (a widely used metric that takes into account both user's and producer's accuracy) is much higher for RAPT relative to MCD64 for most reference maps in Table 1, these two products should be considered complementary and used together to understand the complete picture. In particular, RAPT and MCD64 operate on different input data and use different algorithms as a result of which some kind of fires might be easier for one algorithm to detect in comparison to the other. For example, short lasting fires that recover in a few weeks might be hard for RAPT to detect (as it uses 8-day MODIS composites), but could still be detected by MCD64 since it relies on daily data. However, long lasting fires (occluded by smoke in the short run) can be difficult to detect for MCD64 as it needs to observe on high quality data in the vicinity of the fire date. Another major issue with MCD64 is that it uses a single set of parameters for the entire globe. Although these parameter choices work well for in areas like North America, they are not necessarily optimal for the tropics where the spectral characteristics of fires are significantly different and noise issues are more prevalent. Hence, a data

adaptive scheme like RAPT could be used to compensate for the large number of such omissions of MCD64 in these regions.

6. Concluding Remarks

We presented results on a new MODIS-based burned area product that has comparable user's accuracy but considerably better producer's accuracy compared to the state-of-art MCD64 in the tropical forests in South America and South-east Asia. Though there is considerable uncertainty in translating burned area into carbon emissions, the significantly larger burned area identified by RAPT will help addressing the current uncertainties in tropical carbon budgets. Moreover, this product can be used to identify and mitigate practices such as slash-and-burn that are often used to illegally clear forested locations for plantations and other commercial activity.

Even though this paper presented the RAPT results only for the tropical forests of Amazon and South-east Asia, the algorithm offers opportunities for global-scale fire mapping. Since, the presented algorithm does not rely on hand-crafted rules or parameters, and automatically adjusts its model parameters to the specifics of different spatial regions and land cover classes, it can be adapted to identify burned areas in other land classes and geographical regions.

In the RAPT framework, the predictive model is trained on homogeneous land cover and geographical regions. The presented implementation of RAPT uses MODIS tile boundaries to address geographical heterogeneity and relies on the MODIS land classification for a good quality forest mask. However, there may be considerable variability within a MODIS tile, and the MODIS land cover labels are coarse (many sub-classes) and may also be incorrect. The ability of the RAPT framework to identify fires with high accuracy is impacted by these issues. Hence, further extensions of this framework are needed that can perform automated data-driven grouping of pixels, instead of relying on land classification products and MODIS tile boundaries.

Supplementary Materials: In the supplementary document, we present the annual burned area reported by RAPT, MCD64 and Active fire for each MODIS tile in our study area. In addition, composite images for Landsat scenes used for generating the reference maps are also provided.

Acknowledgments: This research was supported in part by the University of Minnesota Doctoral Dissertation Fellowship, National Science Foundation under Grants IIS-1029711 and NASA grant NNX12AP37G. Access to computing facilities was provided by the University of Minnesota Supercomputing Institute and NASA Earth Exchange (NEX).

Author Contributions: Varun Mithal, Guruprasad Nayak and Vipin Kumar developed the algorithms, designed the experiments, performed the analyses and wrote the paper. Ankush Khandelwal prepared Landsat-based reference maps using high resolution imagery and contributed to analysis. Ramakrishna Nemani and Nikunj C. Oza provided valuable insights to the paper's results and contributed to discussions.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Appendix A. Details of the Classification Procedure

Appendix A.1. Data Pre-Processing

1. Level-2 multispectral data from MODIS (MOD09A1) along with Active Fire hotspot data (MOD14A2) and landcover data (MCD12Q1) is obtained for each MODIS tile from http://earthexplorer.usgs.gov/.
2. The multispectral reflectance data and Active Fire are available at a temporal resolution of per 8 day (46 time steps in a year). Landcover data is available as a yearly product from 2001 to 2012. Landcover map is used to determine the forested pixels each year. Since, burning might alter the signature of a pixel, we use the landcover map from the previous year to determine forested pixels. Landcover classes from 1 to 5 correspond to different kinds of forests, but we train a single

model for all of them. Also, since the landcover maps are only available until 2012, for 2013 and 2014, landcover map from 2012 is used.

3. The multispectral reflectance data and landcover data are available at a spatial resolution of 500 m, while Active Fire data is available at a spatial resolution of 1 km. In our algorithm, since Active Fire is used as labels to train the classifier over features at 500 m resolution, we downscale Active Fire data from 1 km resolution to 500 m resolution i.e., at each of the 46 time steps in the year, a 500 m pixel gets an Active Fire if the corresponding 1 km pixel had an Active Fire at that time step.
4. To handle the errors of MODIS landcover product, we only consider pixels that the product has consistently masked forests for multiple years. We define a confident forest mask consisting of pixels that have been marked as forests in all of the first 4 years (2001-04). All maps are produced on the set of pixels that belong to this set of confident forests and were also marked forests in the year before the year of consideration.
5. Each of the 7 bands in the reflectance data are z-normalized i.e the mean and the standard deviation for each of the 7 bands is computed across all pixels (and time steps) in the tile and the reflectance values are transformed as $r = (r - m)/s$, where r is the reflectance value in a certain band and m and s are the mean and the standard deviation in the reflectance values in that band for all pixels in the scene across time.

Appendix A.2. Training the Classification Model

1. One model is trained for each MODIS tile. The classifier operates on the set of features corresponding to all 46 time steps of a year for a given pixel and outputs the probability that the pixel experienced a burn activity in that year.
2. The training set consists of equal number of samples of both kinds-Burned (positive class) and unburned (negative class). Every sample corresponds to data for a 500 m pixel for some year in the 14 year time window (2001-14). For each sample I in the training set, we provide the collection of reflectance data for all 46 time steps $X_1^I, \cdots, X_{46}^I$ and a label Y_I which is 1 if an Active Fire hotspot was observed at the pixel during that year and 0 otherwise.
3. To limit the effect of confusion from boundary pixels, while constructing the training set, we mask out pixels that are either at the boundary of a spatially coherent Active fire occurrence or that belong to an Active Fire patch that is too small (<10,500 m pixels in size).
4. The classification model has 2 parameters-β and w. Given the training set, we learn the parameters using the following procedure

    (a) Initialize β to a 8 (7 dimensional feature space + bias term) dimensional vector drawn from a standard Gaussian distribution (mean = 0, std. deviation = 1). Similarly, initialize w to a 47 (46 time steps + bias term) dimensional vector drawn from a standard Gaussian distribution (mean = 0, std. deviation = 1).

    (b) Until the change in the magnitude of β and w from the previous iteration is at less than 1%, repeat the following

        i. $\beta = \beta + learning\_rate \times gradient1$
        ii. $w = w + learning\_rate \times gradient2$

    $gradient1$ is the gradient of the objective function O with respect to $\beta \leq \beta_0, \beta_1, \cdots, \beta_7 >$ and is computed as,

$$\frac{\partial O}{\partial \beta_d} = \sum_{I=1}^{N} (F^I - Y_I) \left(\sum_{t=1}^{46} (w_t f_t^I (1 - f_t^I) X_t^I(d)) \right) - \lambda T \beta_d$$

*gradient*2 is the gradient of the objective function O with respect to $w \leq w_0, w_1, \cdots, w_{46} >$ and is computed as,

$$\frac{\partial O}{\partial w_t} = \sum_{I=1}^{N} (F^I - Y_I)(f_t^I) - \lambda w_t$$

In the above update equations, F^I is the probability of the I^{th} location in the training set being burned in the given year and f_t^I is the per time step score for the I^{th} location at time step t. Given the values of β and w from the previous iteration, these can be computed as,

$$f_t^I = \frac{1}{1 + e^{\left(-\sum_{d=1}^{7} \beta_d X_t^I(d) + \beta_0\right)}}$$
$$F^I = \frac{1}{1 + e^{\left(-\sum_{t=1}^{46} w_t f_t^I + w_0\right)}}$$

learning_rate is a parameter and is fixed using cross-validation. For our analysis, we found 0.01 to work well. Similarly the regularization parameter λ is fixed using cross validation. For our analysis, we found 0.01 to work well.

Appendix A.3. Prediction

1. Once the model parameters have been learned, the probability that a pixel burned in a given year is given by F^I.
2. For each forested pixel, we now have the probability that it burned during each of the 14 years from 2001-14.

Appendix A.4. Fixing the Threshold

1. The stage 1 classification maps for each year are a binary output obtained by thresholding the probability values at each pixel for that year. This threshold is chosen so as to optimize the product of user's and producer's accuracy on the training set.
2. For each threshold γ from a set of threshold choices 0.01, 0.02, ..., 0.99, the product (scaled up to a constant) of user's and producer's accuracy is estimated to be,

$$[\Pr(Y^I = 1 | F^I > \gamma) - noise\_rate]^2 \Pr(F^I > \gamma)$$

where, the three quantities $\Pr(Y^I = 1 | F^I > \gamma)$, *noise_rate* and $\Pr(F^I > \gamma)$ are estimated as,

$$\Pr(Y^I = 1 | F^I > \gamma) = \frac{|\{I | I \in \{1, \cdots, N\} \wedge Y^I = 1 \wedge F^I > \gamma\}|}{|\{I | I \in \{1, \cdots, N\} \wedge F^I > \gamma\}|}$$
$$\Pr(F^I > \gamma) = \frac{|\{I | I \in \{1, \cdots, N\} \wedge F^I > \gamma\}|}{|\{I | I \in \{1, \cdots, N\}|}$$
$$noise\_rate = \frac{\sum_{Y_I \in S} Y^I}{|S|}$$

where S is the set of all instances that lie in the bottom 5% in the training set in terms of their predictions F^I.

3. The threshold that maximizes the product is used to binarize the probability maps to produce the stage 1 classification output.

Appendix A.5. Stage 2

The stage 2 maps is obtained by taking a logical AND of the stage 1 output and the Active Fire map for that year i.e., for each pixel, it is marked burned in stage 2 map only if it was marked burned in stage 1 and it had an Active Fire hotspot sometime in that year. If any of these two conditions is not met, the pixel is marked unburned.

Appendix A.6. Stage 3

The stage 2 output identifies the 'seeds' of the burn activity across the tile. Stage 3 uses these seeds to gather other burned pixels in their spatial proximity that have strong enough burned scar signature in the spectral space. For a given year, if stage 1 marked a pixel as burned as it was less than 5 pixel distance away from a seed pixel, it is marked burned in stage 3.

References

1. Tanimoto, H.; Kajii, Y.; Hirokawa, J.; Akimoto, H.; Minko, N.P. The atmospheric impact of boreal forest fires in far eastern Siberia on the seasonal variation of carbon monoxide: Observations at Rishiri, a northern remote island in Japan. *Geophys. Res. Lett.* **2000**, *27*, 4073–4076.
2. Fuller, D.O.; Fulk, M. Burned area in Kalimantan, Indonesia mapped with NOAA-AVHRR and Landsat TM imagery. *Int. J. Remote Sens.* **2001**, *22*, 691–697.
3. Chen, Y.; Randerson, J.T.; Morton, D.C.; DeFries, R.S.; Collatz, G.J.; Kasibhatla, P.S.; Giglio, L.; Jin, Y.; Marlier, M.E. Forecasting fire season severity in South America using sea surface temperature anomalies. *Science* **2011**, *334*, 787–791.
4. Randerson, J.T.; Chen, Y.; Werf, G.R.; Rogers, B.M.; Morton, D.C. Global burned area and biomass burning emissions from small fires. *J. Geophys. Res. Biogeosci.* **2012**, *117*, doi:10.1029/2012JG002128.
5. Schultz, M.G. On the use of ATSR fire count data to estimate the seasonal and interannual variability of vegetation fire emissions. *Atmos. Chem. Phys.* **2002**, *2*, 387–395.
6. Smith, A.M.S.; Lentile, L.B.; Hudak, A.T.; Morgan, P. Evaluation of linear spectral unmixing and DNBR for predicting post-fire recovery in a North American ponderosa pine forest. *Int. J. Remote Sens.* **2007**, *28*, 5159–5166.
7. Sukhinin, A.I.; French, N.H.; Kasischke, E.S.; Hewson, J.H.; Soja, A.J.; Csiszar, I.A.; Hyer, E.J.; Loboda, T.; Conrad, S.G.; Romasko, V.I.; et al. AVHRR-based mapping of fires in Russia: New products for fire management and carbon cycle studies. *Remote Sens. Environ.* **2004**, *93*, 546–564.
8. Giglio, L.; Loboda, T.; Roy, D.P.; Quayle, B.; Justice, C.O. An active-fire based burned area mapping algorithm for the MODIS sensor. *Remote Sens. Environ.* **2009**, *113*, 408–420.
9. Fraser, R.H.; Li, Z.; Cihlar, J. Hotspot and NDVI differencing synergy (HANDS): A new technique for burned area mapping over boreal forest. *Remote Sens. Environ.* **2000**, *74*, 362–376.
10. Bastarrika, A.; Alvarado, M.; Artano, K.; Martinez, M.P.; Mesanza, A.; Torre, L.; Ramo, R.; Chuvieco, E. BAMS: A tool for supervised burned area mapping using Landsat data. *Remote Sens.* **2014**, *6*, 12360–12380.
11. Loboda, T.V.; Csiszar, I.A. Reconstruction of fire spread within wildland fire events in Northern Eurasia from the MODIS active fire product. *Glob. Planet. Chang.* **2007**, *56*, 258–273.
12. Pu, R.; Gong, P. Determination of burnt scars using logistic regression and neural network techniques from a single post-fire Landsat 7 ETM+ image. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 841–850.
13. Pu, R.; Gong, P.; Li, Z.; Scarborough, J. A dynamic algorithm for wildfire mapping with NOAA/AVHRR data. *Int. J. Wildl. Fire* **2004**, *13*, 275–285.
14. Roy, D.P. Multi-temporal active-fire based burn scar detection algorithm. *Int. J. Remote Sens.* **1999**, *20*, 1031–1038.
15. Mithal, V.; Nayak, G.; Khandelwal, A.; Kumar, V.; Oza, N.C.; Nemani, R. Rapt: Rare class prediction in absence of true labels. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2484–2497.
16. Vermote, E.F.; Kotchenova, S.Y.; Ray, J.P. MODIS Surface Reflectance User's Guide. *MODIS Land Surface Reflectance Science Computing Facility*, version 1; 2011. Available online: http://modis-sr.ltdri.org/guide/MOD09_UserGuide_v1_3.pdf (accessed on 4 January 2018)

17. Giglio, L.; Van der Werf, G.R.; Randerson, J.T.; Collatz, G.J.; Kasibhatla, P. Global estimation of burned area using MODIS active fire observations. *Atmos. Chem. Phys.* **2006**, *6*, 957–974.
18. Padilla, M.; Stehman, S.V.; Chuvieco, E. Validation of the 2008 MODIS-MCD45 global burned area product using stratified random sampling. *Remote Sens. Environ.* **2014**, *144*, 187–196.
19. MOD09A1, MOD14A2, MOD12Q1, MCD64A1 Data Products Were Retrieved from the Online Data Pool, Courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. Available online: https://lpdaac.usgs.gov/data_access/data_pool (accessed on 4 January 2018).
20. Friedl, M.A.; McIver, D.K.; Hodges, J.C.; Zhang, X.Y.; Muchoney, D.; Strahler, A.H.; Woodcock, C.E.; Gopal, S.; Schneider, A.; Cooper, A.; et al. Global land cover mapping from MODIS: Algorithms and early results. *Remote Sens. Environ.* **2002**, *83*, 287–302.
21. Nayak, G.; Mithal, V.; Jia, X.; Kumar, V. Classifying multivariate time series by learning sequence-level discriminative patterns. In Proceedings of the 2018 Society for Industrial and Applied Mathematics (SIAM) International Conference on Data Mining, San Diego, CA, USA, 3–5 May 2018.
22. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94.
23. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sens. Environ.* **2015**, *159*, 269–277.
24. Giglio, L.; Randerson, J.T.; Werf, G.R. Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *J. Geophys. Res. Biogeosci.* **2013**, *118*, 317–328.
25. Padilla, M.; Stehman, S.V.; Ramo, R.; Corti, D.; Hantson, S.; Oliva, P.; Alonso-Canas, I.; Bradley, A.V.; Tansey, K.; Mota, B.; et al. Comparing the accuracies of remote sensing global burned area products using stratified random sampling and estimation. *Remote Sens. Environ.* **2015**, *160*, 114–121.
26. Roy, D.P.; Jin, Y.; Lewis, P.E.; Justice, C.O. Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data. *Remote Sens. Environ.* **2005**, *97*, 137–162.
27. Tansey, K.; Bradley, A.; Smets, B.; van Best, C.; Lacaze, R. The Geoland2 BioPar burned area product. In Proceedings of the EGU General Assembly Conference Abstracts, Vienna, Austria, 22–27 April 2012; Volume 14.
28. Chuvieco, E. ESA CCI ECV Fire Disturbance—Product Specification Document. ESA Fire-CCI project 2013. Available online: https://www.esa-fire-cci.org/webfm_send/932 (accessed on 4 January 2018).