

Research Article

Development of a Point-based Method for Map Validation and Confidence Interval Estimation: A Case Study of Burned Areas in Amazonia

Liana Oighenstein Anderson^{1,2,3*}, David Cheek², Luiz EOC Aragão^{3,4}, Luaê Andere³, Brenda Duarte³, Natália Salazar³, André Lima⁵, Valdete Duarte³ and Egidio Arai³

¹National Center for Monitoring and Early Warning of Natural Disasters – Cemaden, Technological Park of São José dos Campos, Dr. Altino Bondensan Road, 500, São José dos Campos - São Paulo, Brasil

²Environmental Change Institute, University of Oxford, Oxford, OX1 3QY, UK

³National Institute for Space Research – INPE. Av dos Astronautas, 1.751. Jd. Da Granja, São José dos Campos, São Paulo, Brazil

⁴College of Life and Environmental Sciences, University of Exeter, Exeter, EX4 4RJ, UK

⁵Department of Geographical Sciences, University of Maryland, 4321 Hartwick Rd Suite 400 College Park, MD, USA

Abstract

Forest fires and their associated emissions are a key component for the efficient implementation of the Reducing Emissions from Deforestation and Forest Degradation (REDD+) policy. The most suitable method for quantifying large scale fire-associated impacts is by mapping burned areas using remote sensing data. However, to provide robust quantification of the impacts of fire and support coherent policy decisions, these thematic maps must have their accuracy quantitatively assessed. The aim of this research is to present a point-based validation method developed for quantifying the accuracy of burned area thematic maps and test this method in a study case in the Amazon. The method is general; it can be applied to any thematic map consisting of two land cover classes. A stratified random sampling scheme is used to ensure that each class is represented adequately. The confidence intervals for the user's accuracies and for both overall accuracy and area error are calculated using the Wilson Score method and Jeffrey Perks interval, respectively. Such interval methods are novel in the context of map accuracy assessment. Despite the complexity of calculation of the confidence intervals, their use is recommended. A spreadsheet to calculate point and interval estimates is provided for users.

Keywords: Accuracy assessment; Uncertainty; Confidence interval; Burned area; Amazon

Introduction

It must be appreciated that thematic maps are not perfect depictions of reality. The increase in both freely-available remote sensing data, and software for image processing and classification, has boosted thematic mapping in the last 10 years. Due to the key role that thematic maps play for land cover dynamics science and for supporting policy decisions and planning, it is imperative that the accuracy of such maps are formally assessed. The overarching philosophy of a thematic map accuracy assessment is standard: compare the map result with some "true" reference data at selected sample locations, in order to estimate various measures of map accuracy. Stephen and Czaplewski described three basic components of a map accuracy assessment: (i) sampling design, the procedure for selecting sample locations on the map; (ii) response design, the procedure for determining reference land cover at these sample locations, with attention to spatial autocorrelation [1] and (iii) analysis of the data. This framework is widely used, but within this framework there is no standard procedure [2].

Within the sampling design, there is flexibility for choice of sampling method, sampling unit, and sample size. Several sampling methods are available and commonly used in map accuracy assessments, such as simple random, stratified random, clustered, systematic, and non-random sampling. A variety of sampling units can be used which include points [3,4], pixels [5,6], and polygons [7-9]. Sample size is arbitrary.

Three main techniques have been identified for the accuracy assessment of burned areas products [10]: (i) Pixel-level error matrices derived from the comparison of the research map output with some reference map, (ii) Area-level comparison of the proportions of research output map using regression analysis and (iii) the analysis of the landscape patterns. Despite all these methods, it is argued that the generation of accurate maps by one-class classifier is challenging [11]. Moreover, the use of medium-to-low spatial resolution imagery for large scale burned area mapping and the use of higher spatial resolution data for accuracy assessment are likely to underestimate and/-or overestimate the presence or the real extent of the mapped feature [12]. This is due to the inherent characteristics of the sensor, the landscape configuration and the size of burned patches or even the intensity and duration of the fire. To overcome known limitations on the validation of burned area maps, the use of an expert humaninterpreter has been suggested for visual interpretation of burn scars [5,13,14] or more sophisticated methods based on decisions with multiple conflicting objectives, known as Pareto Boundary method [7].

In the early 2000s, many initiatives for mapping global burned area started (GLOBSCAR project, Global Burned Area GBA-2000, ATSR-2 World Fire Atlas, Global VGT burnt area product - L3JRC, MODIS burned area product - MCD45, etc.). However great differences have been reported among these products [13,15-17]. For instance, according to GLOBSCAR, 4,333 km² of forests were burned in Brazil in the year 2000, while based on the GBA-2000, only 846 km² were

*Corresponding author: Liana Oighenstein Anderson, National Center for Monitoring and Early Warning of Natural Disasters – Cemaden, Technological Park of São José dos Campos, Dr. Altino Bondensan Road, 500, São José dos Campos - São Paulo, Brasil, Tel: +551232050154; E-mail: liana.anderson@cemaden.gov.br

Received January 16, 2017; Accepted March 07, 2017; Published March 10, 2017

Citation: Anderson LO, Cheek D, Aragão LE, Andere L, Duarte B (2017) Development of a Point-based Method for Map Validation and Confidence Interval Estimation: A Case Study of Burned Areas in Amazonia. J Remote Sensing & GIS 6: 193. doi: 10.4172/2469-4134.1000193

Copyright: © 2017 Anderson LO, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

J Remote Sensing & GIS, an open access journal ISSN: 2469-4134

detected [18]. The differences in the estimates provided by these two datasets make evident the high uncertainties related to the detection of burned areas. Today, global-long term systematic-mapping of burned areas are provided by MODIS MOD45 and MOD64 products [19]. These products have been validated globally [20] and regionally by different research groups [21-24]. However, the performance of these algorithms for mapping sub-canopy burning, particularly in tropical regions, is deficient in accurately detecting the timing and extent of understory forest fires. The main factors affecting the detection are associated with (i) persistent cloud cover, (ii) limitation on the revisit time of the satellite, (iii) spatial resolution, (iv) lack of consistent field data for depicting and characterizing the impacts on the ground and (v) surface reflectance, which is a function of the interaction between forest structure fire characteristics, such as intensity and duration [25]. It is recognized that there is a need for increasing the quality of the routines used for validation of burned area products [7]. However, the choice of the validation methods depends upon the aimed application of the product [26]. In the context of data for supporting climate change mitigation policies, such as the decree number 7.390/2010, which regulates the Brazilian National Plan on Climate Change (BNPCC), validation routine must be suitable for large scale, operational application. Fires in the Amazon have been quantified as a major source of carbon to the atmosphere [27,28]. During the 2010 anomalous drought, Amazonian fires were responsible for emitting 0.51 ± 0.12 Pg C to the atmosphere [29], a value close to the targeted emission by all sectors proposed by the Brazilian government for the year 2020.

Because of the complex mosaic of land cover changes that emerge rapidly through the use of fire in Amazonia, the performance of automated global burned area methods to detect fire affected area is undermined [30-32]. Facing the challenge of providing burned area information for the Amazon region a great effort has been made for developing a methodology based on the Linear Spectral Mixture Model (LSMM) that more accurately detects burned area in both productive lands and forests, aiming to support the construction of an operational burned area product for the region [28,30,32-34]. The experiences developed in the above mentioned studies pointed out that the global and regional validation schemes, such as the SAFNet protocol developed in Africa burned areas [14], are not adequate for the complexity and dynamics of the Amazonian burn scar characteristics. SAFNet encompasses a (i) spatial sampling of about 3% of the mapped area (burned polygons) and (ii) about 2 Landsat images per year. The characteristics of burned areas in the Amazon region are not only related to land cover change conversion, but also wild fires in forests and savannahs, seasonally flooded grasslands and productive lands. Most of these land covers burns yearly maintaining a complex landscape mosaic. Therefore, a comprehensive validation scheme based on area or polygons would be operationally unfeasible due to time limitation, when compared with the evaluation of random points, which allows a rapid interpretation by the human interpreter. Moreover, the characteristics of the land cover types require multiple images throughout the dry season to fully capture the burned areas dynamics. In 2008, the new policy adopted by the United States Geological Survey (USGS) have made the geocorrected Landsat archive freely available to any user [34], overcoming the limitation on the restricted number of images suitable for validation purposes. Following the above arguments for establishing a burned area monitoring system to subsidize quantification of carbon emission, to guide local action and mitigation policies on fire prevention, such as law enforcement, it is mandatory that thematic burned area maps are associated with a robust accuracy assessment able to explicitly report uncertainties in estimation.

This study, therefore, has two objectives: First, we aim to present a point-based validation method developed for quantifying the accuracy of burned area thematic maps. Secondly, we test this method in a study case of burned areas in the Amazon. The methodology for detecting burned areas in the Amazon is built on the previous experiences of the research group [28,30,32-34]. The development of an operational burned area product with accuracy assessment and confidence interval is a joint initiative in Brazil: the National Institute for Space Research, INPE, and the National Center for Monitoring and Early Warning of Natural Disasters - Cemaden.

Methods

In this section we first provide a detailed description of the accuracy assessment method proposed in this study and subsequently present the study case, where the updated methodology for burned area mapping is described and the validation method applied.

Accuracy assessment

Detecting fire and fire scars in closed canopy tropical forests is more challenging and difficult than in lower biomass land covers such as cerrado and productive lands [7]. Therefore, the accuracy assessment of the burned area product is performed in two steps. In the first step we apply the validation routine to quantify uncertainties in the map of burned area affecting forests. In the second step we repeat the procedure to other land cover types burned. This strategy provides two independent map accuracy assessments.

The method proposed for the accuracy assessment of burned forests follows six steps:

- Generate a forest mask. In the case of the Brazilian Legal Amazon, a forest map can be provided yearly by the INPE-PRODES project [35].
- ii) Select the Landsat scenes where any burned forest is detected.
- iii) Generate 150 random points within the burned forest classification.
- iv) Generate 150 random points within the forest limits excluding the mapped burned forest area within the Landsat scenes selected.
- v) Each random point is evaluated by two interpreters and in a spreadsheet where the points are identified it is marked as correct or incorrect. This spreadsheet also contains the formulas for accuracy assessment (details in next section). Landsat scenes from June to October are used for this procedure.
- vi) The third interpreter (auditor) performs the inspection of the points classified and determines the correct interpretation where the two interpreters' information is different (correct and incorrectly classified). Usually the auditor is a more skilled interpreter.

It should be noted that steps (i) to (iv) are trivial to perform. Whereas step (v) and (vi) requires significant labour cost to implement, as a dataset with reference images and two images interpreters in addition to one auditor is required.

Sampling design rationale

Sampling method: Examples of non-probabilistic sampling are common in map accuracy assessments. However probabilistic sampling is preferable to non-probabilistic because it enables probability

theory to be applied [36]. Implementing a probabilistic sampling design contributes to a scientifically defensible accuracy assessment. Probabilistic sampling designs include systematic, clustered, simple random and stratified random.

Systematic sampling is commonly employed because it is easy to implement. It can be dangerous however if there is periodicity in the population [37]. Also, in general it is difficult to estimate the sampling error [38]. Therefore, systematic sampling is not adopted in the proposed methodology. Clustered sampling is typically motivated by cost. It can reduce the cost of data collection, for example by reducing distance travelled in the case of ground visits [38]. For the purpose of validating the current burn scar detection methodology based on MODIS data, Landsat 5, Landsat 7, Landsat 8 and LISS-3 images are suitable for use as reference data [39]. Therefore, an option would be to use clustered sampling as a means to reduce the total number of images required. It was judged, however, that the slight advantage of fewer Landsat images did not outweigh the disadvantages which comprise increased complexity of the sample design and analysis, and the potential reduction of precision estimate.

Therefore, simple or stratified random sampling is appropriate for this method. Simple random sampling is the simplest to implement and analyse. But it has the flaw that if a particular class has a relatively small area then this class will be underrepresented in the sample. This has the consequence that the class specific accuracy is estimated with relatively poor precision. In the case study presented the burnt area is significantly smaller than the not-burnt area. To overcome this problem, this methodology was constructed based on stratified random sampling based on the map classes (burnt and not-burnt) to ensure that each class is represented adequately.

Sample size: The stage of classifying the sample points according to the reference data has a workload directly proportional to sample size. This workload is the major expense for operationalising the burned area product, and therefore there is a strong motivation to keep the sample size small. On the other hand, a larger sample size would deliver results of superior precision. The sample size of 150 sample points per stratum gives an achievable workload in terms of time for the interpreter to perform the validation, while delivering an acceptable margin of error in estimation. It was calculated that 150 sample points for each class gives user's accuracy with a margin of error of less than 5% with 95% confidence level, providing the user's accuracy is at least 90%. In total 600 random points are generated for assessing the burned area product: (i) forests: burned (150 points) and unburned (150 points) and (ii) non-forests: burned (150 points) and non-burned (150 points).

The margin of error is calculated as follows:

Suppose the sample size within a map class is n and the user's accuracy of the class is p.

Then the margin of error in the estimate of p is (Equation 1):

$$1.96\sqrt{p(1-p)/n} \tag{1}$$

with 95% confidence. If we take n=150 and assume p>0.9, then the error is less than (Equation 2):

$$1.96\sqrt{0.9(1-0.9)/150} = 0.048$$
 (2)

Sample unit: Of the commonly used sample units, points were chosen [1]. The problem with using pixels or polygons as sampling units is that land cover is in fact continuous. Pixels and polygons have spatial extent so may include both classes of land cover. Therefore it could be unclear how a particular pixel or polygon should be classified. This issue is not insuperable, for example, by making use of fuzzy set theory [38,40], but it complicates both the response design and analysis stages.

Page 3 of 9

Using points as sampling units provides no such problem, provided that the class boundaries are well defined. In the case of points, the mathematics involved in the analysis stage is based on a sampling from an infinite population.

Point and interval estimation: The following notation was used for the point and interval estimation procedures:

 n_i is the number of points sampled from map class i;

 X_{ij} is the number of points sampled from map class i which fall in reference class j;

 γ_i is the area of map class i as a proportion of the total map area;

 P_{ii} is the area of the intersection of map class i and reference class j as a proportion of the total map area;

Point estimates: For the accuracy assessment, we estimate the following parameters:

Overall accuracy (Equation 3):

$$p_{11} + p_{22}$$
 (3)

$$\frac{p_{11}}{\gamma_1} \text{ for class 1, } \frac{p_{22}}{\gamma_2} \text{ for class 2}$$
(4)

Producers accuracy (Equation 5):

$$\frac{p_{11}}{p_{11} + p_{21}} \text{ for class 1, } \frac{p_{22}}{p_{12} + p_{22}} \text{ for class 2}$$
(5)

Area estimation error(Equation 6):

ĺ

$$p_{12} - p_{21}$$
 (6)

 p_{ii} / γ_i is the area of the intersection of map class *i* and reference class j, as a proportion of the area of map class i. Thus, one point randomly sampled from map class *i* has probability p_{ij} / γ_i of falling in reference class *j*. Therefore if n_i points are randomly sampled from map class *i*, then the number of these points falling in reference class *j*, n_i , follows a binomial distribution with parameters n_i and p_{ii}/γ_i .

It follows that the maximum likelihood estimate for p_{ii} / γ_i is x_{ii} / n_i This implies the maximum likelihood estimate for p_{ii} is (Equation 7):

$$\hat{p}_{ij} = \gamma_i x_{ij} / n_i \tag{7}$$

All accuracy parameters are functions of the P_{ij} . Using the functional invariance of maximum likelihood estimation, maximum likelihood estimates of all the accuracy parameters can be found by replacing each instance of p_{ij} with \hat{p}_{ij} . For example the producer's accuracy of class 1, $\frac{p_{11}}{p_{11}}$, has maximum likelihood estimate given by (Equation 8): $p_{11} + p_{21}$

$$\frac{\hat{p}_{11}}{\hat{p}_{11} + \hat{p}_{21}} = \frac{\gamma_1 x_{11} / n_1}{\gamma_1 x_{11} / n_1 + \gamma_2 x_{21} / n_2} \tag{8}$$

All parameters with their maximum likelihood estimates are estimated as follows:

J Remote Sensing & GIS, an open access journal

(9)

Overall accuracy (Equation 9): $\frac{\gamma_1 x_{11}}{\gamma_1 x_{11}} + \frac{\gamma_2 x_{22}}{\gamma_2 x_{22}}$

Using accuracy (Equation 10):

$$\frac{x_{11}}{n_1}$$
 for class 1, $\frac{x_{22}}{n_2}$ for class 2 (10)

Producers accuracy for class 1 (Equation 11):

$$\frac{\gamma_1 x_{11} / n_1}{\gamma_1 x_{11} / n_1 + \gamma_2 x_{21} / n_2} \tag{8}$$

Producers accuracy for class 2 (Equation 12):

$$\frac{\gamma_2 x_{22} / n_2}{\gamma_1 x_{12} / n_1 + \gamma_2 x_{22} / n_2}$$
(12)

Area estimation error (Equation 13):

$$\frac{\gamma_1 x_{12}}{n_1} - \frac{\gamma_2 x_{21}}{n_2} \tag{13}$$

Under very general conditions, maximum likelihood estimators possess a number of desirable properties, including consistency, asymptotic normality and efficiency [36].

The overall accuracy describes the proportion of total map area which has been correctly mapped. While this is an overall indicator of map quality it should be noted that on its own it can be misleading. It is possible to have a map of very high overall accuracy with one class poorly represented. User's accuracy describes the proportion of a mapped class area which has been correctly mapped, giving an idea of a class's accuracy from the perspective of the map's user. Producer's accuracy describes the proportion of a true class area which has been correctly mapped, giving an idea of a class's accuracy from the perspective of the map's producer. Note that while user's and producer's accuracy may at first glance seem similar, they are in fact quite distinct and both necessary; the former is a measure of commission error and the latter is a measure of omission error. The area error describes the proportion of total map area by which the map overestimates the area of class 1. Which measures are most important will depend upon the application of the map. For example, if the map is used to estimate the area of burnt forest then the area error will be particularly important [41].

Confidence intervals: The provision of confidence intervals is rare in map accuracy assessments [42-45]. A parameter estimate without an interval can be misleading because it can convey a false impression of certainty. A confidence interval is very useful because it provides valuable information regarding the margin of error in estimation and gives a range of probable values for the parameter. Foody [42] and Strahler et al. [38] recommend that estimates in map accuracy assessments should be accompanied by confidence intervals.

The user's accuracies, p_{ii} / γ_i for class *i*, are simply binomial proportions. Typically accuracy assessments which provide confidence intervals employ the Wald interval method for a binomial proportion. The Wald interval is widely used because of the simplicity of its calculation [46]. However, the Wald method can provide poor coverage and be misleading, especially for proportions close to 0 or 1. The Wilson Score method has been demonstrated to provide intervals of superior coverage and works well even for extreme parameter values [47]. Therefore confidence intervals for the user's accuracy of class 1 and class 2 were calculated using the Wilson Score method.

The Wilson Score interval for class i user's accuracy is given by

(Equation 14):

$$\frac{1}{1 + \frac{1}{n_i} z_{1-\alpha/2}^2} \left[\frac{\hat{p}_{ii}}{\gamma_i} + \frac{1}{2n_i} z_{1-\alpha/2}^2 \pm z_{1-\alpha/2} \sqrt{\frac{1}{n_i} \frac{\hat{p}_{ii}}{\gamma_i} \left(1 - \frac{\hat{p}_{ii}}{\gamma_i} \right) + \frac{1}{4n_i^2} z_{1-\alpha/2}^2} \right] (14)$$

Page 4 of 9

Where:

 $Z_{1-\alpha/2}$ is the $(1-\alpha/2)th$ quantile of the standard normal distribution.

The overall accuracy and area error are the weighted sums of independent binomial proportions. Decrouez and Robinson [48] documented the performance of various confidence interval methods for this case and recommend the Jeffrey Perks interval for a computationally simple method.

Suppose q and r are independent binomial proportions, estimated using sample sizes of n_1 and n_2 respectively. And suppose α is a constant. Then as described by Decrouez and Robinson (2012), the Jeffrey Perks interval for $\alpha q + r$ is defined as follows (Equation 15):

$$\frac{\theta + \frac{z_{1-\alpha/2}^{2}}{2}(u(\alpha+1) + v(\alpha-1) - 2\psi v)}{1 + z^{2}u} \pm \frac{z_{1-\alpha/2}(\theta + z^{2}\Delta)^{\frac{1}{2}}}{1 + z^{2}u}$$
(15)

Where (Equations 16-21):

$$\theta = \alpha \hat{q} + \hat{r} \tag{16}$$

$$u = \frac{1}{4} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$
(17)

$$v = \frac{1}{4} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$
(18)

 $\Delta = u^{2} \left(\frac{1}{4}(\alpha + 1)^{2} + \psi(\alpha - 1 - \psi)\right) + v^{2} \left(\frac{1}{4}(\alpha - 1)^{2} - \psi(\alpha - 1 - \psi)\right) + \frac{1}{2}uv(\alpha + 1)(\alpha - 1) (19)$

$$\psi = \frac{\alpha(n_1\hat{q} + \frac{1}{2})}{n_1 + 1} - \frac{n_2\hat{r} + \frac{1}{2}}{n_2 + 1}$$
(20)

$$\mathcal{G} = u \big((\alpha + 1 - \theta)\theta + (\alpha - 1 - \psi)\psi \big) + v \big(\theta(\alpha - 1) + \psi(\alpha + 1) - 2\theta\psi \big) (21)$$

The Jeffrey Perks interval for overall accuracy is then given by taking $q = p_{11}$, $r = p_{22}$, $\alpha = \frac{\gamma_1}{\gamma_2}$ and then multiplying the limits by γ_2 . The Jeffrey Perks interval for area error is given by taking $q = p_{12}$, $r = p_{21}$, $\alpha = \frac{-\gamma_1}{\gamma_2}$ and then multiplying the limits by $-\gamma_2$.

The producer's accuracies are of the form $\frac{aq}{aq+(1-a)r}$ where q and r are binomial proportions, and $a \in (0,1)$ is a constant. To the authors' best knowledge, a comparison of the performance of different confidence interval methods has not been published for parameters of

The Wald type confidence interval for class i producer's accuracy is given by (Equation 22):

this form. We derived a Wald type confidence interval based on the

first order approximation of the estimator variance.

$$\frac{\hat{p}_{ii}}{\hat{p}_{ii} + \hat{p}_{ji}} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}_{ji}^2 \hat{p}_{ii} (\gamma_i - \hat{p}_{ii})}{(\hat{p}_{ii} + \hat{p}_{ji})^4 n_i}} + \frac{\hat{p}_{ii}^2 \hat{p}_{ji} (\gamma_i - \hat{p}_{ji})}{(\hat{p}_{ii} + \hat{p}_{ji})^4 n_j}}$$
(22)

Simulations were carried out to investigate the performance of this

J Remote Sensing & GIS, an open access journal ISSN: 2469-4134

Page 5 of 9

interval for various fixed parameter values. The simulations indicated that this interval generally performs well enough at our chosen sample size. However for extreme parameter values (i.e., P_{11} and P_{22} close to 1) the interval may provide poor coverage. In order to keep this paper brief the results of the simulations are not detailed here. We recommend for further research that the performance of the Wald type interval is compared with alternative interval methods.

Case study

The Mato Grosso state, southern Brazilian Amazon, the epicentre of the 2010 drought [49] (Figure 1), was selected as a test site for testing the validation procedure for burned areas mapped during the dry season, covering the months from June to mid-October. The burned areas classification followed the methods described below.

Burned area mapping: The methodology for operationally detecting burned area in Amazonia is well established [28,30,34]. In summary, the method uses daily surface reflectance products MOD09GA and MOD09GQ as well as 8-day surface reflectance product, MOD09Q1 and MOD09A1, collection 5 from MODIS dataset. The dates of the images are selected based on the latest day of the month with nadir view and cloud free images for the area of interest. The analysis includes the red (band 1, 620–670 nm) and near infrared (NIR - band 2, 841–876 nm) reflectance bands from MOD09GQ 9 (daily) and MOD09Q1 (8day) products and shortwave infrared (SWIR - band 6, 1628-1652 nm) from MOD09GA (daily) and MOD09A1 (8-day) products.

Although the images derived from MOD09-GQ/Q1 have an original spatial resolution of 231.7 meters and MOD09-GA/A1 have a spatial resolution of 463.3 meters in size, all bands are resampled to 250 m spatial resolution using a nearest neighbor algorithm and standardized to the geotiff format in geographic projection with WGS 84 Datum using the MODIS Reprojection Tool. The high quality of the sub-pixel geolocation accuracy of the MODIS land products could allow the minimum detected area to be 1 pixel. However, soils patches, cloud shade, farm pond or a real burn scar are difficult to differentiate by an image interpreter in isolated pixels. Moreover, for an operational product that may be used for on-the-ground verification by authorities or for official statistics, false positive detections may be more costly as human and financial resources are limited in the Amazon. Therefore, the minimum area in this method is assumed to be approximately 25

ha (4 pixels of 250 m x 250 m).

The spectral bands are used for the Spectral Mixture Analysis (SMA). This technique is based on the selection of endmembers (pure pixels) to generate fraction images by the application of spectral mixture models. The endmembers are selected directly in the image [50]. Unconstrained least-squares solution is used to unmix the MODIS data into three fractions for each date. Each resulting fraction image represents the fractional abundance of endmembers in each pixel. The method for detecting burned areas uses a model based on three endmembers: vegetation, soil and shade [30,34]. For detecting burned areas, which present low reflectance, the shade fraction image is used.

The burned area classification is generated by using image segmentation coupled with region-growing algorithm followed by unsupervised classification and post-classification image edition [34]. The intervention of an image analyst in the post-classification procedure is particularly important for an operational burned area monitoring system for generating a product with minimized omission and commission errors [5], as it will be potentially used for guiding on the ground actions by fire managers, for supporting policies, for official statistics and for the fire science community. Finally, the regions mapped as burned areas in one month are used as a mask in the following dates. Consequently, there is no spatial information on areas that burnt more than once, and the final map represents the cumulative burnt areas over one period, usually the dry season.

We mapped burned areas using daily surface reflectance products and as well as 8-day surface reflectance product, collection 5 from MODIS dataset. Six MODIS tiles are necessary to cover the study area: H11-V9, H11-V10, H12-V9, H12-V10, H13-V9 and H13-V10. The dates of the images were selected based on the latest day of the month with nadir view and cloud free images for the study area (Table 1).

In addition to the images used in the classification procedure, MODIS daily data were also acquired for supporting the burned area interpretation in regions where there were clouds in the daily data, as considerable interpretation skill is needed due to the spectral similarity with burned areas [13], or when the classified polygon were not clear for the interpreter. Three main regions have been identified with a particular necessity of supporting daily images: (i) in the eastern



Figure 1: (a) Land cover classes in 2010 for Mato Grosso state. The forest class encompasses forests and forests regrowth, while non-forests are all the other land cover types; (b) Mapped burned areas from June to mid-October 2010.

J Remote Sensing & GIS, an open access journal ISSN: 2469-4134

boundary of the state, where there is a seasonally flooded area (known as the Bananal floodplain) [51] and wildfires are common [52], (ii) in the Xingu Indigenous reserve, the largest continuous forest fragment in the state, where burned areas occur close to the seasonal rivers due to the location of the traditional communities and (iii) some areas in southern and western areas of the state, where dark soil patches are observed. These images were used for checking burned areas and for guiding the manual edition of the burn scars. This procedure corrected the misclassification of water bodies, cloud shadows and areas with low confidence, minimizing commission and omission errors.

Higher spatial resolution than the used for generating the burned area classification is recommended for reference data as well as the same time period of the data being validated [11]. Fifty two Landsat 5 and Landsat 7 path/rows were selected for the validation procedure, totalling 208 images covering from June to October 2010 (Figure 1).

Accuracy assessment: The accuracy assessment of the 2010 burned area map for Mato Grosso State was performed following the methodology described in section 2.1.

Results and Discussion

Burned area

The cumulative burned area during the dry season in 2010 drought in Mato Grosso affected a total of 96,855 km² (Figure 1). The total area of forests and non-forests (Brazilian Cerrado and productive lands) affected by fires corresponds to 13,773 km² and 83,082 km² respectively (Table 2). A detailed analysis about the burned area and its associated carbon emissions during the 2010 drought for Mato Grosso is presented by Anderson et al. [52].

By considering the area uncertainty provided by the validation scheme, the old growth forest area burned ranged from 13,678 km² to 13,929 km². The extent of the 2010 forest fires in Mato Grosso are close to southern Amazonian intact forests that were affected by fires in 2010 for the first time (13,570 km²), or 73% to 75% of all understory forest fires mapped for 2010 in southern Amazonia (18,499 km²) [25].

Non-forest burned areas were estimated to have affected from 78,089 km² to 83,107 km², an area approximately 6 times larger than the estimates of the 2010 burned forests. It has been estimated that fires

affect Mato Grosso's Cerrado and productive lands on average 16,141 km² yr⁻¹ between 2001 and 2005, summing 80,705 km² over this period [53], a value closer to the 2010 cumulative burned area estimates of this study.

A comparison with the MODIS MOD45 burned area product [20] for the same period detected approximately 69,000 km², about 30% less than the estimate from our analysis. This may be associated with two factors. The first refers to the spatial resolution of the MODIS burned area product: 500 m. The second is likely to be related with the algorithm employed in the product. It has been shown that in areas with Leaf Area Index (LAI) higher than 5, there is a misdetection of burned area worldwide [20].

Accuracy assessment

Two image interpreters and one auditor were selected to carry out the validation scheme, proposed above. The Landsat scenes path/row and the random points used in the accuracy assessment are presented in Figure 2. The results of the audition of sample points are summarised in Table 3. The final result, as it shows in the excel spread sheet is presented in Table 4.

In this method it is possible to identify intuitively, point by point,

| Month | Daily images | 8-day Mosaic | | |
|-----------|----------------------------------|----------------------------------|--|--|
| | Products: MOD09GA and MOD09GQ | Products: MOD09Q1 and MOD09A1 | | |
| June | 28 th | 26 th | | |
| July | 25 th | 28 th | | |
| August | 24 th | 29 th | | |
| September | 26 th | 30 th | | |
| October | 13 th | 16 th | | |

| | Table 1: Dates | of the images | used for ma | apping burned | areas in 2010 |
|--|----------------|---------------|-------------|---------------|---------------|
|--|----------------|---------------|-------------|---------------|---------------|

| | | Burned area (km²) | | | | | |
|------------|-------|-------------------|--------|-----------|---------|--------|--|
| Land cover | June | July | August | September | October | Total | |
| Forest | 157 | 201 | 1,906 | 10,819 | 689 | 13,773 | |
| Non-forest | 6,767 | 3,780 | 15,331 | 55,014 | 2,189 | 83,082 | |

 Table 2: Monthly burned area from June to October in Mato Grosso during the 2010 drought.



Figure 2: Data used for the accuracy assessment. (a) Landsat 5 and Landsat 7 scenes path/row used for the validation scheme – for each path/row at least 3 images were used to cover the entire period; (b) Random points location.

J Remote Sensing & GIS, an open access journal ISSN: 2469-4134

Page 6 of 9

Page 7 of 9

| Forest | Reference burnt | Reference not burnt | Total |
|------------------|-----------------|---------------------|-------|
| Mapped burnt | 126 | 24 | 150 |
| Mapped not burnt | 0 | 150 | 150 |
| Non-Forest | Reference burnt | Reference not burnt | Total |
| | | | |
| Mapped burnt | 134 | 16 | 150 |

 Table 3: Number of sample points falling in each class.

| | Forest areas | | | Non-forest areas | | |
|-------------------------------------|---------------------|---------|---------|---------------------|--------|--------|
| | Confidence interval | | | Confidence interval | | |
| | Estimate | Lower | Upper | Estimate | Lower | Upper |
| Burnt class user's accuracy | 84.00% | 77.30% | 89.01% | 89.33% | 83.38% | 93.33% |
| Not burnt class user's accuracy | 100.00% | 97.50% | 100.00% | 96.67% | 92.43% | 98.57% |
| Burnt class producer's accuracy | 100.00% | 100.00% | 100.00% | 58.52% | 37.55% | 79.48% |
| Not burnt class producer's accuracy | 99.16% | 98.86% | 99.47% | 99.42% | 99.16% | 99.69% |
| Overall accuracy | 99.20% | 97.67% | 99.48% | 96.30% | 92.88% | 98.55% |
| Area error | 0.80% | -0.69% | 1.13% | -2.63% | -6.01% | 0.03% |

Table 4: Final result of the map accuracy for the two area types evaluated: forests and non-forests.

whether the error is due to a misclassification or whether it is due to the mixed land characteristic of that each point. All the points where the classification failed were evaluated with further detail and the type of error was identified. A summary of the type of errors is presented below:

Misclassification due to the difference in spatial resolution from Terra/MODIS to Landsat/TM and ETM+: 20 points

Agricultural areas with dark soils: 4 points;

Burned areas over water bodies (water mask incomplete): 2 points

Landsat scene available prior to the MODIS image used for mapping the burned areas: 1 point;

Burn scar on the Landsat data was not as clear as in MODIS data due to the distance between the date of burning and the day of the available LT data: 2 points;

Point located exactly in the boundary between burned and unburned area: 2 points;

Errors not associated with any identified pattern/justification: 14 points

It is expected that, at least to some extent, a better accuracy may be achieved in the estimation of the burned area as each pixel is decomposed into subpixel fraction through the unmixing method used in this research. It would be necessary to carry out a detailed comparative analysis in order to determine the proportion of the shade fraction depicted in the medium spatial resolution data that could be attributed to one burned pixel or pixels classified as burned in a higher resolution image. The definition of the shade fraction thresholds for configuring burned area in the lower resolution data, however, would have limitations such as the thresholds could not be generalized for the entire Amazon region, it would be dependent on the endmember [54], date of the image, as well as the fire intensity and the time between the fire occurrence and the image acquisition. These same limitations have been observed in the detection of selective logging using the unmixing technique [3]. The selection of the endmember is directly related to the resulting fraction image generated through the unmixing technique. The endmember is a reflex of the reflectance of a pixel as a function of the homogeneous land cover class on the ground, which itself can present a variety of features with distinct spectral response, especially in moderate and low resolution images [54]. This land cover class heterogeneity is likely to be minimized in our method as large water bodies with clear water, used as endmember pixels for the shade fraction image are common in the Amazon.

The use of an error matrix for assessing the accuracy of maps derived from medium and low resolution data validated with high resolution data may not be adequate due to the mixture of classes presented in the low resolution pixel. Exploratory studies aiming the understanding of the implications of the spatial resolution on a dichotomous classification could use the method proposed by Boschetti et al. [14]. In this research, we estimate that approximately 7% of the points misclassified can be attributed to the differences in the spatial resolution of the burned area map, derived from MODIS, in relation to the validation, based on Landsat data.

Conclusions

The methodology for map accuracy assessment proposed in this study has been designed to be simple to implement. It is applicable in a variety of scenarios. Moreover it gives precise estimation of accuracy parameters along with confidence intervals. As far as we know, our proposed confidence interval method is novel in the context of map accuracy assessments, which brings valuable information on the extent of the uncertainty of the map provided. The only "cost" of providing the improved intervals is in the complexity of their calculation, which is minimized by the provision of a spreadsheet along with this paper to the interested users.

An example of the value of the improved confidence intervals can be seen in the user's accuracy for the not-burnt class of the forest region. The Wilson-Score interval used implies that this burnt class user's accuracy is probably between 97.5% and 100%. However, if the

J Remote Sensing & GIS, an open access journal ISSN: 2469-4134

Page 8 of 9

standard Wald confidence interval were used, the lower and upper bounds would both be 100%, conveying the false impression that the user's accuracy is a perfect 100%. For producer's accuracy we use a Wald interval due to a lack of alternatives. A problem can be seen with the interval for the producer's accuracy of the burnt class in the forest area, which has lower and upper bounds of 100%.

The anomalously large width of the interval for the producer's accuracy of the burnt class in the not-forest area is not due to a problem with the confidence interval method. In fact, a large imbalance in the size of burnt and not burnt areas causes the variance of this estimate to be very high. This large width demonstrates the importance of confidence intervals. It brings to our attention our large degree of uncertainty in the producer's accuracy.

An assumption of the accuracy assessment is that the true land cover comprises of exactly 2 classes, and that the boundaries between these classes are well defined. For our case study, a map with classes of burnt and not-burnt, seems a reasonable assumption. However, there could be a sample point whose reference classification is ambiguous. Therefore it would be helpful to have standard protocols to deal with such a situation.

Another assumption we make, which is indeed a fundamental assumption of all accuracy assessments, is that the reference data represents the true land cover. The collection of reference data is a stage in which, if not careful, the integrity of a map accuracy assessment can be compromised. If too many errors creep into the reference data then the work ceases to be an accuracy assessment, instead simply becoming a comparison of two different data sources.

The identified possible causes of reference classification error encompass the temporal and spatial resolution of the images, misleading images and interpreter error. Finally, the method proposed in this research attends the requirements proposed for good practice recommendations for assessing accuracy and estimating area.

Acknowledgments

This work was supported by UK NERC Amazonica grant NE/F005482/1 (http:// www.geog.leeds.ac.uk/projects/amazonica/), Brazil MCTI-PCI 513 (302541/2014-4), CNPq grant 458022/2013-6, NASA-IDS grant NNX14AD31G, the University of Oxford and Internships Office for funding D.C., and the Pan-Amazônia II project http://www.dsr.inpe.br/laf/panamazonia/. The authors thank Joaquim Godoi Filho and Silvio Coimbra for their support as image interpreters in the validation phase. The first author thanks CNPq for the productivity scholarship (309247/2016-0).

Author Contributions

Liana O Anderson, David Cheek and Luiz EOC Aragão conceptualized the research; Valdete Duarte and Egidio Arai processed the MODIS images and performed data quality check; André Lima, Luaê Andere, Brenda Duarte and Natalia Salazar performed the burned area classification and validation. Liana O. Anderson and David Cheek wrote the paper; David Cheek developed the mathematical framework. All authors contributed to the final version of this manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ 37: 35-46.
- Tsutsumida N, Comber AJ (2015) Measures of spatio-temporal accuracy for time series land cover data. Int J Appl Earth Obs Geoinf 41: 46-55.
- Souza Jr C, Siqueira J, Sales M, Fonseca A, Ribeiro J (2013) Ten-Year Landsat Classification of Deforestation and Forest Degradation in the Brazilian Amazon. Remote Sens 5: 5493-5513.

- Adami M, Mello MP, Aguiar DA, Rudorff BFT, Souza AF (2012) A Web Platform Development to Perform Thematic Accuracy Assessment of Sugarcane Mapping in South-Central Brazil. Remote Sens 4: 3201-3214.
- Schroeder W, Ruminski M, Csiszar I, Giglio L, Prins E, et al. (2008) Validation analyses of an operational fire monitoring product: The Hazard Mapping System. Int J Remote Sens 29: 6059-6066.
- Vanonckelen S, Lhermitte S, Van Rompaey A (2013) The effect of atmospheric and topographic correction methods on land cover classification accuracy. Int J Appl Earth Obs Geoinf 24: 9-21.
- Giglio L, van der Werf GR, Randerson JT, Collatz GJ, Kasibhatla P (2006) Global estimation of burned area using MODIS active fire observations. Atmos Chem Phys 6: 957-974.
- Padilla M, Stehman S, Litago J, Chuvieco E (2014) Assessing the Temporal Stability of the Accuracy of a Time Series of Burned Area Products. Remote Sens 6: 2050-2068.
- Mueller-Warrant GW, Whittaker GW, Banowetz GM, Griffith SM, Barnhart BL (2015) Methods for improving accuracy and extending results beyond periods covered by traditional ground-truth in remote sensing classification of a complex landscape. Int J Appl Earth Obs Geoinf 38: 115-128.
- Mack B, Roscher R, Waske B (2014) Can I Trust My One-Class Classification? Remote Sens 6: 8779-8802.
- Foody GM (2008) Harshness in image classification accuracy assessment. Int J Remote Sens 29: 3137-3158.
- 12. Simon M (2004) Burnt area detection at global scale using ATSR-2: The GLOBSCAR products and their qualification. J Geophys Res 109: 1-16.
- Roy DP, Frost PGH, Justice CO, Landmann T, Le Roux JL, et al. (2005) The Southern Africa Fire Network (SAFNet) regional burned area product validation protocol. Int J Remote Sens 26: 4265-4292.
- Boschetti L, Flasse SP, Brivio PA (2004) Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: The Pareto Boundary. Remote Sens Environ 91: 280-292.
- Tansey K (2004) Vegetation burning in the year 2000: Global burned area estimates from SPOT VEGETATION data. J Geophys Res 109: 1-22.
- Jain AK (2007) Global estimation of CO emissions using three sets of satellite data for burned area. Atmos Environ 41: 6931-6940.
- Chuvieco E, Opazo S, Sione W, Del Valle H, Anaya J, et al. (2008) Global burned-land estimation in Latin America using MODIS composite data. Ecol Appl 18: 64-79.
- Boschetti L, Roy D, Hoffmann A, Humber M (2009) MODIS Collection 5.1 Burned Area Product - MCD45. User's Guide p: 30.
- Boschetti L, Roy DP, Justice CO, Giglio L (2010) Global assessment of the temporal reporting accuracy and precision of the MODIS burned area product. Int J Wildl Fire 19: 705.
- Roy DP, Boschetti L (2009) Southern Africa Validation of the MODIS, L3JRC, and GlobCarbon Burned-Area Products. IEEE Trans Geosci Remote Sens 47: 1032-1044.
- Araújo FM, Ferreira LG (2015) Satellite-based automated burned area detection: A performance assessment of the MODIS MCD45A1 in the Brazilian savanna. Int J Appl Earth Obs Geoinf 36: 94-102.
- Tsela P, Wessels K, Botai J, Archibald S, Swanepoel D, et al. (2014) Validation of the Two Standard MODIS Satellite Burned-Area Products and an Empirically-Derived Merged Product in South Africa. Remote Sens 6: 1275-1293.
- Ruiz J, Lázaro J, Cano I, Leal P (2014) Burned Area Mapping in the North American Boreal Forest Using Terra-MODIS LTDR (2001–2011): A Comparison with the MCD45A1, MCD64A1 and BA GEOLAND-2 Products. Remote Sens 6: 815-840.
- 24. Freeborn P, Cochrane M, Wooster M (2014) A Decade Long, Multi-Scale Map Comparison of Fire Regime Parameters Derived from Three Publically Available Satellite-Based Fire Products: A Case Study in the Central African Republic. Remote Sens 6: 4061-4089.
- Morton DC, Le Page Y, DeFries R, Collatz GJ, Hurtt GC (2013) Understorey fire frequency and the fate of burned forests in southern Amazonia. Philos. Trans R Soc Lond B Biol Sci vol: 368
- 26. MODIS Collection 5.1 Burned Area Product MCD45.

Page 9 of 9

- Gatti LV, Gloor M, Miller JB, Doughty CE, Malhi Y (2014) Drought sensitivity of Amazonian carbon balance revealed by atmospheric measurements. Nature 506: 76-80.
- Anderson LO, Aragão LEOC, Gloor M, Arai E, Adami M (2015) Disentangling the contribution of multiple land covers to fire-mediated carbon emissions in Amazonia during the 2010 drought. Global Biogeochem Cycles 29: 1739-1753.
- Gatti LV, Gloor M, Miller JB, Doughty CE, Malhi Y (2014) Drought sensitivity of Amazonian carbon balance revealed by atmospheric measurements. Nature 506: 76-80.
- 30. Anderson LO, Aragão LEOC, de Lima A, de Shimabukuro YE (2005) Detecção de cicatrizes de áreas queimadas baseada no modelo linear de mistura espectral e imagens índice de vegetação utilizando dados multitemporais do sensor MODIS/TERRA no estado do Mato Grosso, Amazônia brasileira. Acta Amaz 35: 445-456.
- Piromal RAS, Rivera-Lombardi RJ, Shimabukur YE, Formaggio AR, Krug T (2008) Utilização de dados MODIS para a detecção de queimadas na Amazônia. Acta Amaz 38: 77-84.
- Cardozo F, Pereira G, Shimabukuro Y, Moraes E (2014) Analysis and Assessment of the Spatial and Temporal Distribution of Burned Areas in the Amazon Forest. Remote Sens 6: 8002-8025.
- 33. Anderson LO, Shimabukuro YE, Defries RS, Morton D (2005) Assessment of Deforestation in Near Real Time Over the Brazilian Amazon Using Multitemporal Fraction Images Derived from Terra MODIS. IEEE Geosci Remote Sens Lett 2: 315-318.
- 34. Shimabukuro YE, Duarte V, Arai E, Freitas RM, Lima A, et al. (2009) Fraction images derived from Terra Modis data for mapping burnt areas in Brazilian Amazonia. Int J Remote Sens 30: 1537-1546.
- 35. INPE (2011) PRODES Project: Monitoramento da Floresta Amazônica brasileira por satélite.
- Kendall M, Stuart A (1979) Advanced Theory of Statistics: Inference and Relationship. 4th edn, Hodder Arnold, London, UK, pp: 733.
- Armitage P, Berry G, Matthews JNS (2001) Statistical Methods in Medical Research. 4th edn, Wiley-Blackwell, London, UK, pp: 832.
- 38. Strahler AHL, Boschetti GM, Foody MA, Friedl MC, Hansen M et al. (2006) Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps. Luxembourg: Office for Official Publications of the European Communities.
- Wulder MA, Masek JG, Cohen WB, Loveland TR, Woodcock CE (2012) Opening the archive: How free data has enabled the science and monitoring promise of Landsat. Remote Sens Environ 122: 2-10.

- Gopal S, Woodcock C (1994) Theory and methods for accuracy assessment of thematic maps using fuzzy sets. Photogramm Eng Remote Sensing (United States) Vol: 60.
- 41. de Klerk HM, Wilson AM, Steenkamp K (2012) Evaluation of satellite-derived burned area products for the fynbos, a Mediterranean shrubland. Int J Wildl Fire 21: 36-47.
- 42. Foody GM (2004) Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. Photogramm Eng Remote Sensing 70: 627-633.
- Kelly A, Kelly N (2014) Validating the Remotely Sensed Geography of Crime: A Review of Emerging Issues. Remote Sens 6: 12723-12751.
- 44. Weiss M, Baret F, Block T, Koetz B, Burini A (2014) On Line Validation Exercise (OLIVE): A Web Based Service for the Validation of Medium Resolution Land Products. Application to FAPAR Products. Remote Sens 6: 4190-4216.
- 45. Olofsson P, Foody GM, Stehman SV, Woodcock CE (2013) Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sens Environ 129: 122-131.
- Wald A, Wolfowitz J (1939) Confidence limits for continuous distribution functions. The Annals of Mathematical Statistics 10: 105-118.
- 47. Agresti A, Coull BA (1998) Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. Am Stat 52: 119-126.
- Decrouez G, Robinson AP (2012) Confidence Intervals for the Weighted Sum of Two Independent Binomial Proportions. Aust N Z J Stat 54: 281-299.
- Marengo JA, Tomasella J, Alves LM, Soares WR, Rodriguez DA (2011) The drought of 2010 in the context of historical droughts in the Amazon region. Geophys Res Lett 38: 1-5.
- 50. Haertel V, Shimabukuro YE, Almeida-Filho R (2004) Fraction images in multitemporal change detection. Int J Remote Sens 25: 5473-5489.
- 51. Hamilton SK (2002) Comparison of inundation patterns among major South American floodplains. J Geophys Res 107: 8038.
- 52. Neves LF, Marimon BS, Anderson LO (2017) Dinâmica de queimadas no Parque Estadual do Araguaia, zona de transição Amazônia-Cerrado, Acta Amazonica. Aticle Under Review.
- DeFries RS, Morton DC, van der Werf GR, Giglio L, Collatz GJ, et al. (2008) Fire-related carbon emissions from land use transitions in southern Amazonia. Geophys Res Lett 35: 1-5.
- Ichoku C, Karnieli A (1996) A review of mixture modeling techniques for subpixel land cover estimation. Remote Sens Rev 13: 161-186.